GABRIEL NIERO DE CARVALHO

PROPOSTA DE TRATAMENTO E MODELAGEM DE DADOS ESPACIAIS PARA USO EM INFRAESTRUTURA DE DADOS ESPACIAIS - IDEs: ESTUDO DE CASO DE MACROBENTOS PARA A ÁREA COSTEIRA DA BAIXADA SANTISTA

São Paulo
2013

GABRIEL NIERO DE CARVALHO

PROPOSTA DE TRATAMENTO E MODELAGEM DE DADOS ESPACIAIS PARA USO EM INFRAESTRUTURA DE DADOS ESPACIAIS - IDEs: ESTUDO DE CASO DE MACROBENTOS PARA A ÁREA COSTEIRA DA BAIXADA SANTISTA

Dissertação de Mestrado apresentada à Escola Politécnica da Universidade de São Paulo para obtenção do título de Mestre em Engenharia.

Área de Concentração:
Engenharia de Transportes – Geoprocessamento

Orientador: Prof. Dr. José Alberto Quintanilha

São Paulo
2013

GABRIEL NIERO DE CARVALHO

PROPOSTA DE TRATAMENTO E MODELAGEM DE DADOS ESPACIAIS PARA USO EM INFRAESTRUTURA DE DADOS ESPACIAIS - IDEs: ESTUDO DE CASO DE MACROBENTOS PARA A ÁREA COSTEIRA DA BAIXADA SANTISTA

Dissertação de Mestrado apresentada à Escola Politécnica da Universidade de São Paulo para obtenção do título de Mestre em Engenharia.

Área de Concentração:
Engenharia de Transportes – Geoprocessamento

Orientador: Prof. Dr. José Alberto Quintanilha

São Paulo
2013

Este exemplar foi revisado e alterado em relação à versão original, sob responsabilidade única do autor e com a anuência de seu orientador.

São Paulo, de Setembro de 2013

Assinatura do autor _____

Assinatura do orientador _____

FICHA CATALOGRÁFICA

Carvalho, Gabriel Niero de
Proposta de tratamento e modelagem de dados espaciais para uso em infraestrutura de dados espaciais – IDEs: estudo de caso de macrobentos para a área costeira da Baixada Santista / G.N. de Carvalho. – versão corr. -- São Paulo, 2013.
120 p.

Dissertação (Mestrado) - Escola Politécnica da Universidade de São Paulo. Departamento de Engenharia de Transportes.

1.Sistema de informação geográfica 2.Geoprocessamento - Baixada Santista (SP);Macrobento – Baixada Santista (SP) I.Universidade de São Paulo. Escola Politécnica. Departamento de Engenharia de Transportes II.t.

Dedicado aos meus pais.

AGRADECIMENTOS

A toda minha família que sempre me apoiou em todos os aspectos.

A minha namorada por ter muita paciência na elaboração deste mestrado.

Ao Prof. Dr. José Alberto Quintanilha primeiramente por acreditar na minha capacidade, além do apoio e amizade.

A Profa. Dra. Sílvia Sartor, que me ajudou diretamente a definir o tema do meu mestrado e me ensinar um pouco da área ambiental, Macrobentos, etc.

A Profa. Dra. Karla Borges, por suas sábias palavras e conhecimento que muito agregaram neste trabalho.

A Mari, do LabGeo (USP), pela transmissão de conhecimento e incentivo nos estudos e a Patrícia Santana, do Programa de Pós Graduação PTR, por sempre me ajudar nas questões de matrícula, etc.

Gostaria também agradecer a algumas pessoas específicas que me ajudaram de alguma forma na construção deste trabalho, em especial: Wolmar Sabino – me ensinou o "mundo" de SIG; Eduardo Nakamura - grande amigo e altamente capaz; Eduardo Kenzo – pela ajuda em vários momentos de dificuldade técnica; André Strauss – por sempre me incentivar a concluir esta tese.

Por fim, à todos os amigos do Laboratório de Geoprocessamento, Sisgraph e Google que participaram desta árdua jornada.

RESUMO

As zonas costeiras são áreas complexas que contemplam ambientes terrestres e marinhos que, além de possuírem enorme riqueza ambiental, também são áreas atrativas aos seres humanos por oferecer alimentos, lazer, negócios, transporte, entre outros. Algumas dificuldades de gerenciamento ocorrem pela complexidade, conflito de interesses e pelo fato de não haver padronização no levantamento de dados e disponibilização para a comunidade científica, órgãos públicos, etc. O uso de geotecnologias pode auxiliar na organização, padronização e compartilhamento destas informações em Atlas Web além de apoiar no planejamento e tomada de decisão pois agregam, em um único ambiente, diversos dados provenientes de fontes distintas. A construção de um modelo de dados espacial voltado à área ambiental, para ser utilizada em Infraestrutura de Dados Espaciais (IDE) é exemplificada a partir da modelagem de um bioindicador, Macrobentos, de qualidade de sedimentos. Este trabalho apresenta as etapas necessárias para a construção de modelo de dados espacial de Macrobentos e emprega a Região Metropolitana da Baixada Santista como referência, além de ilustrar e discutir as principais dificuldades para organizar os dados não padronizados. Conclui-se que a estruturação do conhecimento quando se trabalha com dados ambientais em um modelo é essencial para sua posterior integração em IDE. Constatou-se no processo de modelagem que questões metodológicas relativas ao processo de coleta podem dificultar ou até mesmo inviabilizar a integração de dados provenientes de diferentes estudos. A construção de um modelo de dados espacial e sua posterior publicação via Geoportal, como o apresentado neste estudo, poderá ser utilizado como referência para novas pesquisas com objetivos semelhantes.

Palavras-Chave: IDE. Geoportal. SIG. Baixada Santista. Biomonitoramento. Macrobentos. Modelo de Dados Espaciais. OMT-G.

ABSTRACT

Coastal zones are complex areas that include marine and terrestrial environments. Besides its huge environmental importance, they also attract humans because they provide food, recreation, business, transportation, among others. Some difficulties to manage these areas are related with their complexity, diversity of interests and the absence of standardization to collect and share data to scientific community, public agencies, among others. The use of geo-technologies can be used in the organization, standardization and sharing of this information through Atlas Web and assists planning and decision making issues because it aggregates different files from distinct sources. The construction of a spatial database integrating the environmental business, to be used on Spatial Data Infrastructure (SDI) is illustrated by a bioindicator, Macrobenthos, that indicates the quality of the sediments. This research shows the required steps to build Macrobenthos spatial database based on Santos Metropolitan Region as a reference. Besides, it tries to illustrate the problems related to organize non standardized data. It can be concluded, when working with environmental data, that the structuring of knowledge in a conceptual model is essential for their subsequent integration into the SDI. During the modeling process it can be noticed that methodological issues related to the collection process may obstruct or make impracticable the data integration from different studies of the same area. The development of a database model and its subsequent publication in a Geoportal can be used as a reference for further research with similar goals.

Keywords: SDI. GIS. Geoportal. Macrobenthos. Biomonitoring; Spatial Database Model. OMT-G.

LISTA DE ILUSTRAÇÕES

Figura 1. Hierarquia de IDE, adaptado de (Rajabifard, 2006). 9
Figura 2. Diagrama representando as etapas da IDE, adaptado de (Hjelmager et al., 2008). 10
Figura 3. Arquitetura de uso de serviços, adaptado de (OGC, 2011).....13
Figura 4. Mapa resultante da requisição no *browser* de um serviço WMS.....14
Figura 5. Arquitetura de requisição de serviço, adaptado de (OGC, 2011). 16
Figura 6. Arquitetura resumida de um Geoportal dentro de uma IDE. 18
Figura 7. Composição das seções referentes ao padrão FGDC e suas obrigatoriedades, adaptado de (FGDC, 2011). 20
Figura 8. Processo de Abstração para a construção do Modelo de Dados (Câmara et al., 2005). 27
Figura 9. Níveis de Abstração de Dados Geográficos (Câmara et al., 2005). 27
Figura 10. Notação gráfica do Diagrama de Classe UML (Lisboa Filho, 2000).....29
Figura 11. Diferenciação entre Classes Geográficas e Convencionais (Câmara et al., 2005). 30
Figura 12. Exemplos de Representações de Geo-Campo (Câmara et al, 2005). 30
Figura 13. Exemplo de Representação de Geo-Objeto (Câmara et al., 2005). 30
Figura 14. Armazenamento de Dados Espaciais utilizando CAD (Davis e Oliveira, 2002). 31
Figura 15. Sistema Gerenciador de Banco de Dados. 33
Figura 16. Fluxo de um ETL, adaptado de (Trujillo e Luajn-Mora, 2003). 36
Figura 17. Aplicação de ETL para dados espaciais (Farley, 2009). 37
Figura 18. Exemplo de utilização de ETL Espaciais (Pestana e Silva, 2003). 38
Figura 19. ETL como padronizador de dados para IDE (Hintz, 2009). 39
Figura 20. Contexto Espacial da Região Metropolitana da Baixada Santista no Estado de São Paulo composta pelos municípios: Bertioga, Cubatão, Guarujá, Itanhaém, Mongaguá, Peruíbe, Praia Grande, Santos e São Vicente. 43
Figura 21. Resumo da Metodologia. 49
Figura 22. Esquema Conceitual de dados de Macrobentos. 51
Figura 23. Parte 1 do Esquema Físico de Dados dos Macrobentos. 54
Figura 24. Parte 2 do Esquema Físico de Dados dos Macrobentos 55
Figura 25. Catálogo de Metadados da INDE.....64

Figura 26. Interface Administrativa da ferramenta *GeoNetwork*. 66
Figura 27. Perfil MGB sumarizado (CONCAR, 2009). 67
Figura 28. Interface de pesquisa no *GeoNetwork*. 71
Figura 29. Resultado de requisição URL feita ao Catálogo de Metadados *GeoNetwork*. 72
Figura 30. Arquitetura do Geoportal e suas principais funcionalidades (GeoMedia SDI, 2012). 73
Figura 31. Consultas simultâneas da ferramenta (GeoMedia SDI, 2012). 74
Figura 32. Associação de feições espaciais a um único registro de metadados. 76
Figura 33. *Menu* acesso online integrando o registro de metadados com o dado espacial via URL OGC WMS. 77
Figura 34. Visão Geral do Geoportal. 78
Figura 35. Interface de busca de metadados e janela com descrição detalhada. 78
Figura 36. Adição de serviços externos do IBGE ao Geoportal. 79
Figura 37. Imagens do *Google Maps* como plano de fundo somado aos pontos de coleta de Macrobentos. 80
Figura 38. Imagens do *Google Maps* como plano de fundo somado aos pontos de coleta de Macrobentos. 80
Figura 39. Medições de distância entre determinados pontos de coleta. 81
Figura 40. Arquitetura desenvolvida para o Geoportal. 82

LISTA DE TABELAS

Tabela 1. Estrutura de armazenamento de dados espaciais na tecnologia Oracle. ..34
Tabela 2. Exemplo de criação e armazenamento de dados espaciais em Oracle, adaptado de (Murray, 2009). 35
Tabela 3. Detalhamento do esquema conceitual segundo OMT-G. 51
Tabela 4. Lista de fontes de dados utilizadas. 57
Tabela 5. Tabela original ilustrando os pontos de coletas e alguns parâmetros: coordenadas, volume, sedimento e profundidade (Codesp, 2008). 58
Tabela 6. Tabela original ilustrando os subpontos de coletas e o Táxon por ponto amostral (Codesp, 2008). 59
Tabela 7. Tabela original ilustrando os pontos de coletas no inverno (i) e verão (v). (Heitor, 2002). 60
Tabela 8. Volume coletado para as amostras nas fontes de dados. 62
Tabela 9. Ilustração de termos padronizados (Codesp, 2008). 63

LISTA DE ABREVIATURAS E SIGLAS

AMA - African Marine Atlas

AMBIS - Australian Marine Boundary Information System

AODCJF - Australian Ocean Data Centre Joint Facility

BCG - The Boston Consulting Group

BI - Business Intelligence

CEMG - Comitê de Estruturação de Metadados Geoespaciais

CGDI - Canadian Geospatial Data Initiative

CSDI - Coastal Spatial Data Infrastructure

CMARIS - Coastal and Marine Resource Information System

CONCAR - Comissão Nacional de Cartografia

CSW - Catalog Service Web

DW - Data Warehouse

EML - Ecological Metadata Language

EPUSP - Escola Politécnica da Universidade de São Paulo

ER - Entidade-Relacionamento

ET-EDGV - Especificação Técnica para a Aquisição de Dados Geoespaciais Vetoriais

ETL - Extract Transform and Load

Emplasa - Empresa Paulista de Planejamento Metropolitano

FGDC - Federal Geographic Data Committee

FOSS - Free and Open Source Software

Funai - Fundação Nacional do Índio

GA - Geoscience Australia

GCE - Georgia Coastal Ecosystems

GML - Geography Markup Language

GISER - Geographic Information System Entity Relational

GPS - Global Positioning System

GSDI - Global Spatial Data Infrastructure Association

GUI - Graphical User Interface

HTTP - Hypertext Transfer Protocol

IBGE - Instituto Brasileiro de Geografia e Estatística

ICAN - International Coastal Atlas Network

ICMBIO - Instituto Chico Mendes de Conservação da Biodiversidade

IDE - Infraestrutura de Dados Espaciais

IFO - Modelo Semântico Formal de Banco de Dados

INDE - Infraestrutura Nacional de Dados Espaciais

INSPIRE - European Commission's INfrastructure for SPatial InfoRmation in Europe

IOC - International Oceanographic Data

IODE - Information Exchange Committee of the Internacional Oceanographic Commission

ISO - International Organization for Standardization

LTER - Long Term Ecological Research Network

MBAR - Monterrey Bat Aquarium Research Institute

MDIP - Marine Data and Information Partnership

MEDIN - Marine Environmental Data and Information Network

MGDI - Marine Geospatial Data Infrastructure

MIDA - Marine Irish Data Atlas

MIML - Marine Information Mark-up Language

MMA – Ministério do Meio Ambiente

MSCC - Marine Science Coordination Committee

NCEAS - National Center for Ecological Analysis

NOAA - National Oceanic and Atmospheric Administration

OBIS - Ocean Biogeographic Information System

Odinafrica - Ocean Data and Information Network for Africa

OGC - Open Geospatial Consortium

OLAP - On-line Analytical Processing

OMG - Ocean Mapping Group

OMT-G - Object Modeling Technique for Geographic Applications

OO - Orientado a Objeto

OWL - Web Ontology Language

OWS - OGC Web Services

RDF - Resource Description Framework

RMBS - Região Metropolitana da Baixada Santista

SAD69 - South American 1969

SDI - Spatial Data Infrastructure

SGBD - Sistema Gerenciador de Banco de Dados

SIG - Sistema de Informações Geográficas

SOA - Service Oriented Architeture

UDDI - Universal Description, Discovery and Integration Protocol

UML - Unified Modeling Language

URI - Uniform Resource Identifier

URL - Uniform Resource Locator

UTM - Universal Transversal de Mercator

WCS - Web Coverage Service

WFS - Web Feature Service

WFS-T - Web Feature Service Transaction

WGS-84 - World Geodetic System

WMS – Web Map Service

WMTS - Web Map Tile Service

WPS - Web Processing Service

WSDL - Web Services Definition Language

XML - Extensible Markup Language

SUMÁRIO

1 INTRODUÇÃO 1

1.1 APRESENTAÇÃO 1

1.2 OBJETIVOS 3

1.3 JUSTIFICATIVA 3

1.4 ESTRUTURA DO TRABALHO 5

2 REVISÃO BIBLIOGRÁFICA 7

2.1 INFRAESTRUTURA DE DADOS ESPACIAIS 7

2.2 ARQUITETURA BASEADA EM SERVIÇOS 11

2.2.1 Serviços Web 11

2.2.2 Geo Serviços 12

2.2.3 Geoportal 16

2.3 CONCEITO E IMPORTANCIA DE METADADOS 18

2.3.1 Metadados na área ambiental 22

2.4 BANCO DE DADOS GEOGRÁFICOS 25

2.4.1 Modelagem Conceitual de Dados Geográficos 26

2.4.2 Implementação Física do Banco de Dados Espacial 31

2.4.3 Transformação e Importação de dados 35

2.5 REVISÃO DE APLICAÇÕES ESPACIAIS PARA A ÁREA AMBIENTAL 40

2.5.1 Gerenciamento de Áreas Costeiras e Marinhas 40

2.6 DEFINIÇÃO DA ÁREA DE ESTUDO 42

2.6.1 Região Metropolitana da Baixada Santista 42

2.6.2 Contextualização e importância de Macrobentos 46

3 MATERIAIS E MÉTODOS 48

3.1 INTRODUÇÃO 48

3.2 METODOLOGIA 48

3.3 CONSTRUÇÃO DO MODELO DE DADOS 50

3.3.1 Modelo Conceitual de Dados 50

3.3.2 Modelo Físico 54

3.4 IMPORTAÇÃO DE DADOS GEOESPACIAIS 56

3.4.1 Características dos Dados 56

3.4.2 Dificuldades Práticas 61

3.5 DEFINIÇÃO E PREENCHIMENTO DE METADADOS 64

3.5.1 Criação de Catálogo de Metadados 64

3.5.2 Preenchimento de Metadados 66

3.5.3 Compartilhamento de Informações 70

3.6 CONSTRUÇÃO DO ATLAS AMBIENTAL (GEOPORTAL) 72

3.6.1 Características Gerais 72

3.6.2 Criação de Serviços Geoespaciais 75

3.6.3 Navegação pelo Portal 77

4 RESULTADOS 82

4.1 INTRODUÇÃO 82

4.2 CONSIDERAÇÕES 83

5 CONCLUSÃO 85

6 CONSIDERAÇÕES FINAIS 88

REFERÊNCIAS BIBLIOGRÁFICAS 91

APENDICE A – DESCRIÇÃO DE METADADOS PARA (CODESP, 2008) 102

1 INTRODUÇÃO

1.1 APRESENTAÇÃO

As Zonas Costeiras são áreas complexas, compostas por ambientes terrestres e marinhos, que não somente constituem reservas de biodiversidades como também oferecem aos seres humanos desde alimentos até oportunidades de lazer, negócios, transporte marítimo, pesca, etc.

Segundo Rodríguez e Windevoxhel (1998), a Zona Costeira é:

> "o espaço delimitado pela interface entre o oceano e a terra, ou seja, a faixa terrestre que recebe influência marítima e a faixa marítima que recebe influência terrestre" (tradução do autor).

Já o Atlas Universal do Oceano *Un Atlas (2012*), estima-se que 60% da vida humana encontra-se nestas regiões e que a taxa de crescimento da população humana nestas zonas costeiras é cerca do dobro em relação ao crescimento global, o que só faz aumentar a pressão sobre estas áreas.

Este cenário também intensifica as atividades vinculadas à degradação ambiental, variabilidade climática, entre outros. Como exemplo, o aumento da industrialização na Região Metropolitana da Baixada Santista potencializa a grande quantidade de agentes tóxicos que se acumulam no sedimento e influenciam diretamente as populações faunísticas que vivem nesta região. O impacto em um elo da cadeia trófica, neste caso em específico, Macrobentos, interfere diretamente na atividade pesqueira, por exemplo. Segundo o mesmo *Un Atlas (2012*), estima-se que cerca de um bilhão de pessoas dependem da pesca como principal fonte de proteínas animais.

A informação geográfica em zonas costeiras é essencial, pois agrega o contexto espacial, permitindo a realização de análises. Estas indicam distribuição, concentração e tendências de indicadores ambientais que facilitam o planejamento costeiro. Entretanto, para ser possível utilizar esta potencialidade é imprescindível que fontes de dados sigam normas e padrões garantindo assim o correto compartilhamento e análise.

Uma das principais formas de se compartilhar dados costeiros é através da criação de um portal de dados. Um Geoportal nada mais é do que uma "porta de entrada"

para a busca de uma informação que, quando pensado especificamente na área espacial, se trata de um conteúdo geográfico (geometria, imagem de satélite, metadado, etc.). Possibilita também que organizações, comunidades, etc., possam interagir distribuindo informações padronizadas com o principal objetivo de auxiliar nas tomadas de decisões.

As iniciativas de organização e compartilhamento de informações costeiras, principalmente as internacionais (Estados Unidos, Canadá, Irlanda, Reino Unido e Austrália), estão em processo de desenvolvimento e cada vez mais buscam construir uma ferramenta que possa auxiliar na gestão e na tomada de decisão, por meio da integração de dados e a sua rápida disponibilização por meio, por exemplo, de portais.

No Brasil as iniciativas para construção de Infraestruturas de Dados Espaciais (IDE) que contemplem as questões Costeiras e Marinhas ainda são incipientes. As Cartas de Sensibilidade Ambiental ao Óleo desenvolvidas pela Petrobras (Araujo et al., 2007) e pelo Ministério do Meio Ambiente Brasileiro (MMA), (Ghehardi et al., 2008), avançaram na questão, mas não tratam de aspectos relativos à padronização no levantamento de dados ambientais e disponibilização ampla para facilitar o planejamento costeiro.

A pesquisa desenvolvida pela Escola Politécnica da Universidade de São Paulo (EPUSP), pertencente ao projeto denominado “Atlas Ambiental e Socioeconômico da Baixada Santista, financiado pela FAPESP (Processo nº 2006/51780-2), no Programa de Políticas Públicas”, teve como objetivo desenvolver um atlas ambiental Web contemplando as principais fontes de informações existentes para a Região Metropolitana da Baixada Santista, voltado à fiscalização e licenciamento ambiental, baseado em tecnologias de IDE.

Diretamente relacionado com esta pesquisa, Carvalho et al. (2012), discute sobre a construção de um modelo de dados para ser disponibilizado através do Atlas Ambiental desenvolvido pela EPUSP.

1.2 OBJETIVOS

O objetivo desta pesquisa é propor uma metodologia para organizar dados ambientais, espaciais e alfanuméricos, sem padronização, para que possam ser visualizados por meio de uma interface Web, denominado Atlas Ambiental (Geoportal). Para tanto, esta pesquisa pretende discutir e exemplificar a respeito das etapas necessárias para a construção de um modelo de dados espacial, além das principais dificuldades encontradas no entendimento e importação de dados ambientais.

Para ilustrar este processo foi eleito o tema Macrobentos como estudo de caso, considerado um indicador ambiental, a partir de fontes de dados distintas e sem padrão de codificação nos levantamentos. Além disso, é também objetivo demonstrar as etapas necessárias para a elaboração de um Atlas Ambiental, desde a importação de dados, preenchimento de metadados, construções de serviços geoespaciais até a disponibilização das informações de forma padronizada, seguindo os conceitos de uma Infraestrutura de Dados Espaciais.

A área escolhida para a construção do modelo de dados foi a Região Metropolitana da Baixada Santista (RMBS), por ser relevante sob diversos aspectos ambientais e socioeconômicos (Sartor et al., 2007 e 2009).

O modelo de dados consiste em uma forma estruturada de representar os dados que, posteriormente, pode ser utilizada como referência e estendida para outras regiões e/ou temas ambientais, que não necessariamente, apenas Macrobentos.

1.3 JUSTIFICATIVA

Os dados ambientais quando produzidos, normalmente estão vinculados a projetos específicos e raramente são compartilhados com outras organizações ou instituições, como também, não há um padrão de levantamento, dificultando o acesso e o compartilhamento entre os mesmos, como destaca Rafabifard et al. (2008). O advento de tecnologias como o Sistema de Informações Geográficas (SIG)

se torna um fato importante neste contexto, pois como mostra Rodríguez et al. (2008), é possível realizar a integração de diversos tipos de dados, de diferentes épocas e contexto possibilitando uma análise conjunta com fácil acesso às informações.

Especificamente no caso de Macrobentos, há uma escassez muito grande de estudos, principalmente que abordem a padronização de levantamentos, organização, documentação e disponibilização dos dados. A maioria das informações existentes na Região Metropolitana da Baixada Santista encontra-se em formato analógico (teses de mestrado e/ou doutorado, estudos técnicos específicos) ou formato digital (Microsoft Excel ou PDF), elaborados para propósitos específicos.

O fato de existir esta enorme lacuna na quantidade e disponibilização de informações sobre o tema Macrobentos na região, aumenta o interesse em propor uma metodologia que possibilite a agregação dos estudos existentes e também sirva como referência para futuras pesquisas com o intuito de respaldar avaliações e análises vinculadas à fiscalização, licenciamento, análise temporal de evolução da degradação ambiental, etc., por parte de órgãos públicos, cientistas, técnicos, entre outros.

A comunidade de organismos Macrobentos, estudo de caso desta pesquisa, é muito importante em estudos ambientais, pois é utilizada como indicador do grau de degradação/ conservação do ambiente. Por meio destes seres vivos é possível inferir se determinada área possui uma concentração de poluição, por exemplo. É essencial, portanto, que se entenda como se dá sua distribuição ao longo da Zona Costeira da Baixada Santista com o intuito de se planejar melhor políticas públicas vinculadas à preservação do Meio Ambiente.

Umas das formas encontradas para disponibilizar estas informações à sociedade em geral, que normalmente não é familiarizada com os conceitos de Sistema de Informações Geográficas, é através das Infraestruturas de Dados Espaciais. Esta estrutura deve ser a mais transparente de forma que seja possível compartilhar dados, processos, análises, entre outros.

Entretanto, um dado proveniente de um estudo ou levantamento específico para ser compartilhado por meio de um Atlas Ambiental Web, seguindo os conceitos de uma IDE, deve seguir uma série de etapas até que esteja apto para ser utilizado pela

comunidade, representada por órgãos públicos, pesquisadores e população em geral. A construção de uma metodologia se faz necessária para ilustrar as diversas etapas na construção de um modelo de dados espacial referente ao tema Macrobentos. Esta informação, uma vez organizada, poderá ser integrada por qualquer Geoportal que siga os conceitos da IDE.

1.4 ESTRUTURA DO TRABALHO

O capítulo introdutório busca resumir o que é apresentado na pesquisa, já o capítulo “2 Revisão Bibliográfica” aborda a revisão bibliográfica e a fundamentação teórica da pesquisa. Foi feita uma divisão em seis grandes temas expressos pelos itens: “2.1 Infraestrutura de Dados Espaciais” que trata dos conceitos envolvidos em uma IDE e as formas de estruturá-la; “2.2 Arquitetura baseada em serviços”, que discute sobre as principais tecnologias e arquiteturas utilizadas em geoserviços; “2.3 Conceito e importância de Metadados”, onde é feito um resumo sobre o conceito de metadados e alguns dos principais padrões de armazenamento com enfoque em dados espaciais; “2.4 Banco de Dados Geográficos” aborda, por meio de subitens, os conceitos envolvidos na construção e no armazenamento de dados espaciais em Sistema Gerenciador de Banco de Dados, além de destacar processos de transformação de dados, inerentes ao processo, quando se deseja integrar diversas fontes de dados distintas; “2.5 Revisão de aplicações espaciais para a área ambiental” é feita uma abordagem específica de algumas iniciativas de Infraestrutura de Dados Espaciais para a área ambiental; “2.6 Definição da Área de Estudo” é discutido acerca da área escolhida e os principais fatores que podem causar degradação ambiental, além da contextualização e destaque dos organismos Macrobentos como bioindicador ambiental.

O capítulo “3 Materiais e Métodos” aborda a metodologia utilizada como referência para a construção da pesquisa, além da apresentação do modelo teórico e do modelo físico de banco de dados para Macrobentos. Já o capítulo “4 Resultados” expressa as etapas realizadas para a construção do Atlas Ambiental desde a importação dos dados, criação dos geoserviços e a interface do Geoportal. O capítulo “5 Conclusão” reflete sobre os resultados obtidos e por fim, o capítulo “6

Considerações Finais" aborda temas não discutidos diretamente nesta pesquisa, mas que podem ser um ponto de partida para futuras pesquisas.

2 REVISÃO BIBLIOGRÁFICA

2.1 INFRAESTRUTURA DE DADOS ESPACIAIS

A era da informação está em desenvolvimento; o rápido acesso a informações pode ser feito praticamente em qualquer lugar do mundo que exista acesso a Internet. Embora este seja um cenário interessante, a mudança de paradigma faz com que algumas novas dificuldades, nunca antes pensadas, se exacerbem e as resoluções das mesmas são parte da evolução da tecnologia.

Diante deste contexto, o dado geográfico se apresenta como um dos mais críticos elementos para rápido consumo entre diferentes pessoas, organizações, cientistas, etc. Interessante discussão é abordada no estudo realizado pelo grupo *The Boston Consulting Group* (BCG), que ilustra como a demanda por dados geoespaciais é crescente no mercado americano, não só ambiente corporativo, mas também na população que utiliza de informação geográfica para tomar decisões rotineiras baseadas no contexto espacial, seja para buscar uma melhor rota para sua residência, um bom restaurante ou até mesmo a localização geográfica de entes próximos (Henttu et al., 2012).

Fatores como este exposto anteriormente só ilustram o aumento da demanda por dados geográficos, todavia, em um passado recente, a maioria das organizações provedoras de dados espaciais não tinham a real preocupação em produzir informações de uma forma que pudessem ser consumidas por pessoas que não as construíram. Fatores como: documentação falha de processos para construção de dados, armazenamento em formatos proprietários de *softwares* e, muitas vezes desorganizados, fazem com que exista uma real dificuldade no compartilhamento e troca de informações espaciais.

Diante deste contexto, o advento do *Spatial Data Infrastructure* (SDI), ou em português, Infraestrutura de Dados Espaciais, tem como principal objetivo, facilitar e coordenar a troca e o compartilhamento de dados e serviços espaciais para melhor atingir as diversas necessidades em diferentes níveis político e administrativo (Hjelmager et al., 2008; Rajabifard et al., 2006).

Uma definição interessante sobre uma IDE pode ser encontrada no Global Spatial Data Infrastructure (GSDI), uma associação que promove a cooperação internacional e colabora com a criação de IDEs em diferentes escalas. Produz também *GSDI Cookbook*, espécie de guia para a construção de uma IDE, conforme citação abaixo:

> "O termo Infraestrutura de Dados Espaciais (IDE), é frequentemente usado para designar a coleção de tecnologias, políticas e arranjos institucionais que facilitem a viabilidade e o acesso aos dados geográficos. A IDE fornece uma base para a descoberta de dados geográficos, avaliação e aplicação para os usuários e provedores em todos os níveis de governo, setor comercial, setor sem fins lucrativos, universidades e cidadãos em geral." (GSDI, 2004, tradução do autor).

Uma IDE é muito mais do que simplesmente um conjunto de amostras ou banco de dados e vai além do levantamento e mapeamento; na prática fornece um ambiente onde organizações/ nações interagem, via tecnologia, para promover o uso, gerenciamento e produção de mecanismos para busca, acesso à dados, além de serviços ou *softwares* para suportar a comunicação (GSDI, 2004; Rajabifard et al., 2006).

Inicialmente, as Infraestruturas de Dados Espaciais contavam com uma metodologia de modelagem baseada na estrutura *top-down (*Rajabifard et al., 2006). Após alguns desenvolvimentos, percebeu-se que esse modelo não mais satisfazia às necessidades intrínsecas a uma IDE. Neste momento houve uma evolução para a segunda geração de IDE, na qual a modelagem transformou-se à uma orientação a processos e sua visão denominada *bottom-up*, a partir de Infraestruturas de Dados Espaciais locais para subsequente construção de uma estrutura nacional ou global (Rajabifard et al., 2006). Crompvoets et al. (2004) discorrem sobre algumas experiências desenvolvidas em diversos países e reafirmam esta tendência da mudança de estratégia da primeira geração de IDEs, orientada a produtos e dados, para uma estrutura mais orientadas às aplicações e seus usuários.

Iniciativas que privilegiam os conhecimentos locais são valorizadas pela riqueza de aplicações e diversidade de interesses impondo novos requisitos às IDEs. Diferentemente das questões relativas às propostas de IDE nacionais, as propostas locais requerem um maior detalhamento e, com isso, um maior acesso a diferentes fontes de dados, mantidas por diversos provedores (Davis e Alves, 2005).

A tendência de popularização da utilização de dados geográficos, decorrente das facilidades promovidas pelos serviços disponibilizados pela Web, tem feito com que,

recentemente, haja uma convergente demanda por modelos de Infraestruturas de Dados Espaciais locais orientadas a serviços (Davis e Alves, 2005; Vaccari et al., 2008; Chan et al., 2001). Os autores, ao analisarem as diversas definições de IDE, já mencionavam a falta de uma perspectiva de serviços nessas definições.

Masser et al. (2008) mencionam que o desenvolvimento de Infraestruturas de Dados Espaciais efetivas deve servir de forma transparente como suporte para a vasta maioria da sociedade que não é familiarizada com conceitos sobre consultas espaciais. Essas estruturas devem ser aprimoradas de tal forma que seja possível compartilhar, adicionalmente aos dados, estratégias, processos, operações, produtos de valor agregado, dentre outros.

As IDEs são construídas, portanto, para possibilitar a troca de informações em diversas escalas, conforme é possível observar na Figura 1 a seguir:

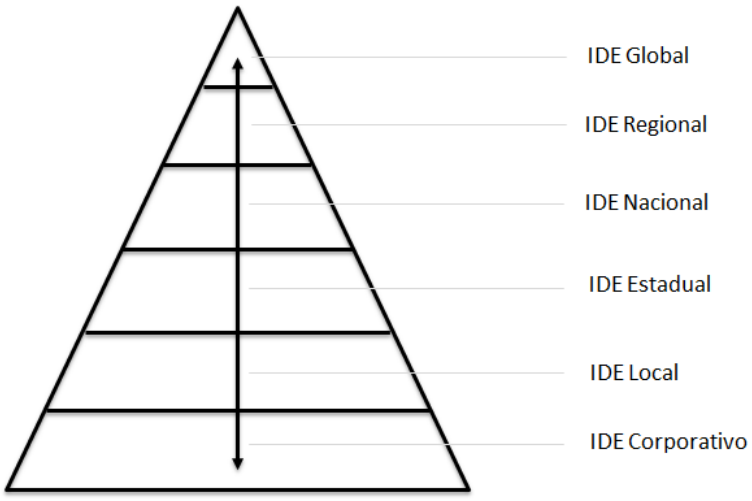


Figura 1. Hierarquia de IDE, adaptado de (Rajabifard, 2006).

Embora esta figura ajude a explicar a estrutura de uma IDE, é importante ressaltar que atualmente não existe uma real hierarquia entre as mesmas; de alguma forma, as Infraestruturas de Dados Espaciais se conectam e são construídas de forma paralelas.

A Figura 2 ilustra as principais etapas existentes na concepção de uma IDE, mas que, dependendo da aplicação, ainda pode conter com mais subdivisões:

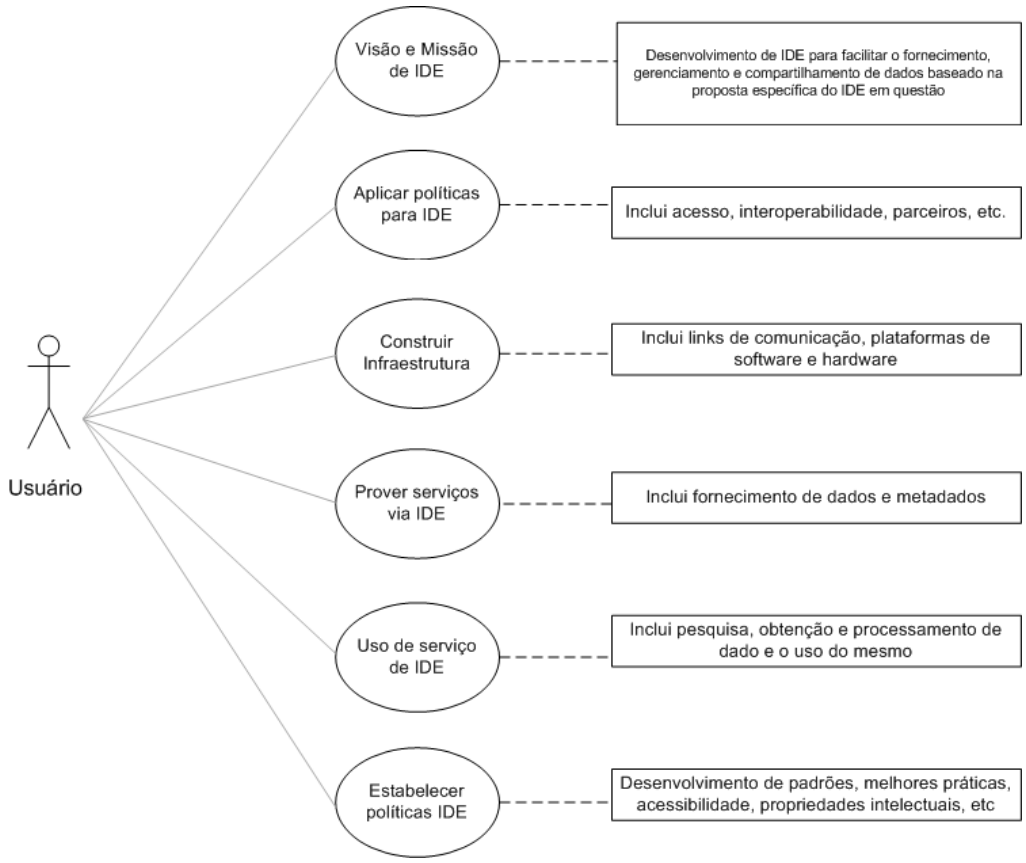


Figura 2. Diagrama representando as etapas da IDE, adaptado de (Hjelmager et al., 2008).

Este usuário destacado na Figura 2 pode assumir diversos papéis, dependendo do tipo de interação que será feita com a IDE, segundo Hjelmager et al. (2008):

- Produtor: usuário que produz dados para uma IDE;
- Provedor: usuário que provê dados ou serviços para ser consumido por uma IDE;
- Desenvolvedor: usuário que adiciona novas funcionalidades para um produto existente ou desenvolve um novo e torna isto disponível na IDE;
- Usuário Final: usuário que utilizará a IDE para um determinado propósito.

Os itens discutidos a seguir neste capítulo são intrínsecos na criação de uma IDE e esta pesquisa utilizou destas metodologias para a construção do modelo de dados, publicação de informações utilizando geoserviços e a documentação por meio de metadados geoespaciais.

2.2 ARQUITETURA BASEADA EM SERVIÇOS

2.2.1 Serviços Web

A construção de uma IDE tem, conforme citado anteriormente, como um dos principais objetivos facilitar o acesso e a troca de informações espaciais provenientes de diversas fontes de dados. Além disso, deve conter ferramentas que permitam ao usuário realizar buscas por informações espaciais específicas, conforme interesse.

A heterogeneidade de dados espaciais atualmente não é uma exceção e pode ser classificada como: sintática (formatos distintos de dados), estrutural (diferenças de esquemas) e semântica (diferença de termos em contextos específicos). Tais informações espaciais são provenientes de diversas bases de dados, que na maioria das vezes, se encontram em distintos formatos (normalmente associados à softwares específicos), escalas e precisão cartográfica diferentes, o que torna a integração e o compartilhamento de dados espaciais algo não tão trivial (Vaccari et al., 2008).

Especificamente em relação à questão sintática, algumas tentativas foram desenvolvidas no sentido de se criar formatos neutros, independentes de tecnologia, como o *OpenGIS* e *Geography Markup Language* (GML), que permitisse que o SIG escolhido dispusesse apenas de um leitor/conversor de um formato de dados específico à outro, como ilustra Câmara e Junior (2002). Todavia, esta metodologia, sem o apoio da organização e controle de um Sistema Gerenciador de Banco de Dados, faz com que múltiplas versões do mesmo dado sejam criadas gerando uma grave ingerência de informação.

Pelos motivos discutidos, tal arquitetura praticamente inviabilizaria qualquer iniciativa de implementação de IDEs. A busca, portanto, por uma plataforma mais flexível e aberta era necessária e o uso de arquiteturas de sistemas distribuídos, baseados em serviços e independentes de tecnologia SIG é exatamente o principal alicerce da criação e desenvolvimento de IDEs.

Arquitetura Orientada a Serviços ou, em inglês, *Service Oriented Architeture (SOA*), tem sido bastante utilizado por facilitar a integração dos serviços entre vários sistemas e também a possibilidade de escalonar o sistema sem modificar o núcleo da aplicação. Davis e Alves (2005) exemplificam este conceito:

> "Serviços, acompanhados de suas descrições e operações fundamentais, tais como descoberta, seleção e chamada, constituem a base da SOA. A arquitetura suporta sistemas grandes com compartilhamento de dados e de capacidade de processamento, através da alocação distribuída de aplicações e recursos computacionais através de redes de computadores. (...) Serviços podem ser primários, se programados em uma linguagem de programação qualquer e independentes de outros serviços, ou compostos, se parte do seu processamento depender de outros serviços".

Tais serviços Web ou também conhecidos como, em inglês, *Web Services* são parte integrante desta arquitetura, utilizando os principais padrões abertos adotados mundialmente na Internet, e facilitam, portanto, o intercâmbio de dados entre sistemas, aplicativos, etc. Alguns dos principais formatos são:

- *Hypertext Transfer Protocol* (HTTP): protocolo para conexão, distribuição, colaboração, transferência e comunicação (HTTP, 2013).
- *eXtensible Markup Language* (XML): formato baseado em texto simples para representar/descrever informações estruturadas como documentos, dados, configuração, transações, entre outros (XML, 2013).
- *Web Services Definition Language* (WSDL): é um arquivo no formato XML usado para descrever os serviços (WSDL, 2001).
- *Universal Description, Discovery and Integration Protocol* (UDDI): é um protocolo que permite promover, usar e compartilhar serviços via Web (UDDI, 2006);
- *Uniform Resource Identifier* (URI): é o pilar da arquitetura Web pois provê a identificação única que é utilizada por todos na Web (URI, 2004).

2.2.2 Geo Serviços

Especificamente para a questão espacial, o *Open Geospatial Consortium* (OGC), consórcio internacional envolvendo mais de 473 empresas (agências governamentais, universidades, indústria geoespacial), propôs a criação de uma arquitetura aberta para distribuição de informações e funcionalidades geográficas

buscando, principalmente, garantir a interoperabilidade de dados permitindo que diferentes sistemas tenham a capacidade de trocar informações e instruções em tempo real, sem necessidade de transformação de dados.

Tal arquitetura aberta é comumente denominada de *framework* de serviços da OGC e composta por três operações principais, como mostra a Figura 3:

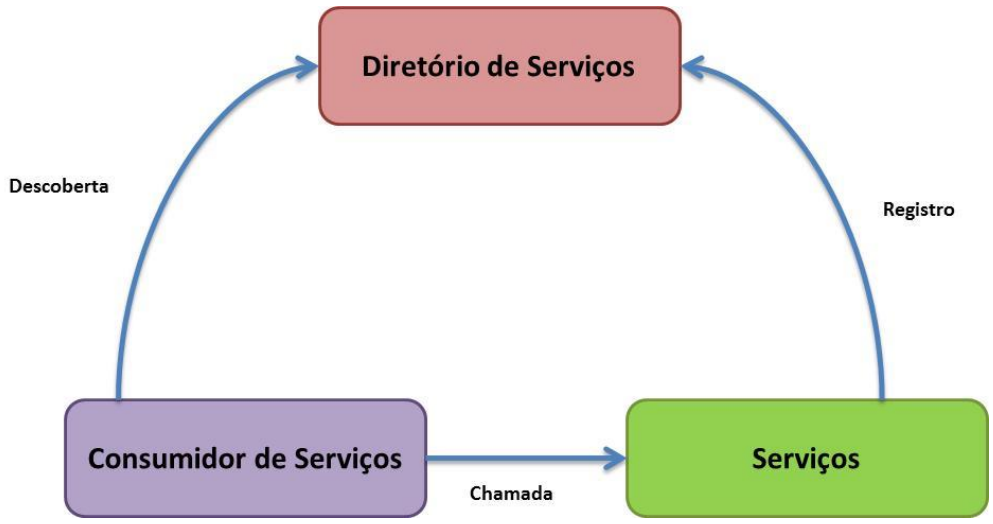


Figura 3. Arquitetura de uso de serviços, adaptado de (OGC, 2011).

- **Serviços**: publicação dos serviços e dados para um diretório (como registro e catálogo) e posterior fornecimento para os consumidores.
- **Consumidor de Serviços**: realiza operações de descoberta no gerenciador (*broker*) para encontrar serviços específicos que necessita. Normalmente a pesquisa é feita sobre os metadados para se buscar aquilo que se deseja.
- **Diretório de Serviços**: uma espécie de intermediário entre o provedor do serviço e o consumidor.

O *framework* da OGC é responsável por identificar serviços, interfaces e intercâmbios de protocolos que podem ser usados por qualquer aplicação que tenha sido desenvolvida para ler determinado padrão. Tais instruções seguem estrutura pré-definida no arquivo XML e as requisições também são padronizadas. Por exemplo, a URL abaixo mostra uma requisição sendo feita na interface cliente seguindo o padrão OGC, para um servidor específico do *software Map Server*:

http://demo.mapserver.org/cgibin/foss4g?&SERVICE=WMS&VERSION=1.1.1&REQUEST=GetMap&LAYERS=OSM_Denver&STYLES=&SRS=EPSG:4326&BBOX=-105.208290,39.542378,-104.769779,39.980889&FORMAT=image/png

A *Uniform Resource Locator* (URL) segue, como supracitado, o padrão OGC e cada parâmetro tem uma função (instrução) específica, conforme detalhado abaixo:

- http://demo.mapserver.org: servidor de onde está sendo feita a requisição;
- SERVICE = WMS: representa o tipo de serviço OGC, neste caso, WMS;
- VERSION = 1.1.1: versão do serviço;
- REQUEST = GetMap: representa a requisição do mapa. Há também outras formas de requisição como o getcapabilities que descreve alfanumericamente as propriedades contidas no serviço específico;
- LAYERS = OSM_Denver: descreve qual é a feição cartográfica desejada;
- SRS = EPSG[1]: 4326: parâmetro que descreve o Sistema de Coordenadas, neste caso em específico, Sistema de Coordenadas Geográfico e Datum WGS84;
- BBOX = -105.208290,39.542378,-104.769779,39.980889: definição da área geográfica desejada;
- FORMAT = image/png: parâmetro que define o formato do resultado, que neste caso é png.

A Figura 4 representa o resultado da requisição:

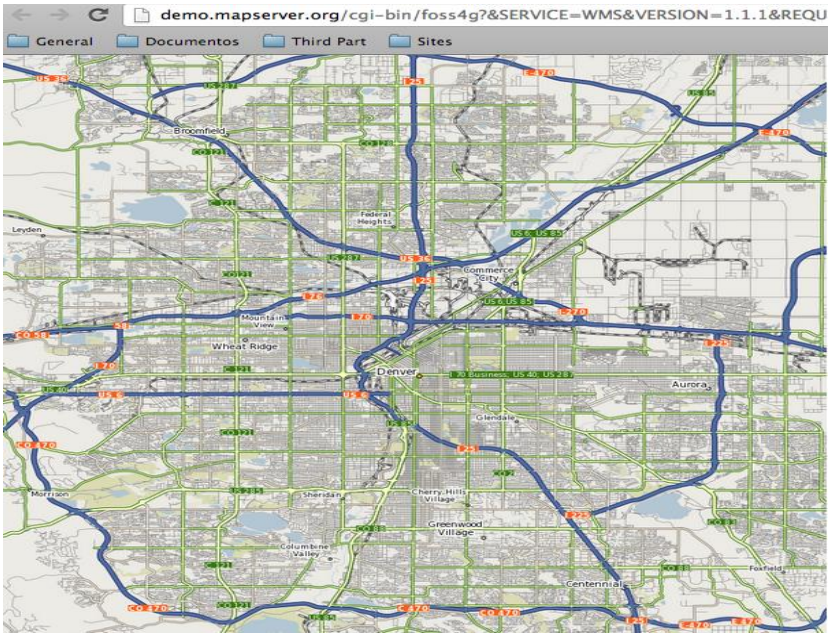


Figura 4. Mapa resultante da requisição no *browser* de um serviço WMS.

[1] Para mais detalhes sobre o EPSG, observar: http://www.epsg.org

Alguns dos principais serviços providos pela OGC, denominados também de OGC Web Services (OWS) ou mesmo Geoserviços, são:

- *Web Map Service* (WMS) - retorna as feições geográficas no formato *raster* (imagem);
- *Web Feature Service* (WFS) - retorna as feições geográficas no formato vetorial;
- *Web Feature Service Transaction* (WFS-T) - é complementar ao WFS, porém suporta operações de edição cartográficas como inserção, edição e exclusão;
- *Catalog Service for Web* (CSW) - provê as informações contidas nos metadados, permitindo a pesquisa, consumo, *download*, etc., de dados;
- *Web Coverage Service* (WCS) - é representado pelo formato *raster* e representam fenômenos com variação no tempo/espaço, como dados climáticos, por exemplo;
- *Web Processing Service* (WPS) - pode-se ser resumido com uma sequência de regras que devem ser seguidas a partir da requisição do usuário final. Tais requisições podem variar de análises *raster*, como uma detecção de mudança temporal ou mesmo o cruzamento espacial de vetores;
- *Web Map Tile Service* (WMTS) - semelhante ao WMS, porém tem como principal objetivo pré-processar o dado *raster* (imagem), também denominado de cache, para garantir o aumento de desempenho no consumo da informação.

Para mais detalhes sobre estes e outros serviços mais específicos no padrão OGC, atentar-se para OGC (2011).

A Figura 5 ilustra a arquitetura de requisição de um serviço específico feito pelo usuário cliente. Uma vez que o usuário, por exemplo, altera o valor de escala se aproximando ou se afastando do mapa, automaticamente uma requisição é enviada para o servidor de aplicações (normalmente um *software* de mercado ou livre como: *Intergraph GeoMedia WebMap*, *Esri*, *MapServer*, etc.) que é responsável em realizar a operação de busca no Banco de Dados Geográfico e retornar o resultado ao Geoportal.

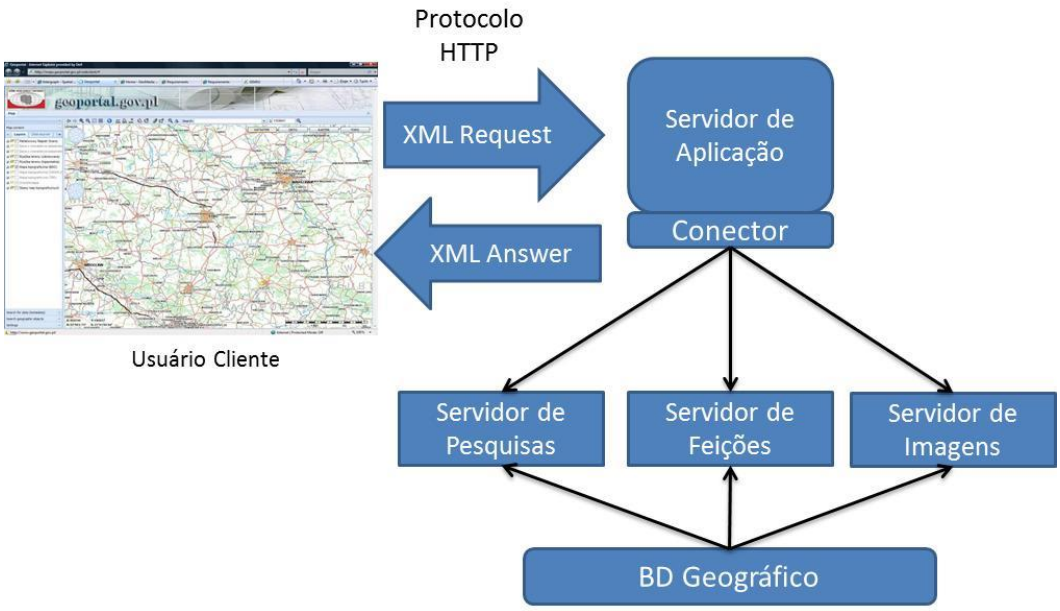


Figura 5. Arquitetura de requisição de serviço, adaptado de (OGC, 2011).

2.2.3 Geoportal

O conceito de Geoportal resume-se basicamente ao que foi previamente discutido no item “2.1 Infraestrutura de Dados Espaciais”, ou seja, como parte dos conceitos intrínsecos da construção de IDEs, pois contempla algumas características elementares como: arquitetura baseada em serviços, funcionalidades básicas para que os usuários possam visualizar, medir, etc., e também com o propósito de distribuir informações espaciais e alfanumérica de forma simples e rápida. De uma maneira geral, denominações como Portal de Informações Espaciais, Atlas Geográfico, Atlas Ambiental, etc., podem ser interpretados como parte dos conceitos de IDEs.

Portais são *websites* que podem ser entendidos como uma "porta de entrada" para a busca das mais diversas informações como dados, notícias, tutoriais, ferramentas, etc., além de serem ambientes Web que permitem que organizações e comunidades possam interagir compartilhando e trocando informações pertinentes a determinado assunto (Maguire e Longley, 2005). Especificamente para a questão espacial, o Geoportal pode ser definido como um *website* que representa um ponto de entrada para a descoberta e visualização de conteúdo geográfico (Tait, 2005).

A criação do *European Commission's INfrastructure for SPatial InfoRmation in Europe* (INSPIRE), espécie de órgão regulador europeu, incentivou a criação de diversas IDEs nas mais diferentes escalas: locais, regionais e nacionais (Masser 2010). Exemplos como (Poland Geoportal, 2013; Nature SDI Plus, 2013; Sakkopoulos et al., 2012; Geoportal de Chile, 2013) apenas ilustram a algumas implementações de portais ao redor do mundo.

Maguire e Longley (2005) chegam a ressaltar a possibilidade de dividir o Geoportal em dois grupos: Geoportais de Aplicações e Geoportais de Catálogo, na qual o primeiro contempla ferramentas de análise e consumo de informações geográficas via *Web Services* e o segundo é referente a organização e gerenciamento das informações de forma estrutural. Para maiores detalhes sobre os conceitos e arquiteturas de Geoportais recorrer a (Bernard et al., 2005; Maguire e Longley, 2005; Tait, 2005).

A Figura 6, a seguir, exemplifica o conceito e arquitetura intrínseca ao desenvolvimento de um Geoportal. Todavia é importante ressaltar que este é apenas um exemplo e não necessariamente todos os Geoportais devem seguir a mesma estrutura já que os tipos de funcionalidades, interface e serviços geoespaciais dependerão exclusivamente do propósito do portal.

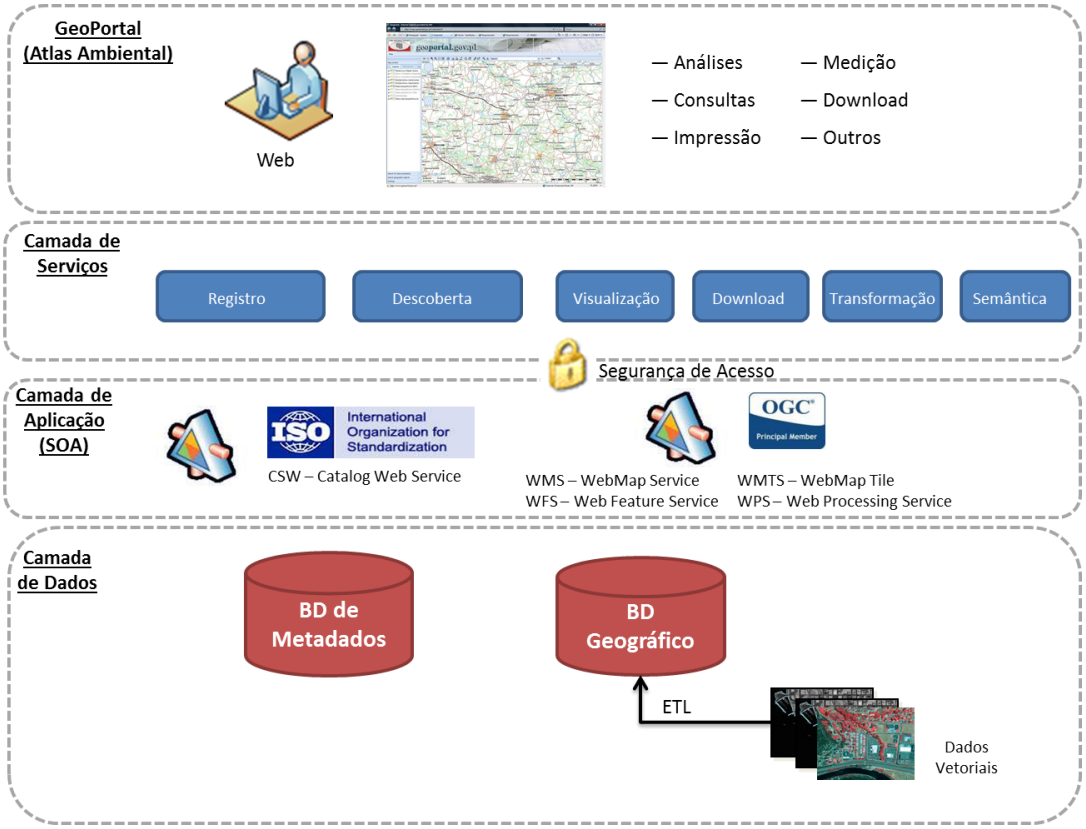


Figura 6. Arquitetura resumida de um Geoportal dentro de uma IDE.

A arquitetura apresentada é composta por dois Banco de Dados, sendo o primeiro alfanumérico (metadados) e o segundo contendo informações geográficas (vetores e imagens de satélite e/ou fotografias aéreas) provenientes de importação via *Extract, Transform and Load* (ETL) e inserções manuais. Tais informações podem ser utilizadas pelo usuário final do Geoportal por meio da interação entre a Camada de Serviços, responsável em disponibilizar as diversas funções e operações baseado em *Web services*, e a Camada de Aplicação (SOA) que recebe as requisições e interage com os bancos de dados, retornando a informação requisitada; sendo possível ainda se definir controle de acesso às informações por parte do usuário.

2.3 CONCEITO E IMPORTANCIA DE METADADOS

O crescente desenvolvimento na produção de dados espaciais aumenta a necessidade de se documentá-los visando seu uso em futuros estudos. Um dado com seu contexto (documentação e metadados) se transforma em informação,

todavia, sem estes detalhes praticamente se torna uma informação não valiosa, sem valor (GSDI, 2004).

Os Metadados representam as informações sobre o dado e qualquer descrição sobre quem, como, quando, porque, produziu um dado são exemplos destes. Segundo o Decreto Lei nº 6.666 de 27 de Novembro de 2008, instituída no âmbito do Poder Executivo Federal, metadados representa o:

> "conjunto de informações descritivas sobre os dados, incluindo as características de seu levantamento, produção, qualidade e estrutura de armazenamento, essenciais para promover a sua documentação, integração e disponibilização, bem como possibilitar sua busca e exploração".

A importância da existência de metadados é novamente abordada por GSDI (2004):

- Ajudam a organizar e manter o investimento de uma organização, como também, provê informações sobre determinado dado;
- Evita a duplicação de esforços;
- Possibilidade de localizar/ disponibilizar dados geoespaciais e associados relevantes; para a área de interesse;
- Provedores de dados são capazes de informar e promover a disponibilidade de seus dados e agregar-se a serviços existentes (relatórios de texto, imagens, mapas na web).

Quando se pensa especificamente em metadados no contexto espacial, alguns outros parâmetros são agregados, como por exemplo:

- Quem criou os dados?
- Qual o conteúdo e porque foi criado?
- Quando foi criado? (levando em consideração a época do levantamento do dado)
- Qual sua localização espacial?
- Qual é o Sistema de Coordenadas utilizado?
- Qual é sua precisão cartográfica?

Estes metadados para serem devidamente utilizados, devem seguir uma padronização de armazenamento internacional. Estes padrões auxiliam no entendimento, no acesso e na interoperabilidade por meio de *softwares* e aplicações já desenvolvidas para compreender as informações armazenadas. Alguns padrões internacionais já foram e estão sendo desenvolvidos ao longo do tempo, cada qual com propósitos e objetivos que, as vezes, são semelhantes em alguns parâmetros,

mas também distintos em outros aspectos como distintas regras de preenchimento (campos obrigatórios e opcionais), por exemplo.

O padrão *Federal Geographic Data Committee* (FGDC, 2000), desenvolvido em 1994 nos Estados Unidos, é comumente adotado no desenvolvimento de IDEs em países como Estados Unidos, Canadá e Reino Unido. Tal padrão organiza os dados em 7 (sete) diferentes seções, conforme Figura 7:

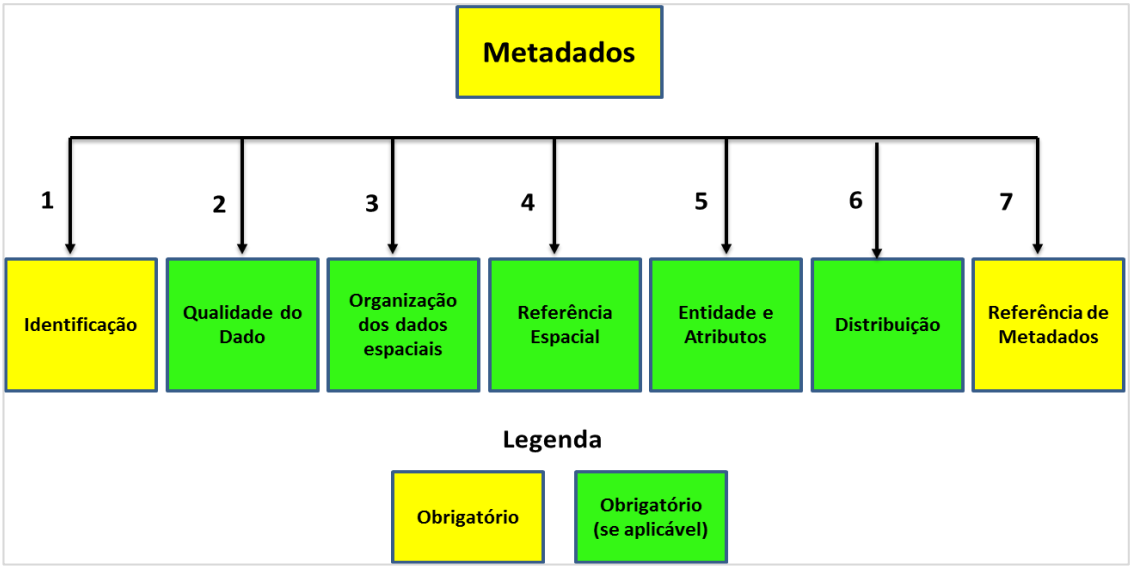


Figura 7. Composição das seções referentes ao padrão FGDC e suas obrigatoriedades, adaptado de (FGDC, 2011).

1. **Identificação**: Informações básicas sobre o conjunto de dados. Onde os dados originaram-se? Como está atualmente o dado? Para quais propostas foram criadas? Qual a área de cobertura do dado?

2. **Qualidade do Dado**: Uma avaliação da qualidade dos dados. Como está a precisão? Quais os passos foram seguidos para coletar os dados? Como foi à entrada de dados?

3. **Organização dos Dados Espaciais**: Como o dado espacial está representado? Quais objetos foram usados para representá-lo?

4. **Referência Espacial**: A descrição do sistema de referência, sistema de coordenadas e parâmetros de projeção utilizados.

5. **Entidade e Atributos**: Quais os tipos de entidade e atributos que fazem a descrição do dado?

6. **Distribuição**: De quem os dados podem ser obtidos? Qual o tamanho do conjunto de dados? Pode ser feito *download*? Há algum custo de aquisição?

7. **Referência dos Metadados**: Informação sobre os metadados. Quando foi que o registro de metadados foi criado? Quem é o responsável? Quando foi a última atualização?

Outro padrão muito popular de metadados é o ISO 19115:2003 (ISO, n2011), especificado pelo Comitê Técnico 211 (TC 211) da ISO. Utiliza de linguagem de modelagem *Unified Modeling Language* (UML*)* para representar suas seções, entidades e elementos de metadados. Segundo CONCAR (2009):

> "A norma ISO 19115:2003 combina aspectos de muitos outros padrões de metadados, visando um padrão universal para o armazenamento e distribuição de metadados geoespaciais. (...) o padrão ISO 19115 apresenta diferenças marcantes em sua implementação, através do recurso de modelagem orientada a objetos (...) é um padrão verdadeiramente internacional; faz parte de um conjunto de normas afins (suíte) concernentes ao armazenamento, troca e manuseio de informações geográficas; prevê o apoio a diferenças culturais e linguísticas, contemplando culturas, áreas de aplicação, profissões, etc., não apenas pela especificação da linguagem dos metadados".

No Brasil, a Infraestrutura Nacional de Dados Espaciais (INDE) está em pleno desenvolvido, sob a tutela da Comissão Nacional de Cartografia (CONCAR) e dentre as diversas padronizações que estão vinculadas à INDE, uma delas é o uso de metadados seguindo a norma ISO 19115:2003, com o objetivo estrutural de garantir e identificar o produtor e responsabilidades técnicas de produção de dados, padronizar terminologia, possibilitar o compartilhamento e transferência de informação, entre outros.

A CONCAR, por meio de um comitê especializado denominado Comitê de Estruturação de Metadados Geoespaciais (CEMG) criou o Perfil de Metadados Geoespaciais do Brasil (Perfil MGB), que nada mais é do que um conjunto básico e obrigatório de elementos que retratem as características de determinada informação geoespacial de uma área específica e garanta sua identificação, avaliação e utilização consistente por todos aqueles que desejarem.

As seções que fazem parte do Perfil MGB (CONCAR, 2009), segundo a norma ISO 19115 são:

- **MD_Metadata** - Informações do Conjunto de Entidades de Metadados: define metadados de um produto e estabelece hierarquia;
- **MD_Identification** - Informações de Identificação: informação básica requerida para identificar univocamente um produto;

- **MD_Constraints** - Informações de Restrições: restrições legais e de segurança no acesso e no uso dos dados;
- **DQ_DataQuality** - Informações de Qualidade dos Dados: descreve sua linhagem (fontes e processos de produção) e qualidade/ teste dos dados;
- **MD_MaintenanceInformation** - Informações de Manutenção dos Dados: descreve práticas de manutenção e atualização;
- **MD_SpatialRepresentation** - Informações de Representação Espacial: descreve mecanismo usado para representar os dados geoespaciais (matricial ou vetorial);
- **MD_ReferenceSystem** - Informações do Sistema de Referência: descreve sistema de referência espacial e temporal usado;
- **MD_ContentInformation** - Informações de Conteúdo: descreve conteúdo do(s) catálogo(s) de abrangência e de feições usado(s) para definir feições de dados geoespaciais;
- **MD_Distribution** - Informações do Distribuidor: informações do distribuidor e métodos de acesso.

Portanto, para o desenvolvimento desta pesquisa foi utilizada a norma ISO 19115:2003, especificamente o Perfil MGB, por se tratar de um padrão verdadeiramente internacional, além de contemplar conjunto de normas de armazenamento, troca e manuseio de informações geográficas e, sem dúvida, ser aderente aos desenvolvimentos que estão sendo realizados em âmbito nacional com a própria INDE.

2.3.1 Metadados na área ambiental

Como dito previamente, metadados são utilizados para documentar e disponibilizar informações sobre algum dado específico. Para se documentar determinado dado é imprescindível seguir algum tipo de padrão, metodologia, etc., mas que não ocorre na maioria das vezes.

Especificamente na área ambiental, (Fegraus et al., 2005; Millard et al., 2005) destacam algumas dificuldades encontradas:

- Não padronização de metodologia no levantamento de dados ecológicos;
- Diferentes propósitos dos estudos;
- Grande variação de extensão, detalhes e qualidades dos dados podem ser encontrados;
- Amostras de tempo e espaço distintas;
- Diversidade de formato de dados e sistemas proprietários;
- Número cada vez maior de dados está sendo levantados por cientistas, instituições, etc., e precisam ser organizadas;
- Dados armazenados informalmente em arquivos Microsoft Excel, anotações mentais ou cadernos, entre outros.

Com o objetivo de superar esta situação, algumas propostas foram estabelecidas. Dentre as diversas ações sobre o tema, vale a pena destacar a iniciativa desenvolvida pelo *National Center for Ecological Analysis* (NCEAS) e *Long Term Ecological Research Network* (LTER), que criaram a chamada Linguagem Ecológica de Metadados (EML). O principal intuito é fornecer um método para formalizar e padronizar os conceitos que são essenciais para descrever os dados ecológicos.

Com a padronização proposta, é possível que cientistas, ecólogos e o público em geral possam ter acesso à levantamentos realizados em diferentes áreas, épocas, projetos, entre outros, de uma maneira simples e rápida, através do principal portal de acesso desenvolvido para o projeto[2].

A projeto desenvolvido pela *The Georgia Coastal Ecosystems* (GCE) se propõe a integrar o modelo EML aos arquivos de metadados existentes no Banco de Dados atual. Além disso, foi criada uma interface chamada *GCE Data ToolBox* com o objetivo de facilitar a busca por metadados, utilizando critérios tanto alfanuméricos (data, tipo, etc.), como espaciais.

Todavia, tal linguagem, por mais que se proponha a se integrar com os diversos padrões estabelecidos (FGDC, ISO), e também tenha possibilidade de armazenar algumas informações espaciais, ainda carece de soluções que agreguem a questão espacial dos dados. Iniciativas realizadas pela Universidade do Colorado e Prefeitura de Seattle (Van Buren, 2002), ambos nos Estados Unidos, tentaram

[2] O portal contém mais informações das instituições que fazem parte do projeto, além de possibilitar a realização de *download* do *software* desenvolvido para a criação e manutenção dos metadados EML, chamado Morpho, e os próprios dados ambientais.

realizar uma integração entre EML e SIG, porém de maneira muito incipiente e os projetos foram descontinuados por questões de investimento econômico no projeto.

Outras iniciativas como o MarineXML, desenvolvido pela *International Oceanographic Data* (IOC) e o *Information Exchange Committee of the Internacional Oceanographic Commission* (IODE) da UNESCO, tem como principais objetivos criar uma linguagem comum que permita a troca de dados entre a comunidade marinha. Millard et al. (2005) descrevem de maneira bem interessante o desenvolvimento da linguagem denominada MarineXML, inclusive com sua integração com padrões espaciais como é o caso do GML. Tal estudo, também destaca os trabalhos desenvolvidos por outros grupos, como: *Monterrey Bay Aquarium Research Institute* (MBARI), *Marine Information Mark-up Language* (MIML) e *Ocean Biogeographic Information System* (OBIS).

Estes padrões de metadados pré-estabelecidos possuem o mesmo fundamento, ou seja, são desenvolvidos a partir de uma linguagem comum XML, o que facilita a interoperabilidade entre eles, por se tratar de uma linguagem aberta. A maioria destes padrões de metadados, sejam eles voltados à área ambiental ou não, normalmente possuem interface para preenchimento de dados e posterior armazenamento. Se necessário for exportá-lo, usa-se a linguagem XML como formato de troca entre informações de distintos padrões.

De qualquer maneira, seja qual for o modelo de metadados utilizado, é necessário que os dados estejam preenchidos e documentados, com o intuito de possibilitar a troca e consumo destas informações em estudos futuros. (Bruce e Kroon, 2006; Mount e Bricher, 2008) seguem neste caminho e ilustram, entre outros assuntos, as informações significativas para serem inseridas nos metadados dos seus projetos. Exemplificam o preenchimento de valores específicos para cada tipo de dado, informando: origem, versão, extensão, sistema de coordenadas, entre outros dos dados existentes, vinculados a área ambiental.

Seeley et al. (2008) mostram, de maneira muito interessante, as melhores práticas para documentar e possibilitar a troca de informações espaciais marinhas. Sugerem quais são os metadados essenciais para as diversas fontes ambientais existentes, ilustrando como trabalhar e organizar arquivos provenientes de formatos *Microsoft Excel* - muito utilizado por pesquisadores para armazenar os frutos dos estudos marinhos e costeiros – imagens de satélite e fotografias aéreas além de Banco de

Dados Espaciais. Entre as informações comuns que, segundo os autores, devem ser preenchidos, estão:

- Nome do Projeto e Data de construção;
- Arquivo Excel de origem;
- Organização que gerou o dado, se aplicável;
- Data da origem;
- Quando o dado foi processado;
- Quem processou o dado;
- Referências usadas para identificar as versões dos dados;
- Quando o dado foi validado;
- Quem validou o dado;
- Data da validação.

2.4 BANCO DE DADOS GEOGRÁFICOS

A evolução da tecnologia proporcionou um acesso mais rápido e fácil a dados espaciais. Este significativo aumento, por um lado, facilita o acesso à dados geoespaciais por parte do público em geral, mas por outro, influencia diretamente na dificuldade de intercâmbio destas informações produzidas, sob diferente perspectivas, àqueles que irão acessá-las, uma vez que estes dados podem estar em diferentes padrões de *software*, sem documentação de sua produção.

Câmara e Júnior (2002) evidenciavam suas preocupações com a busca por padrões no armazenamento e troca de informações espaciais entre os produtores:

> "Um dos desafios mais importantes no uso das geotecnologias é o intercâmbio de dados espaciais, impulsionado principalmente pelo alto custo de produção deste tipo de dado. A falta de modelos conceituais comuns acarreta problemas na troca de dados entre organizações utilizando Sistema de Informação Geográfica (SIGs) distintos, que incluem distorção de dados, comprometimento de qualidade da informação, perda de definições de atributos e georreferenciamento".

A busca pela interoperabilidade entre diferentes dados geoespaciais parte da adoção de um Modelo de Dados que seja aderente a padronização do levantamento de dados e sua posterior disponibilização entre diferentes organizações. No Brasil, a

crescente preocupação por padronização e acessibilidade de dados geoespaciais fez com que fosse criada a INDE. Dentre as atividades em andamento, uma delas é a definição do Perfil MGB para metadados, conforme já supracitado no item “2.3 Conceito e Importância de Metadados”.

Além disso, pode-se destacar a Especificação Técnica para a Aquisição de Dados Geoespaciais Vetoriais (ET-EDGV) que:

> “tem por objetivo padronizar e orientar todo o processo de aquisição da geometria dos vários tipos de dados geoespaciais vetoriais, presentes na Especificação Técnica para Estruturação de Dados Geoespaciais Vetoriais (ET-EDGV), da CONCAR, para qualquer que seja o insumo a ser utilizado (levantamento de campo, fotografias aéreas, imagens de sensores orbitais, etc.), visto que os processos de aquisição são similares.” (Lunardi et al., 2009).

Diante deste contexto, será explanado a seguir a respeito da modelagem conceitual de dados geoespaciais, sua estruturação física e as necessidades de transformações de formatos para a adequação ao modelo definido.

2.4.1 Modelagem Conceitual de Dados Geográficos

Um modelo de dados, seja este com características geográficas ou não, é referente ao processo de abstração na qual apenas os elementos essenciais da realidade para um determinado estudo são considerados (Lisboa Filho e Iochpe, 1999).

O processo de abstração é parte fundamental na criação de sistemas de informações e trata-se, portanto, da transposição de entidades do mundo real e suas interações, para um banco de dados informatizado, realizando a descrição dos possíveis conteúdos dos dados, além de estruturas e regras das entidades escolhidas. Quando pensado no âmbito espacial, esta abstração ainda possui algumas especificidades, referentes à localização espacial, o tempo de observação e a precisão de obtenção e representação das informações geográficas. Um resumo deste processo de abstração pode ser visualizado na Figura 8:

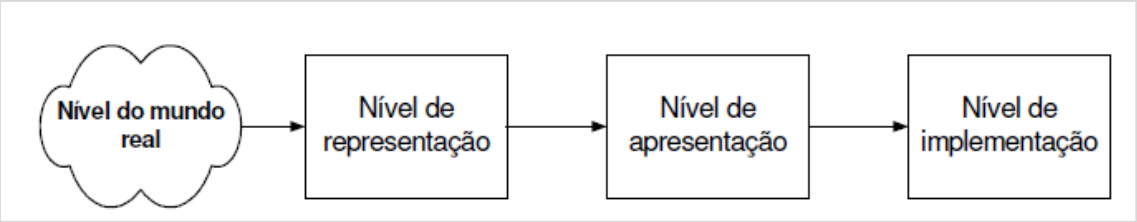


Figura 8. Processo de Abstração para a construção do Modelo de Dados (Câmara et al., 2005).

- Nível do Mundo Real: Contém os fenômenos reais a abstrair, como hidrografia, mancha urbana, vegetação, entre outros.
- Nível de Representação: Conjunto de conceitos formais com os quais as entidades geográficas podem ser modeladas, da forma como são percebidas pelo usuário.
- Nível de Apresentação: Define-se a questão simbológica de apresentação de como os fenômenos geográficos devem assumir em uma aplicação após sua modelagem.
- Nível de Implementação: Ambiente que utiliza os três itens acima descritos para se definir a forma de armazenamento e relacionamento entre a estrutura de tabelas, além de funções e métodos necessários ao sistema. Este é o ambiente desenvolvido no item 2.4.2 Implementação Física do Banco de Dados Espacial.

Uma vez implantado, o modelo de dados se assemelha a Figura 9, a seguir:

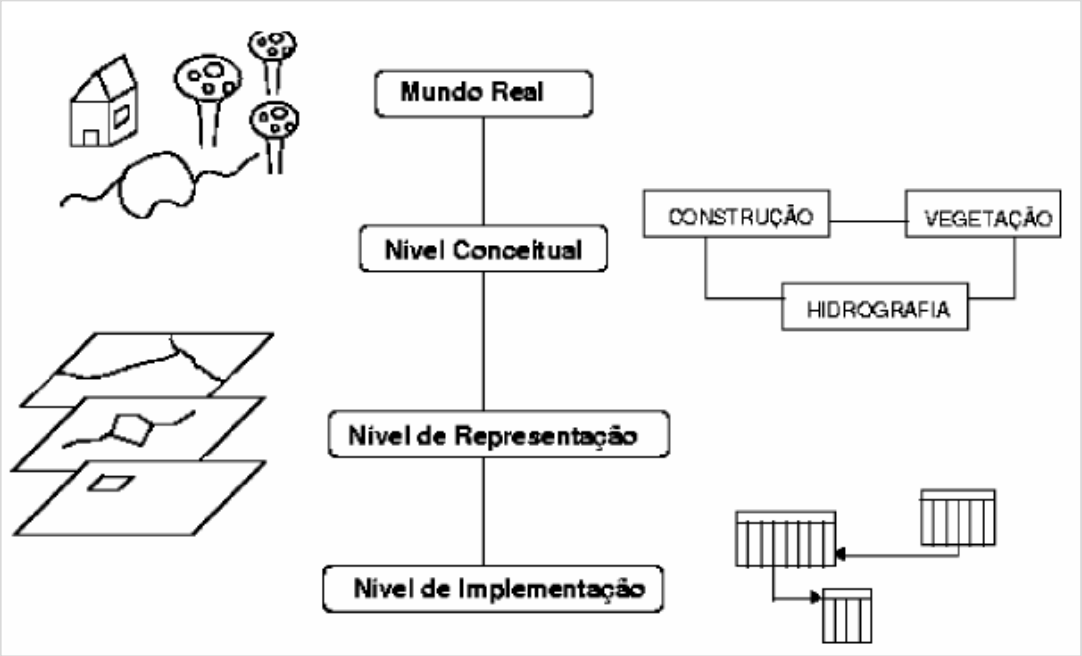


Figura 9. Níveis de Abstração de Dados Geográficos (Câmara et al., 2005).

O modelo conceitual atua diretamente nos tópicos 'Nível de Representação e Apresentação'. Estes modelos têm sido adaptados para atender as necessidades

espaciais desde 1980, uma vez que, bem definidos, auxiliam substancialmente na construção do Banco de Dados Espacial. Algumas propostas de modelagem conceitual para SIG são citadas e discutidas por (Lisboa Filho e Iochpe, 1999; Borges et al., 2001; Câmara et al., 2005) como: Entidade-Relacionamento (ER), Orientado a Objeto (OO), OMT-G (*Object Modeling Technique for Geographic Applications*), IFO (Modelo Semântico Formal de Banco de Dados), GeoFrame, GISER (*Geographic Information System Entity Relational*), entre outros.

Dentre os modelos citados, após revisão bibliográfica, notou-se que os mais utilizados são o GeoFrame e o OMT-G. Pelo fato da própria ET-ADGV ser definida sobre o modelo OMT-G, o presente estudo decidiu-se por seguir o mesmo caminho (Lunardi et al, 2009).

O OMT-G é um Modelo Orientado a Objetos estendido com a parte geográfica baseado no modelo UML e:

> "provê primitivas para modelar a geometria e a topologia dos dados geográficos, oferecendo suporte a estruturas topológicas "todo-parte", estruturas de rede, múltiplas representações de objetos e relacionamentos espaciais. Além disso, o modelo permite a especificação de atributos alfanuméricos e métodos associados para cada classe. Os principais pontos do modelo são sua expressividade gráfica e sua capacidade de codificação, uma vez que anotações textuais são substituídas pelo desenho de relacionamentos explícitos, que denotam a dinâmica da interação entre os diversos objetos espaciais e não espaciais." (Câmara et al., 2005).

Para se construir um diagrama de classes, que tem por objetivo representar um determinado modelo que seja aderente a um assunto específico, como o tema de Macrobentos, por exemplo, é necessário utilizar alguns conceitos, cada qual com sua simbologia específica, conforme Figura 10.

- Objeto: entidade do mundo real;
- Classe de objetos: representa entidades de mesma característica (atributos, operações);
- Associações: relacionamento entre classes;
- Generalização / Especialização: - hierarquia entre classes;
- Agregação: combinação de outras partes.

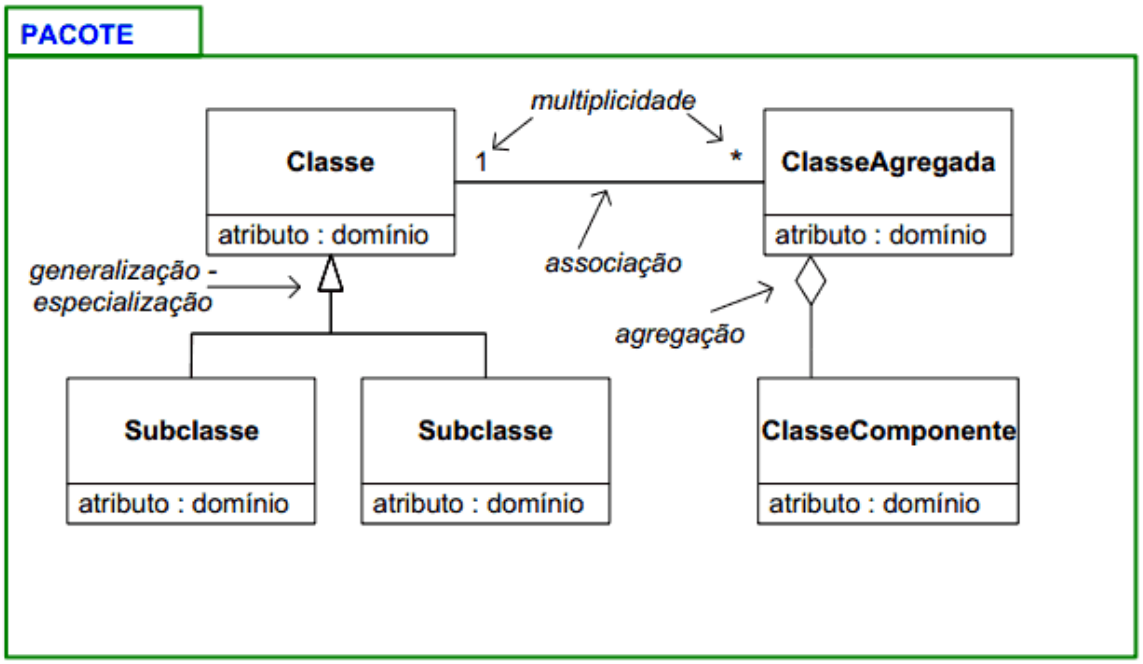


Figura 10. Notação gráfica do Diagrama de Classe UML (Lisboa Filho, 2000).

No modelo OMT-G, existem dois tipos distintos de classes de objetos: convencionais ou georreferenciadas. A primeira descreve um conjunto de objetos que possuem alguma relação com objetos espaciais, mas que não possuem propriedade geométrica e a segunda, descreve um conjunto de objetos que possuem representação espacial.

Ainda para este tema, (Borges et al., 2001; Câmara et al., 2005) sugerem uma especialização da classe georreferenciada, utilizando o conceito de herança da orientação a objetos, em classes do tipo geo-campo e geo-objeto, onde:

- Geo-campo: correspondem a objetos distribuídos continuamente pelo espaço, como topografia (curvas de nível), tipos de solo, geologia, entre outros.
- Geo-objeto: correspondem a objetos geográficos identificáveis e localizáveis que podem ser individualizados, como localização de árvores, eixos de ruas, quadras, entre outros.

O advento do modelo OMT-G faz com que cada subclasse tenha sua representação específica através de símbolos, tornando assim, uma linguagem mais intuitiva e de fácil compreensão. A principal diferença entre classe georreferenciada e convencional é que a primeira possuirá uma simbologia com um retângulo no canto superior esquerdo de cada classe, na qual é inserido o tipo de geometria a ser representada, conforme ilustra a Figura 11 a seguir:

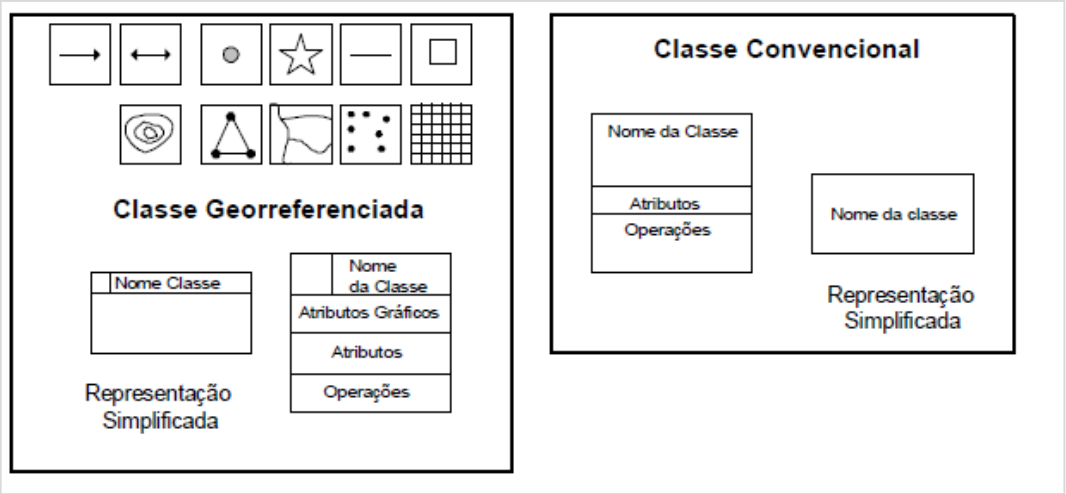


Figura 11. Diferenciação entre Classes Geográficas e Convencionais (Câmara et al., 2005).

As Figura 12 e Figura 13 a seguir, ilustram as representações simbológicas definidas para as subclasses:

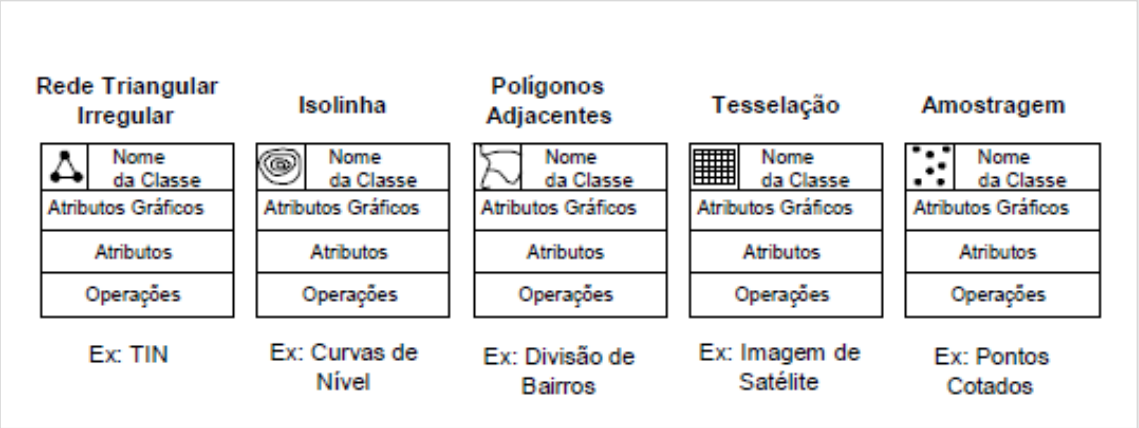


Figura 12. Exemplos de Representações de Geo-Campo (Câmara et al, 2005).

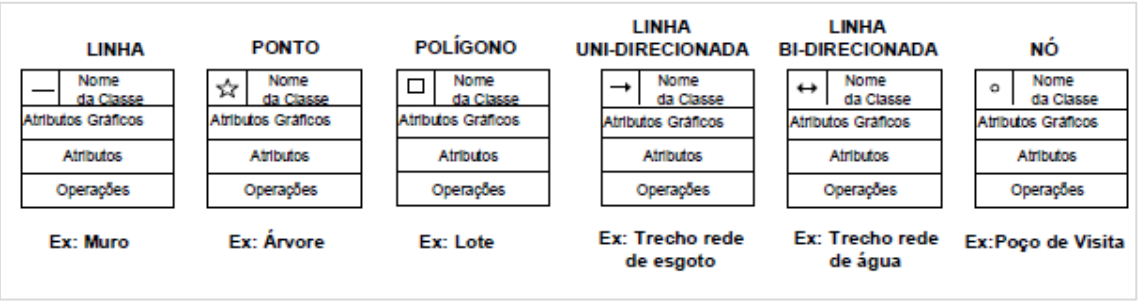


Figura 13. Exemplo de Representação de Geo-Objeto (Câmara et al., 2005).

Cada classe apresentada possui uma característica específica e, portanto, deve ser utilizada em diferentes momentos quando se deseja realizar a Modelagem de Dados. Maiores detalhes podem ser obtidos em (Borges et al., 2001; Câmara et al., 2005).

2.4.2 Implementação Física do Banco de Dados Espacial

Atualmente o termo Banco de Dados é sinônimo para o termo Sistema Gerenciador de Banco de Dados (SGBD) que contempla, além dos próprios dados, uma interface (*software*) para acesso e manipulação das informações. Isto se dá principalmente pelo fato de que a velocidade na produção e a enorme quantidade de dados, seja este espacial ou não, necessita não só um bom ambiente para armazenamento, mas que também possa ser de fácil acesso por parte do usuário final.

A implementação de um SGBD permite que seja feito o compartilhamento de dados, segurança de acesso, *backup* e recuperação à falhas, padronização de dados, controle de acessos simultâneos, regras de integridades, facilidade de manutenção, desempenho, controle de redundância, entre outros (Câmara et al., 2005; Gutting, 1994). Atualmente, quando se imagina desenvolver soluções em ambientes corporativos que tenham estas características, o SGBD é utilizado.

No que tange ao armazenamento de dados espaciais, o uso de aplicações, como citadas anteriormente, fazem parte da arquitetura de um SIG. Na prática, o armazenamento espacial é uma extensão do SGBD que permite com que geometrias sejam armazenadas.

A Figura 14 ilustra um exemplo de uma arquitetura *dual* onde o armazenamento de dados espaciais é integrado com SGBD:

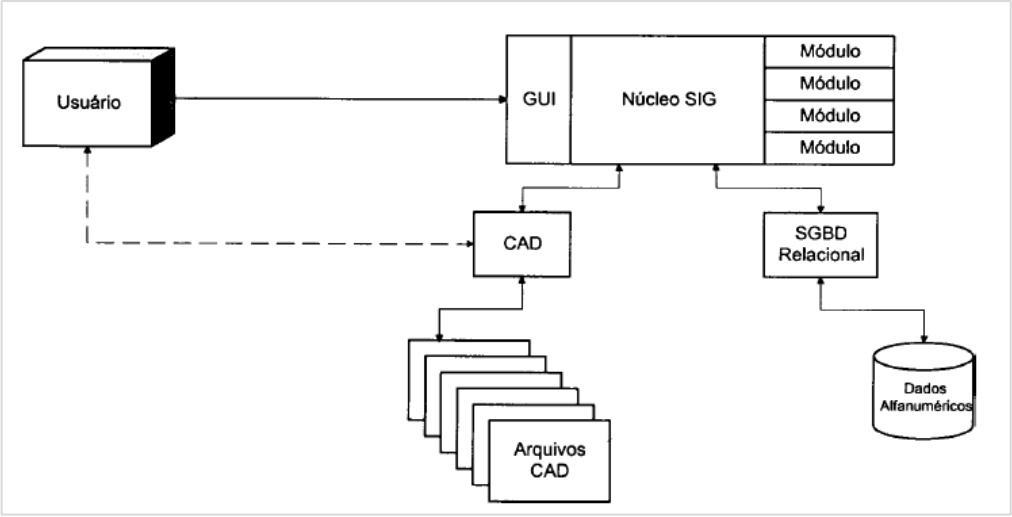


Figura 14. Armazenamento de Dados Espaciais utilizando CAD (Davis e Oliveira, 2002).

Neste exemplo, os dados espaciais são armazenados separadamente dos dados alfanuméricos e a integração é feita utilizando um identificador comum, na qual o Núcleo SIG é responsável pela junção de ambos e a visualização é realizada pelo usuário por meio uma interface gráfica *Graphical User Interface (*GUI). Esta arquitetura foi utilizada durante muito tempo por tecnologias como *Intergraph MGE*, *AutoCAD*, entre outras, e alguns problemas são intrínsecos a este modelo, como (Davis e Oliveira, 2002; Câmara et al., 2005) ilustram:

- Se um usuário acessar diretamente um arquivo CAD, este poderá inserir ou apagar um registro que não terá vínculo com o SGBD (observar linha tracejada entre usuário e arquivos CAD);
- Da mesma forma, se um administrador acessar diretamente um SGBD e inserir ou apagar um registro, este não terá vínculo com o desenho armazenado;
- A única maneira de manter a integridade das informações é por meio do Núcleo SIG;
- O conceito de Indexação e Função espaciais não é gerenciado pelo SGBD;
- Dificuldade de interoperabilidade entre diferentes dados, pois normalmente as geometrias são armazenadas de maneira proprietária, no padrão do software utilizado.

Conforme é possível observar, este modelo implementado possuía alguns problemas que necessitavam ser superados e com a evolução para uma arquitetura integrada, como define Câmara et al., (2005), foi possível realizar a junção das informações espaciais e alfanuméricas em um único SGBD.

Esta mudança possibilitou que dados espaciais, a partir deste momento, também fossem gerenciados pelo Sistema Gerenciador de Banco de Dados e, como vantagem, também usufruírem de todas as capacidades de um SGBD.

A Figura 15 ilustra exatamente a evolução para um SGBD, na qual, tanto camadas geográficas como seus respectivos atributos (informações alfanuméricas) são armazenados em uma mesma estrutura de Banco de Dados. Por meio do GUI, o usuário pode interagir com o Sistema Gerenciador de Banco de Dados realizando entrada de dados espaciais (ponto, linha, área e texto), consultas e análises espaciais, mapas temáticos, impressão e plotagem, além de diversas outras funcionalidades.

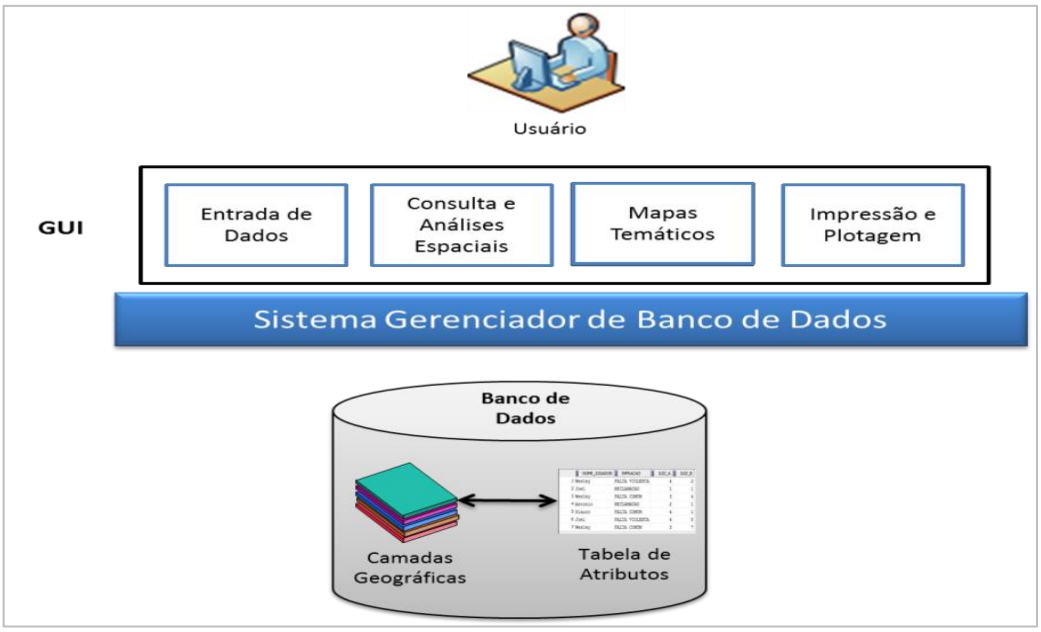


Figura 15. Sistema Gerenciador de Banco de Dados.

A arquitetura ainda pode ser mais bem aproveitada quando se tem a possibilidade de realizar operações de análises espaciais inerentes ao ambiente de SIG como cruzamento, intersecção, etc., no próprio Banco de Dados aumentando assim o desempenho. Sem a necessidade de processamento direto no ambiente cliente (*desktop*), permitirá que o usuário possua um computador com menor capacidade de processamento.

Para evitar que o mesmo problema de interoperabilidade acontecesse nos casos de SGBD espaciais, ou seja, para garantir que as ferramentas não armazenassem as geometrias de uma forma proprietária, foi definido um padrão de armazenamento e acesso aos dados, independente de qual *software* SIG ou banco de dados se esteja utilizando. O responsável pela definição e manutenção deste padrão é o próprio OGC. Este consórcio, também denominado de *OpenGIS* publica em seu Web Site a relação de *softwares* e bancos de dados testados e considerados efetivamente compatíveis (OGC, 2011).

Dentre os diversos Sistemas Gerenciadores de Bancos de Dados com Extensão Espacial estão o *Oracle Spatial*, PostGIS baseado no *Postgresql*, *Microsoft Access*, *Microsoft SQL Server*, *IBM DB2 Spatial Extender* e *MySQL*. Dentre as tecnologias citadas, pode-se destacar o Oracle, por primeiramente, ser oficialmente certificado nas especificações da OGC e também por ser uma das tecnologias mais utilizadas atualmente em ambiente corporativo.

A extensão espacial do Oracle utiliza modelo próprio de dados denominado de MDSYS, que define a forma de armazenamento, sintaxe e semânticas necessárias para suportar armazenamento de dados geométricos, além de mecanismo de indexação espacial, operadores e funções para análises espaciais (junção, intersecção, etc.), modelos topológicos e armazenamento de imagens, denominado *GeoRaster* (Murray, 2009).

O tipo de objeto para representar os dados espaciais é chamado de *SDO_GEOMETRY* e pode ser visualizado na estrutura da Tabela 1 a seguir:

Tabela 1. Estrutura de armazenamento de dados espaciais na tecnologia Oracle.

CREATE TYPE **SDO_GEOMETRY** AS OBJECT (SDO_GTYPE NUMBER, SDO_SRID NUMBER, SDO_POINT SDO_POINT_TYPE, SDO_ELEM_INFO MDSYS.SDO_ELEM_INFO_ARRAY, SDO_ORDINATES MDSYS.SDO_ORDINATE_ARRAY);	Onde cada elemento SDO armazena uma informação referente à entidade geométrica a ser inscrita, como Sistema de Coordenadas, Interpretação da Geometria, Coordenadas X, Y, Z, entre outros[3].

O mais importante aqui é destacar que ambas as informações, espaciais e alfanuméricas, estão juntas em uma mesma tabela. O exemplo abaixo ilustra o armazenamento de uma geometria que são representados pelas colunas:

- Tipo de Geometria: a geometria a ser representa é uma linha simples;
- Criação de Tabela Espacial: a criação física da tabela no Banco de Dados representado, incluindo a coluna geometria;
- Armazenamento SDO_GEOMETRY: Informações preenchidas da geometria do tipo linha no objeto.

[3] O armazenamento de dados espaciais é um assunto muito interessante, embora não seja o principal foco desta pesquisa, mas que pode ser aprofundado analisando a documentação da própria Oracle (Murray, 2009; Câmara et al., 2005).

Tabela 2. Exemplo de criação e armazenamento de dados espaciais em Oracle, adaptado de (Murray, 2009).

Tipo de Geometria	Criação de Tabela Espacial	Armazenamento SDO_GEOMETRY
	CREATE TABLE EIXO_LOGRADOURO (ID FLOAT NOT NULL, GEOMETRY **SDO_GEOMETRY**, NOME VARCHAR2(20), primary key (ID));	GEOMETRY (3002, NULL, NULL, SDO_ELEM_INFO_ARRAY (1, 2, 1), SDO_ORDINATE_ARRAY (1794727.87, 857108.118, 1799786.77, 866438.465))

Com características semelhantes, mas diferentemente do *Oracle*, é gratuito e de código fonte aberto, o *PostGIS*, extensão do *Post*gre*SQL*, desenvolvido pela Universidade da Califórnia em Berkeley, pode ser considerado também um dos bancos de dados espaciais mais utilizados em aplicações SIG.

2.4.3 Transformação e Importação de dados

Um dos processos mais críticos no que se diz a respeito a este tema, é referente a homogeneização de dados descentralizados para um SGBD. Muitas vezes os dados estão espalhados em diversos locais, com diferentes formatos e padrões, como *Microsoft Excel* (**.xls*), Arquivo Texto (**.txt*), *Adobe PDF* (**.pdf*), em diferentes sistemas ou distintos SGBD como *Microsoft SQL Server*, *Oracle*, *PostgreSQL*, entre outros.

O fato é que, com este cenário, é praticamente impossível desenvolver algum tipo de solução que possa acessar diretamente estes dados e estar apta para fornecer ao usuário final algum tipo de informação consistente e confiável. Diante disto, um conceito começou a ser desenvolvido, com o objetivo realizar a leitura/extração, transformação e posterior carga de informações em bancos de dados de destino; este conceito é denominado ETL. Como destaca Vassiliadis et al. (2002), as ferramentas de ETL tem como objetivo homogeneizar dados advindos de diversas fontes, limpá-los e carregá-los em um banco de dados de interesse do processo.

ETL esteve sempre muito associado ao conceito *Data Warehouse* (DW), pois tais tecnologias sempre encontraram grandes dificuldades de conversão de dados. *DWs*

são sistemas complexos que servem como um repositório de dados para uma organização, estando inclusas ferramentas para gerenciar e recuperar metadados, ferramentas de integração de dados e, finalmente, ferramentas de *Business Intelligence* (BI) para a realização de operações analíticas e tomadas de decisão (Simitsis et al., 2010).

Como supracitado, o *DW* é uma ferramenta integradora para a tomada de decisão, entretanto, pensando em um ambiente corporativo, tal solução precisa ter acesso à todas as bases de dados da corporação, mesmo que estes estejam em diversos formatos. O ETL, neste sentido, é responsável por extrair, transformar e carregar estes dados heterogêneos em *DWs*.

As etapas comuns de um ETL estão exemplificadas na Figura 16:

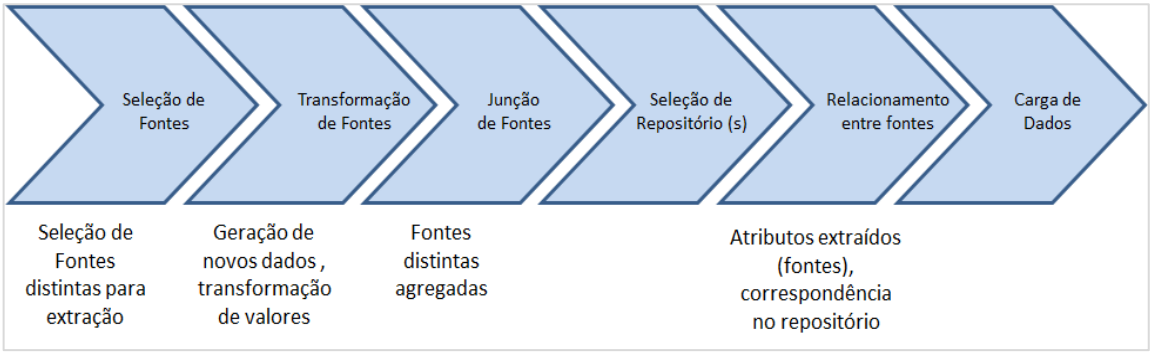


Figura 16. Fluxo de um ETL, adaptado de (Trujillo e Luajn-Mora, 2003).

No contexto espacial ainda é mais complexo este cenário, pois é possível se deparar com diferentes Bases de Dados Espaciais contemplando diversos formatos de dados, sistemas de coordenadas, atributos associados e tipos de geometria.

Há no mercado diversas ferramentas de ETL, como demonstra Ferreira et al. (2010) e algumas mais específicas para tratar a questão espacial. Uma das principais é a da empresa *Safe Software*, denominada *FME*, (FME, 2013).

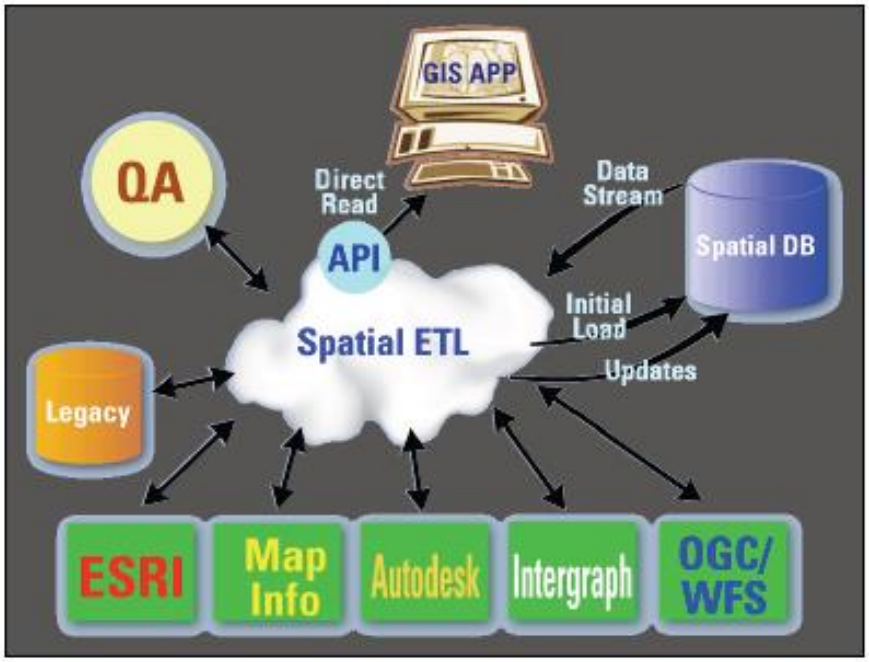


Figura 17. Aplicação de ETL para dados espaciais (Farley, 2009).

A Figura 17 ilustra, por meio de uma nuvem, a ferramenta de ETL Espacial responsável em ler diversos arquivos de dados espaciais, sejam eles de formatos proprietários (*Intergraph, Esri, MapInfo, AutoDesk*), serviços espaciais (WFS), bancos de dados legados e transformados para serem inseridos em Bancos de Dados Espaciais Corporativos e Aplicações Terceiras. O fluxo contrário também é válido, na qual, a partir de Bancos de Dados Corporativos, realiza-se a exportação de dados para formatos proprietários.

Algumas outras aplicações também podem ser citadas, como por exemplo, o ETL Espacial desenvolvido por Pestana e Silva (2003) com o objetivo de espacializar todos os dados provenientes de Bancos de Dados Alfanuméricos. Neste caso, o ETL serve como um processo que define se o dado vai ser inserido diretamente no Banco de Dados GeoDW como vetor, quando o dado de origem for espacial, ou se será necessário geocodificá-lo[4] por pares de coordenadas ou endereço. Por fim, os dados são consumidos por aplicações diversas como *softwares* clientes (*desktop*), *Web Browser* ou ferramentas *On Line Analytical Processing* (OLAP) (Fidalgo, 2005), como ilustra a Figura 18 a seguir:

[4] Geocodificar significa obter uma coordenada geográfica, normalmente do tipo ponto, a partir de valores de coordenadas (x/y ou lat/long) ou endereços que sejam georreferenciados.

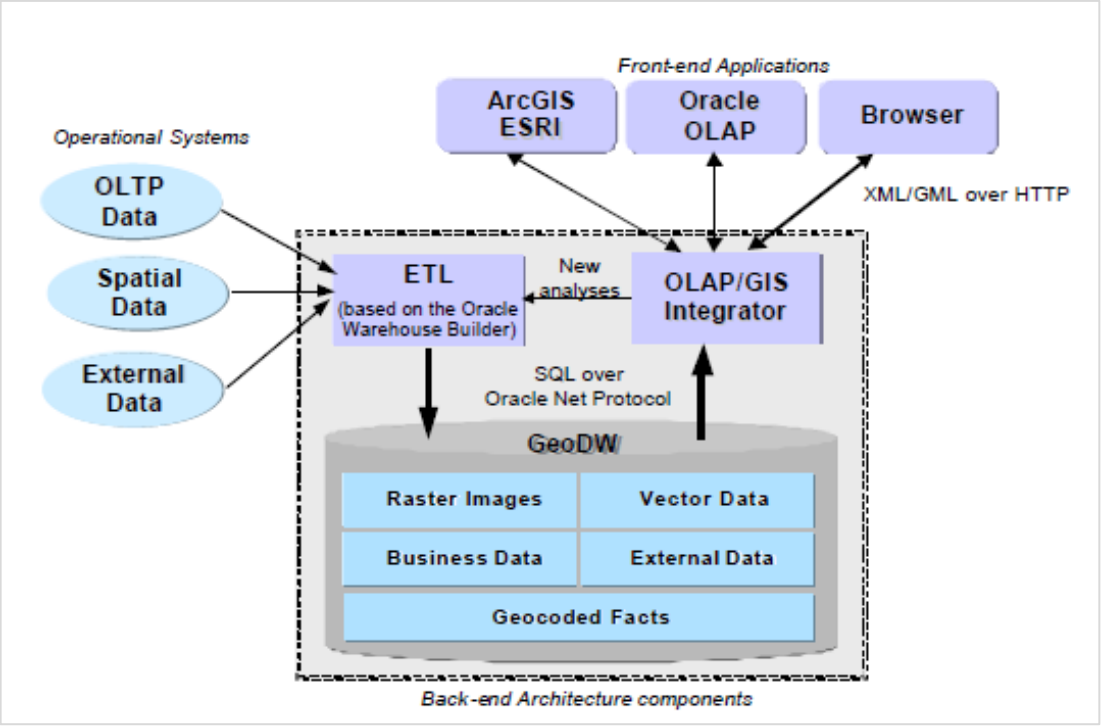


Figura 18. Exemplo de utilização de ETL Espaciais (Pestana e Silva, 2003).

As arquiteturas supracitadas proporcionam maior flexibilidade no desenvolvimento de um sistema e a integração com soluções terceiras. Exatamente por isto, ferramentas como estas estão sendo inseridas cada vez mais em arquiteturas de IDE, uma vez que os provedores de informações poderão continuar a trabalhar nos formatos que escolheram e a ferramenta de ETL será responsável por padronizá-las (transformação de sistema de coordenadas, adequação de atributos, validação geométrica de dados, etc.) para serem integradas por uma IDE, conforme mostra a Figura 19:

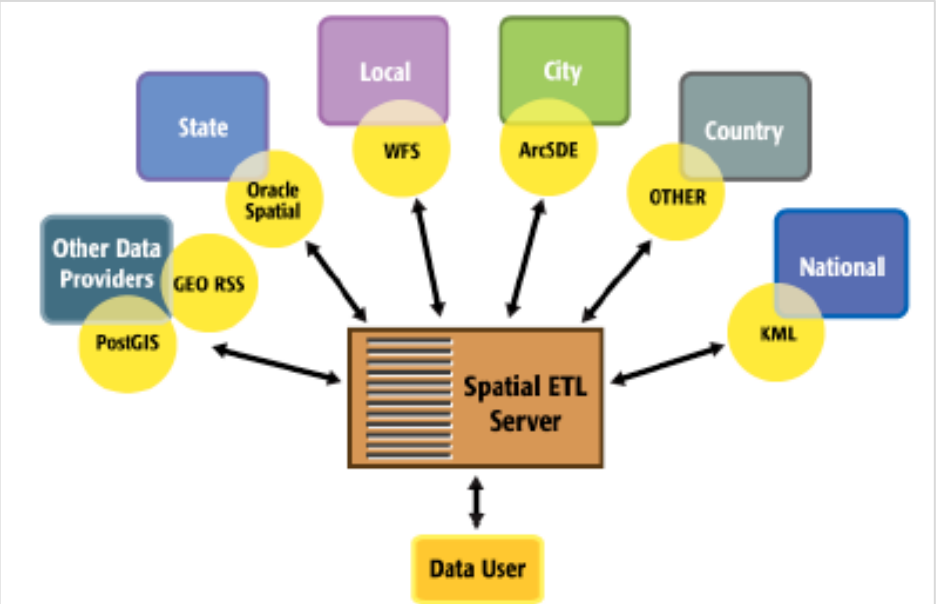


Figura 19. ETL como padronizador de dados para IDE (Hintz, 2009).

Por fim, algumas considerações sobre uma ferramenta ETL são salientadas por Ferreira et al. (2010), no momento da escolha para atendimento de uma demanda específica:

- Facilidade de Uso: deve ser facilmente usada pelo utilizador;
- Apoio funcional: apoio na extração de dados de múltiplas fontes, na limpeza e na transformação de dados, agregação, reorganização e operações de carga;
- Reutilização: deve possuir uma reutilização da lógica de transformação para que o usuário não necessite reescrever a mesma lógica de transformação outra vez;
- Programação: Deve apoiar o agendamento de tarefas ETL aproveitando melhor o tempo, não necessitando de intervenção humana para completar uma tarefa particular. Deve também ter suporte para programação em linha de comandos usando programação externa;
- Paralelismo: apoio às operações de vários segmentos e execução de código paralelo;
- Suportar a carga e o equilíbrio entre servidores e a capacidade de lidar com grandes volumes de dados.

2.5 REVISÃO DE APLICAÇÕES ESPACIAIS PARA A ÁREA AMBIENTAL

2.5.1 Gerenciamento de Áreas Costeiras e Marinhas

(Strain *et al*., 2004; Rajabifard et al., 2008; Bartlett e Smith, 2005) ilustram dificuldades encontradas no gerenciamento de zonas costeiras:

- São governadas por complexas legislações e instituições;
- Agências governamentais em âmbito regional ou nacional são responsáveis por diferentes aspectos da mesma área física e uso da zona costeira, como por exemplo: meio ambiente, pesca, agricultura, transporte, planejamento urbano e cadastro;
- As mudanças do ambiente marinho são muito mais rápidas em relação ao ambiente terrestre;
- Há menos levantamento de dados e estudos que em ambiente terrestre.

Além dos aspectos apresentados há dificuldade na coleta, gerenciamento e disponibilização de dados nesses ambientes. (Strain *et al*., 2004; Rajabifard et al., 2008; Bartlett e Smith, 2005) explicitam, os seguintes fatores a destacar:

- Os dados são dinâmicos e multidimensionais dificultando a delimitação da área de coleta e a atualização dos dados;
- Os dados são coletados para um projeto específico e raramente são compartilhados com outras organizações;
- Não há padronização no levantamento de dados, dificultando assim, a interoperabilidade (troca de informações sem necessidade de transformação);
- As diferenças entre ambientes marinhos e terrestres, além dos métodos de coleta e tecnologias utilizadas para o levantamento de dados, representam grande desafio para a interoperabilidade entre ambos;
- Os estudos sobre políticas para dados espaciais marinhos ainda são incipientes.

Rajabifard *et al*. (2008), após discutir os principais problemas, propôs, de maneira teórica, a criação de uma IDE que englobasse dados provenientes do ambiente marinho, costeiro e terrestre por meio da identificação de quais dados vinculados ao ambiente marinho os usuários necessitavam, verificação e comparação das IDEs desenvolvidas para ambos ambientes, estudos sobre interoperabilidade e troca de dados, desenvolvimento de metodologia para uma IDE comum, teste em uma área piloto e finalmente a implementação.

Especificamente na questão de gerenciamento costeiro e marinho, as diversas atividades internacionais referentes ao assunto ajudaram e subsidiaram o desenvolvimento de soluções para suportar e compartilhar a dimensão espacial de dados marinhos e costeiros (Strain, 2008).

Como também lembra US CoOP (2002):

> "há a necessidade por uma melhor e mais compreensiva maneira para agregar trabalhos provenientes de diferentes disciplinas com o objetivo de oferecer um entendimento mais integrado do meio ambiente marinho e processos que o controlam (...) há a necessidade de se padronizar práticas e procedimentos" (tradução do autor).

Diante deste contexto, algumas IDEs estão sendo desenvolvidas com o enfoque principal no gerenciamento do ambiente costeiro e marinho. Como é possível notar, os casos explicitados a seguir, normalmente fazem parte de uma IDE em maior escala, seja esta regional ou nacional, que demonstra a preocupação de não serem desenvolvidos de maneira isolada e seguem as melhores práticas internacionais de padronização, troca de dados, etc.:

- *Coastal Spatial Data Infrastructure* (CSDI): é gerenciado pelo Centro de Serviços Costeiros ou *The National Oceanic and Atmospheric Administration* (NOAA), sendo desenvolvido como parte do IDE Americano (US NSDI), que permitiu o acesso e consumo de dados geoespaciais costeiros e marinhos (Noaa, 2011);
- *Marine Geospatial Data Infrastructure* (MGDI), iniciado no ano de 1999, como parte integrante do *Canadian Geospatial Data Initiative* (CGDI), agora denominado GeoConnections (Csdi, 2011);
- *Marine Environmental Data and Information Network* (MEDIN) desenvolvido na Inglaterra, o projeto iniciou no ano de 2003 com o nome de *Marine Data and Information Partnership* (MDIP) e em 2008, mudou-se o nome do Órgão que gerenciava tal projeto para *Marine Science Coordination Committee*

*(*MSCC) e consequentemente a solução sofreu uma restruturação para o nome atual Medin (2011);

- *Marine Irish Data Atlas* (MIDA): Segundo Strain (2008), a Irlanda possui uma significativa quantidade de dados relacionados com o ambiente costeiro e marinho, através da integração de 18 diferentes agências em 6 departamentos distintos do governo. O projeto MIDA tem como intuito melhorar o acesso aos dados espaciais por meio do desenvolvimento de um Atlas Marinho (Mida, 2011);
- Há diversas iniciativas na Austrália, sendo a *Australian Marine Boundary Information System* (AMBIS), desenvolvido pelo *Geoscience Australia* (GA) no ano de 2001 (AMBIS, 2011) e o *Australian Ocean Data Centre Joint Facility* (AODCJF), inaugurado em 2005 (Aodcjf, 2011);
- *African Marine Atlas* (AMA): projeto desenvolvido pela O*cean Data and Information Network for Africa* (Odinafrica), responsável por organizar mais de 40 institutos marinhos em 25 países na África, oficialmente lançado em 23 de Fevereiro de 2007, contendo grande quantidade de geoinformação marinha referente ao litoral e oceano africanos (Odinafrica, 2011);
- *International Coastal Atlas Network* (ICAN): é uma iniciativa formada por um grupo informal, que contém mais de 35 organizações, com representações na Europa, América e África. Um dos principais objetivos do grupo é compartilhar experiências e soluções comuns para auxiliar no desenvolvimento de Geoportais, levando em consideração padrões de armazenamento de metadados e troca de informações espaciais (Ican, 2011).

2.6 DEFINIÇÃO DA ÁREA DE ESTUDO

2.6.1 Região Metropolitana da Baixada Santista

A Região Metropolitana da Baixada Santista é formada pelo conjunto de 9 (nove) municípios, sendo eles: Bertioga, Cubatão, Guarujá, Itanhaém, Peruíbe, Praia Grande, Mongaguá, Santos e São Vicente, abrangendo uma área aproximada de

2,373 km^2 e com população em torno de 1.799.086 de pessoas (Emplasa, 2011), conforme mostra Figura 20:

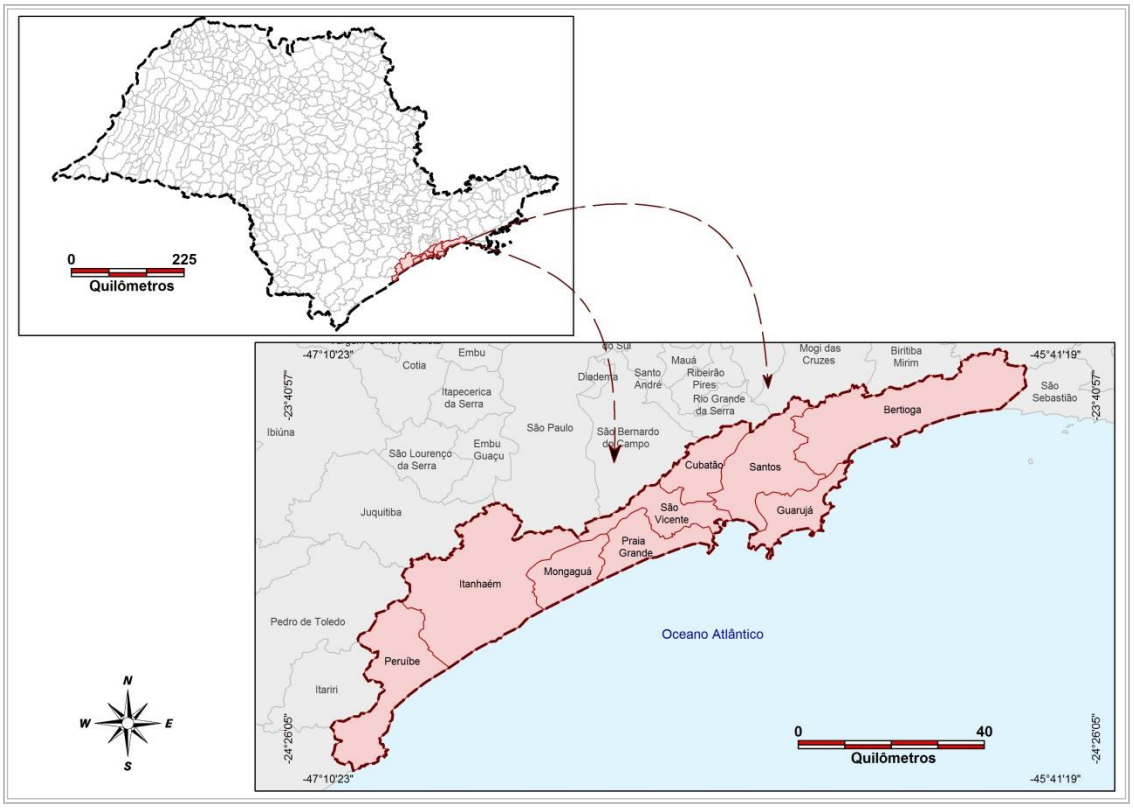


Figura 20. Contexto Espacial da Região Metropolitana da Baixada Santista no Estado de São Paulo composta pelos municípios: Bertioga, Cubatão, Guarujá, Itanhaém, Mongaguá, Peruíbe, Praia Grande, Santos e São Vicente.

Segundo a Empresa Paulista de Planejamento Metropolitano (Emplasa) - órgão pertencente ao Governo do Estado de São Paulo, responsável pelo planejamento urbano das regiões metropolitanas do Estado que inclui São Paulo, Campinas e Baixada Santista - antes mesmo de se definir uma Região Metropolitana é necessário analisar a metrópole (cidade-principal) por meio de duas escalas principais (Emplasa, 1983):

- Nível Regional: Analisar sua área de influência e o poder que esta metrópole possui no sentido de direcionar a organização espacial no nível regional.
- Nível Local: Analisar a articulação da metrópole com a região através da Rede Urbana pertencente à área, como também, no seu poder decisório em relação aos aspectos político-espaciais.

Uma vez definido a importância desta Metrópole para a Região Metropolitana, é possível observar, no Nível Local, um movimento denominado de Processo de Metropolização. Devido a sua evidente importância local, a Metrópole acaba por

aglutinar rapidamente uma grande quantidade de pessoas, indústrias, diversas atividades, entre outros.

O resultado deste processo condiz na distribuição destas atividades para os municípios que cercam a Metrópole, desenvolvendo assim, a mancha urbana metropolitana.

> “(...) Deste processo de Metropolização, gerado a partir da “cidade-mãe” (a metrópole), resulta a estrutura de um novo espaço, composto por um conjunto de cidades (e seus correspondentes municípios) que se agrupam em razão da integração física (através da conurbação) e/ou funcional (estabelecida pela inter-relação socioeconômica) estabelecida entre ela e a metrópole” (Emplasa, 1983).

A partir deste Processo de Metropolização, uma Região Metropolitana pode ser definida como o agrupamento de municípios integrados estruturalmente (relacionado à infraestrutura, recursos de recursos naturais, etc.) e funcionalmente (complementações e suplementações de atividades que ocorrem entre os municípios), liderados por uma cidade líder. Além do que, também é definida por meio do âmbito jurídico, na qual, é necessário que para a existência de uma Região Metropolitana existam mais de um município, serviços comuns entre eles e formem uma comunidade socioeconômicas.

No aspecto econômico, a RMBS é responsável por 3,7% da riqueza estadual, grande parte proveniente do potencial turístico, industrialização (Polo Industrial de Cubatão e Porto de Santos), além de enorme disponibilidade de recursos pesqueiros provenientes dos manguezais existentes na área.

A proximidade da Região Metropolitana de São Paulo e o investimento em infraestrutura na Baixa Santista ao longo dos anos, fez com que a região sofresse continuamente por fatores antrópicos ocasionando a diminuição de áreas de Mata Atlântica e de manguezais, aumento da poluição de rios e córregos causados por efluentes industriais e/ou residenciais, entre outros.

A primeira grande infraestrutura construída na região foi a linha férrea denominada Santos-Jundiaí. A princípio, seria apenas responsável pelo escoamento do café proveniente do interior paulista ao exterior via Porto de Santos, inaugurado no ano de 1892.

A partir da década de 40, mais precisamente 1947, foi finalizada a construção da Via Anchieta, que se tornaria uma das principais responsáveis pelo aumento do fluxo turístico e de serviços à região. Aliado a infraestrutura ferroviária pré-existente, foi possível que se iniciasse o processo de ampliação do Porto de Santos, permitindo

que diversos tipos de navios trazendo carvão ou minérios pudessem atracar no interior do estuário santista.

Esta mesma infraestrutura possibilitou a implantação de diversas indústrias de base (siderurgia, petroquímica, fertilizantes) em meio a uma ampla rede de canais estuarinos e extensos manguezais. O polo industrial instalado no Complexo de Cubatão é responsável em lançar grande quantidade de elementos e substâncias químicas no sistema estuarino. Atualmente contém cerca de 1.100 indústrias que descarregam cerca de 100.000 kg/mês, tais como metais e derivados de petróleo, nas águas adjacentes, sendo que, as três principais indústrias, COSIPA, Ultrafértil e Refinaria de Presidente Bernardes (RPBC – Petrobras), são responsáveis em lançar 90% da carga tóxica no ambiente (Ferreira, 2008).

Esta atividade industrial, de alto potencial poluidor, fez dos estuários de Santos e São Vicente grandes receptores de resíduos tóxicos e efluentes líquidos contaminados. Os poluentes industriais, juntamente com os resíduos e esgotos do Porto de Santos e das cidades da região, provocaram um grave quadro de degradação ambiental, com significativos reflexos na área social e de saúde pública (CETESB, 2001).

Diversos estudos já foram realizados na Baixada Santista com o objetivo de ilustrar a qualidade dos compartimentos água e sedimento influenciados diretamente por estes resíduos tóxicos e esgoto: (CETESB, 2001; Heitor, 2002; Bicego et al., 2006; Martins et al., 2007; Ferreira, 2008), entre outros.

Além disso, outros estudos utilizando conceitos de Sistema de Informações Geográficas, mais especificamente Sensoriamento Remoto, demonstram a influência antrópica e a consequente diminuição das áreas de manguezais na Baixada Santista (Menghini et al., 2007), além de derramamentos de petróleo (Santos et al., 2007).

Cunha-Lignon et al. (2009), por meio da análise temporal de fotografias aéreas e imagens de satélites, ilustraram os impactos ambientes ocorridos em manguezais no Estado de São Paulo, incluindo a Baixada Santista: Canal de Bertioga e Estuário de Santos. Foi notado que em ambas regiões a diminuição das áreas de mangue estão relacionadas com o aumento da infraestrutura e a consequente expansão da mancha urbana.

2.6.2 Contextualização e importância de Macrobentos

Estuários são áreas intensamente utilizadas pelo homem, concentrando grandes centros urbanos, intensa atividade industrial, portuária e comercial (Nybakken, 2001). Essas condições implicam em potenciais impactos ambientais (Kennish, 1997; Rivero et al., 2005).

Biologicamente, os estuários estão entre os ecossistemas mais produtivos do oceano. A Baixada Santista, região que compreende a área estudada na presente pesquisa, engloba o sistema estuarino de Santos e São Vicente. Localiza-se na faixa litorânea do Estado de São Paulo, apresenta complexa estrutura de ocupação da zona costeira que, frequentemente implica em conflito com o equilíbrio dos ecossistemas. Com relação à contaminação química industrial e orgânica, o estuário de Santos é considerado uma área crítica (Cetesb 2001, 2004, 2007 2008 e 2009; Sartor et al., 2000).

Considerando-se a importância ambiental do sistema estuarino de Santos e São Vicente no seu sentido mais amplo: ecológico, social e econômico são muito importantes às contribuições de estudos que auxiliem a um correto direcionamento quanto às ações de expansão concomitante ao controle da poluição e manutenção da capacidade de produtividade do ecossistema.

Comunidades bentônicas têm sido amplamente utilizadas nos programas de monitoramento ambiental. A maioria dos organismos é séssil ou sedentária e podem indicar graus de poluição por alterar a composição da comunidade faunística. Além disso, incluem espécies com diferentes tolerâncias ao agente poluidor como metais pesados, organoclorados, esgoto, entre outros (Warwick, 1986; Bilyard, 1987; Ahn et al., 1995).

A diversidade, abundância e dominância de espécies estão diretamente associadas às condições ambientais, tais como salinidade, temperatura, oxigênio dissolvido, variações de marés, ação de ondas, topografia, granulometria, entre outros (Snelgrove e Butman, 1994; McLachlan, 1996). A interação desses fatores é responsável pela estabilidade do sistema e alterações provocadas direta ou indiretamente pela ação do homem, podem desequilibrá-lo. Se as alterações forem frequentes, o desequilíbrio pode eliminar espécies sensíveis, substituindo-as por

espécies oportunistas, provocando modificações nos níveis tróficos superiores da comunidade faunística Bilyard (1987).

Anelídeos poliquetas são o grupo dominante na região estudada e têm sido amplamente utilizados em programas de monitoramento ambiental, pois a diversidade e abundância relacionadas com os padrões de distribuição podem revelar se um determinado ambiente encontra-se ou não impactado (Pearson e Rosenberg, 1978; Choi e Koh, 1984; Pocklington &, Wells, 1992; Rizzo e Amaral, 2000; 2001 a, b). Dependendo dos fatores hidrodinâmicos e da heterogeneidade do substrato, os poliquetas podem dominar tanto em riqueza quanto em número de indivíduos, sendo representados por mais gêneros do que qualquer outro grupo animal (Brown e McLachlan, 1990; Amaral et al., 1995). Desempenham importante papel na cadeia trófica marinha, servindo como base alimentar para inúmeros organismos. Para algumas espécies de peixes e crustáceos, os poliquetas representam uma proporção superior a 80% do alimento ingerido (Amaral e Migotto, 1980).

3 MATERIAIS E MÉTODOS

3.1 INTRODUÇÃO

O capítulo tem como principal objetivo descrever primeiramente a Metodologia utilizada para agregar as diversas fontes de dados existentes para o tema Macrobentos e como estruturá-las de forma a garantir a aderência aos conceitos e compartilhamento ao Atlas Ambiental. Em segundo lugar, uma abordagem mais detalhada de cada uma das etapas da metodologia também é discutida neste capítulo.

3.2 METODOLOGIA

A Figura 21 ilustra a metodologia empregada no desenvolvimento deste estudo. A construção da Modelagem Conceitual de Dados está intrinsicamente ligada ao conhecimento obtido nas diversas Referências Bibliográficas existentes sobre o tema e ao Levantamento de Necessidades dos usuários do sistema que, em paralelo, subsidiam e refinam o modelo empregado. As referências bibliográficas que foram utilizadas nesta pesquisa são: (Tommasi, 1979; Heitor, 2002; Embraport, 2003; Codesp, 2008; Ferreira, 2008; Codesp, 2010).

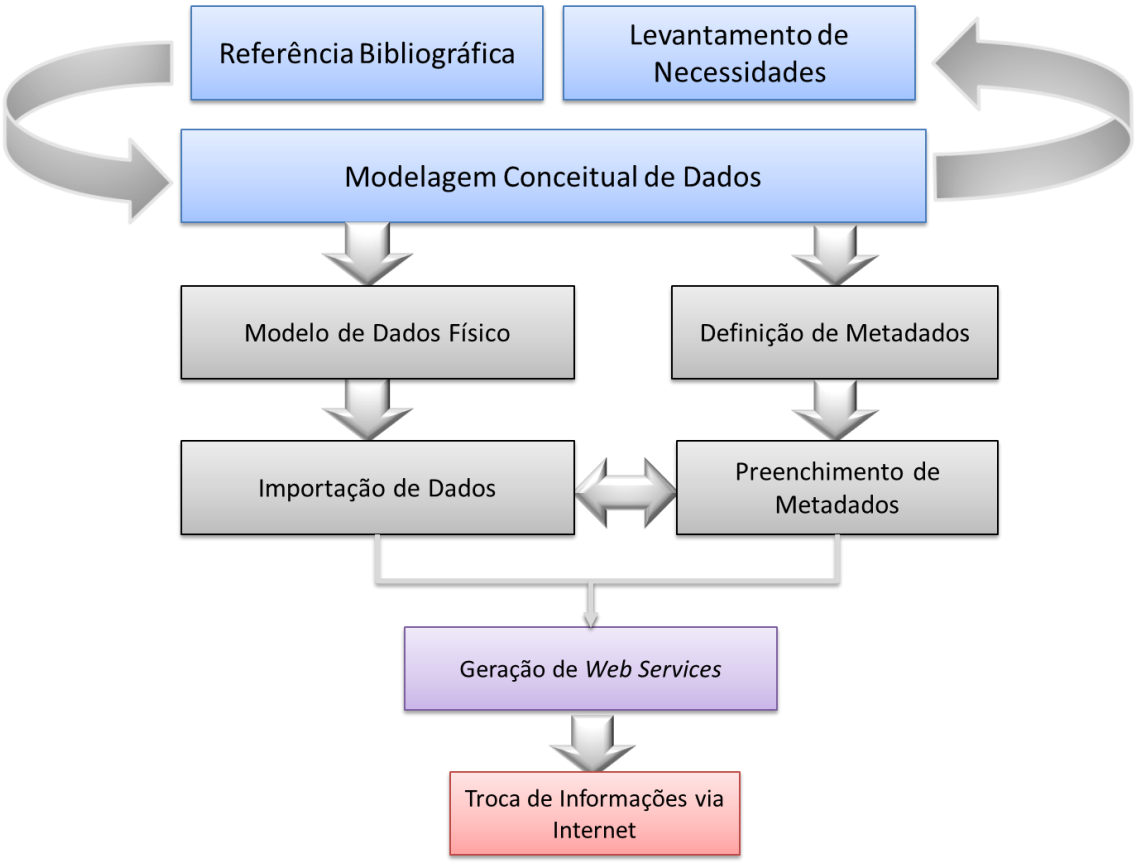


Figura 21. Resumo da Metodologia.

Uma vez definido o Modelo Conceitual adequado às fontes bibliográficas existentes, as necessidades dos usuários e ao tema em questão, foi construído um Modelo de Dados Físico, normalmente representado por um Sistema Gerenciador de Banco de Dados tais como *Oracle*, *Microsoft SQL Server*, *Postgresql*, entre outros, para suportar o processo de Importação de Dados.

Por outro lado, as mesmas referências bibliográficas e necessidades subsidiam a Definição de Metadados, no sentido de avaliar quais campos devem ser ou não obrigatórios, quais informações são imprescindíveis compartilhar para a comunidade, etc.

O Preenchimento de Metadados é essencial para que se possa documentar corretamente todos os passos realizados definindo fontes e adequações nos dados, com o objetivo de facilitar a compreensão e o correto consumo destas informações em uma IDE. Os metadados ajudam a organizar e manter o investimento de uma organização, evita duplicidade de esforços além de facilitar e prover informações para acesso (GSDI, 2004).

Além disso, a correta parametrização de metadados também permite que tais dados, hoje armazenados em bibliotecas públicas e documentos analógicos possam ser compartilhados, via Geração de *Web Services,* de maneira simples e direta como os serviços WMS, WFS, CSW, que é exatamente o que representa a Troca de Informações via Internet garantindo a interoperabilidade entre estas informações e qualquer Geoportal que siga os padrões OGC.
Os itens a seguir tem o propósito de abordar com mais detalhes cada etapa da metodologia.

3.3 CONSTRUÇÃO DO MODELO DE DADOS

Nesta etapa serão descritos os passos realizados para a construção do Banco de Dados Espacial referente ao tema de Macrobentos com o propósito específico de incluí-lo em um Atlas Web Ambiental para a RMBS utilizando os conceitos de IDE como referência no desenvolvimento. Portanto, será discutida a construção da Modelagem Conceitual sobre o tema, sua implementação física e por fim, a etapa desenvolvida para se padronizar/ importar as fontes de dados distintas ao modelo criado.

3.3.1 Modelo Conceitual de Dados

Conforme destacado no tópico *2.4.1 Modelagem Conceitual de Dados Geográficos* do presente trabalho, o modelo conceitual é referente ao processo de abstração da realidade, na qual se levantam os fenômenos intrínsecos ao tema em questão, entende-se, organiza-os de maneira lógica com o objetivo de explicá-los e disponibilizá-los como referência à outros estudos.
A Figura 22 a seguir, destaca o resultado da construção do modelo teórico utilizando o modelo de dados geográfico OMT-G (Borges et al*.,* 2001). Para a construção do modelo foi utilizado o software *Microsoft Visio* 2010, com uma extensão específica aderente ao modelo OMT-G:

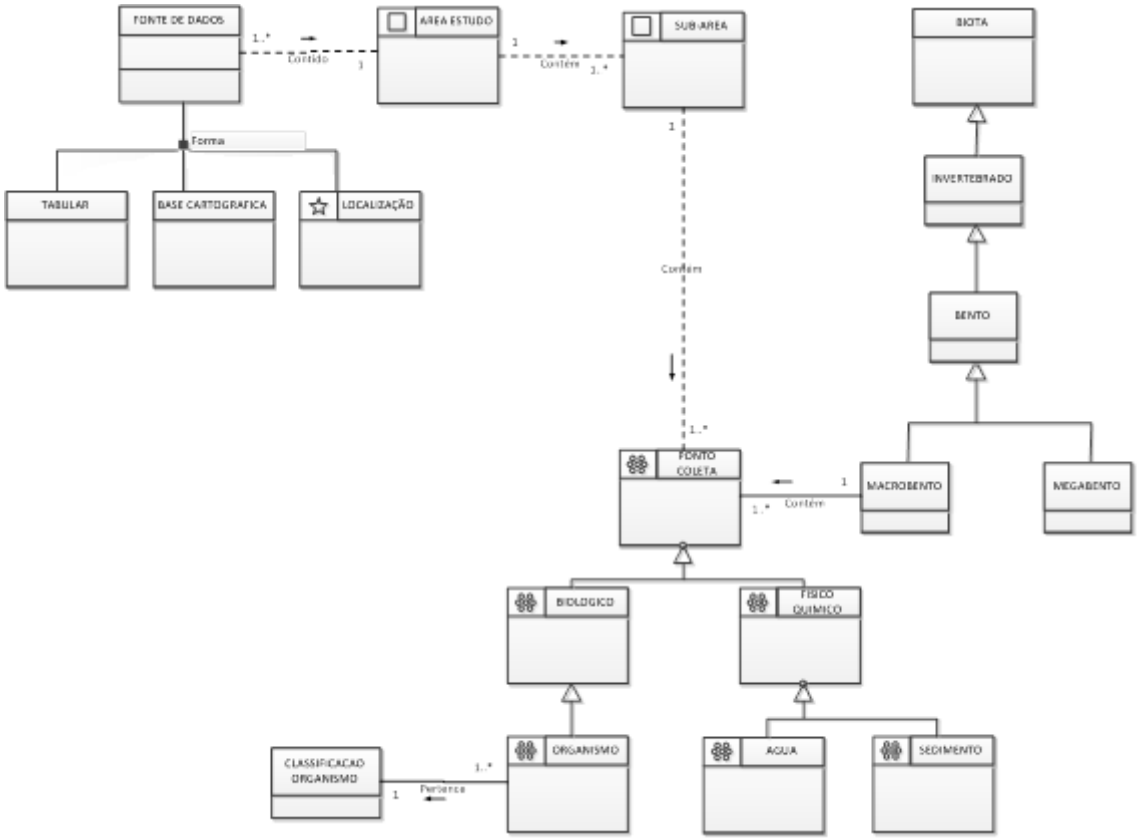
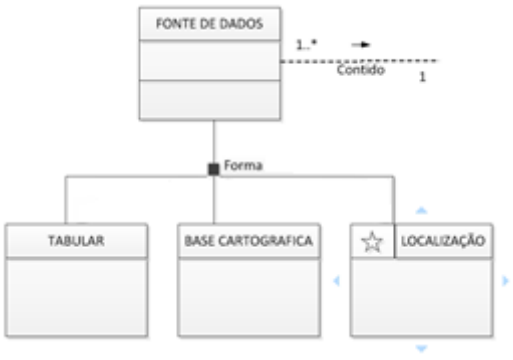


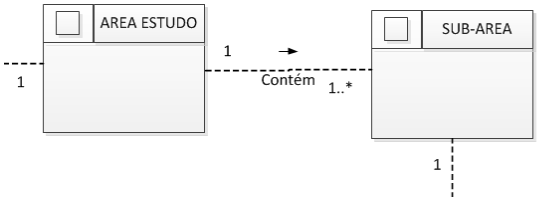
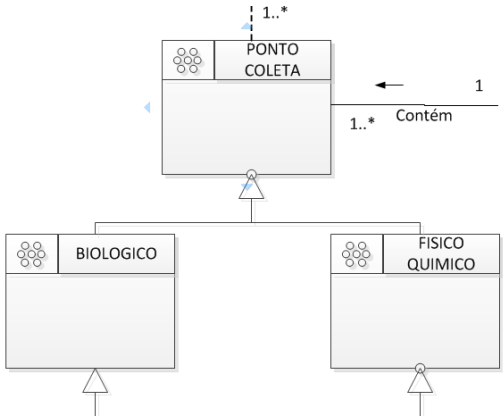
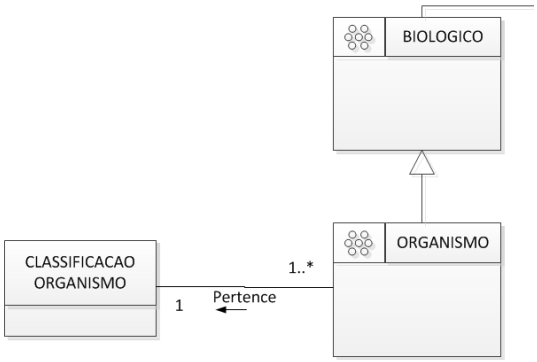
Figura 22. Esquema Conceitual de dados de Macrobentos.

A Tabela 3 ilustra com mais detalhes a composição do esquema conceitual de dados de Macrobentos seguindo o padrão do modelo de dados geográfico OMT-G. Em cada etapa do processo é possível observar os diversos tipos de simbologia que compõem as classes, cada qual com seu significado intrínseco:

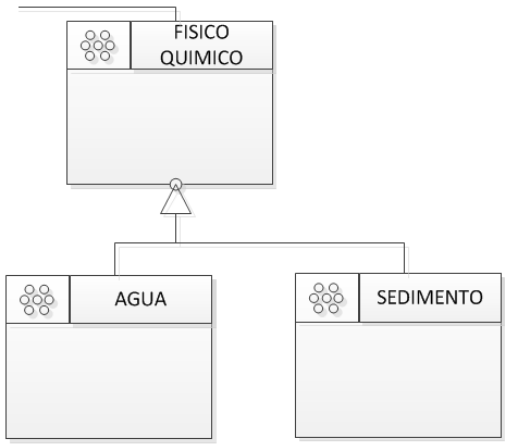
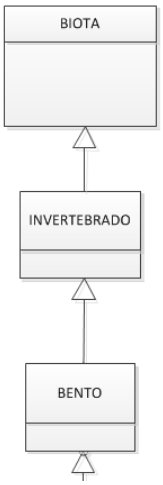
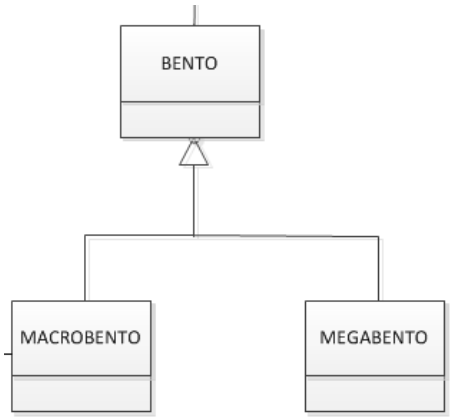
Tabela 3. Detalhamento do esquema conceitual segundo OMT-G.

	A classe denominada FONTE DE DADOS, pode ser generalizada em três subclasses: TABULAR (Estudos Técnicos, Dissertações de Mestrado e Tese de Doutorado que serão usados como subsídio), BASE CARTOGRAFICA (Limites de Município, Estado, etc.) e LOCALIZACAO (Posições dos levantamentos de amostras), sendo estas duas últimas com características espaciais.

continuacão

	A classe AREA ESTUDO delimita espacialmente (restringe) para qual região espacial (RMBS) estas fontes de dados são utilizadas. Esta AREA ESTUDO poderá conter uma ou mais SUB-AREA caso o pesquisador opte por dividir a RMBS em polígonos menores com o objetivo de analisar mais pontualmente cada parcela da região.
	As Classes Área de Estudo e Sub-Áreas supracitadas conterão um ou mais levantamentos amostrais que são representadas pela classe PONTO COLETA. Tal classe sofre uma representação denominada generalização, ou seja, estes Pontos de Coleta podem estar relacionadas às subclasses chamadas BIOLOGICO e FISICO QUIMICO.
	A classe BIOLOGICO é explicada pelos diversos levantamentos amostrais realizados quanto à ocorrência, abundância, diversidade e demais índices ecológicos, como riqueza e AMBI (BORJA et al., 2005) na classe de ORGANISMO, com o objetivo de levantar e classificá-los a partir de parâmetros como Filo, Espécie, Família, etc., que é representado pela classe CLASSICACAO ORGANISMO. Importante salientar que a classificação do organismo no caso do grupo Macrobentos deve ocorrer em nível de espécie, pois os índices ecológicos são baseados na ocorrência e abundância de cada espécie.

continuacão

	O mesmo levantamento amostral também é utilizado para explicar os fatores FISICO-QUIMICO, através da especialização para as duas subclasses denominadas AGUA e SEDIMENTO. No caso da classe AGUA são levantados parâmetros como temperatura, salinidade, profundidade, ph, ecotoxicologia, da água de fundo ou intersticial (existente entre os grãos do sedimento) etc., enquanto a classe SEDIMENTO parâmetros que caracterizam o sedimento como granulometria, teor de matéria orgânica e carbonato de cálcio, entre outros.
	A segunda parte do modelo inicia-se por meio da classe BIOTA, que representa as informações sobre a Fauna que, no caso de Macrobentos encontra-se no grupo do INVERTEBRADO. Os organismos do tipo BENTO (vivem em contato direto com o substrato) são apenas um dos grupos existentes dentro dos Invertebrados, mas de extrema importância para o desenvolvimento deste trabalho.
	Por fim, a classe BENTO pode ser especializada entre MEGABENTO e MACROBENTO, relativos ao tamanho dos organismos, sendo que para este artigo o foco será dado ao segundo. Como se pode verificar, os Macrobentos estão associados à classe Ponto de Coleta. Em um levantamento amostral são realizadas amostragens para o estudo do tema, em localizações específicas.

3.3.2 Modelo Físico

O modelo abstrato do banco de dados serviu como referência para a construção do modelo de implementação física e a posterior inserção das informações provenientes dos estudos supracitados. As classes bem definidas auxiliam o desenvolvedor a pensar e estruturar melhor a organização das tabelas e suas relações, tanto espaciais como alfanuméricas.

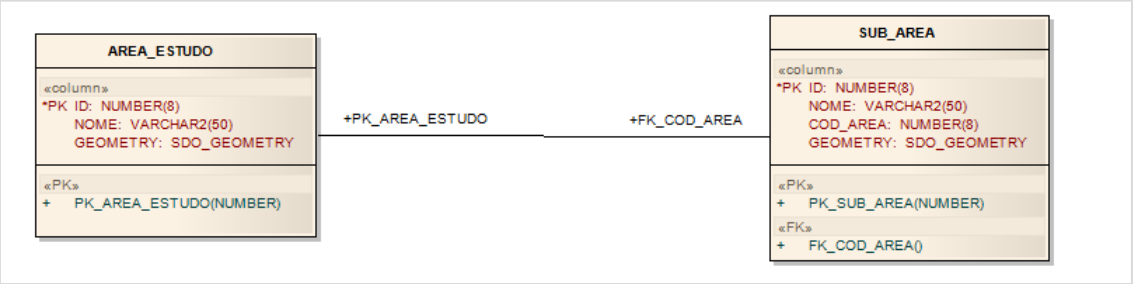


Figura 23. Parte 1 do Esquema Físico de Dados dos Macrobentos.

Nesta primeira parte, na Figura 23, é explicitada apenas a relação entre as tabelas de Área de Estudo e suas possíveis subáreas. Com base no Modelo Conceitual, este relacionamento físico entre as tabelas não existiria já que a relação seria apenas geográfica (dentro de), entretanto, por questões de desempenho, optou-se por utilizar este formato a partir da consulta espacial.

Para a delimitação das áreas (zonas ecológicas) foi levado em consideração os índices ecológicos indicados pelos Macrobentos, já os demais parâmetros como a contribuição dos recursos hídricos, da salinidade e hidrodinamismo (especialmente regime de marés), subsidiam o entendimento sobre as áreas impactadas ou conservadas.

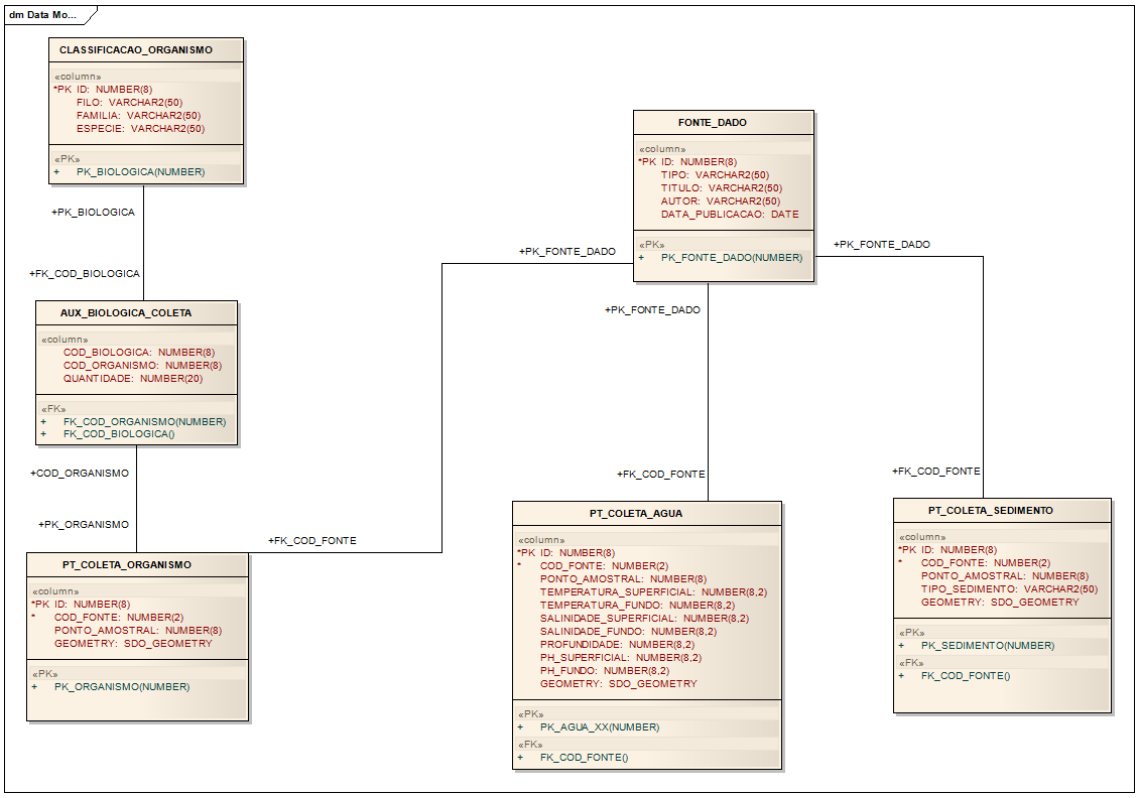


Figura 24. Parte 2 do Esquema Físico de Dados dos Macrobentos

A segunda parte do esquema físico, ilustrada pela Figura 24, transcreve a classe Biota na visão do tema Macrobentos. Uma característica específica para este tema condiz com o fato de que os estudos realizados nem sempre possuem as mesmas informações levantadas ou padronizadas.

Nota-se que nas diversas referências bibliográficas o pesquisador realiza, por exemplo, a coleta de dados utilizando um valor de medida de volume que atende apenas o propósito daquele estudo, não seguindo uma padronização já que nunca houve a preocupação em compartilhar a informação de forma computacional; este tipo de situação é muito crítico quando se tem o objetivo de agregar as diversas fontes e compartilha-las de uma forma que o usuário final tenha uma interpretação igualitária dos dados. Portanto, é imprescindível que neste caso haja um pré-processamento com o objetivo de equalizar os volumes das amostras antes de compartilhá-los.

3.4 IMPORTAÇÃO DE DADOS GEOESPACIAIS

3.4.1 Características dos Dados

Uma das principais etapas na construção de um banco de dados, seja este espacial ou não, condiz com a dificuldade em se homogeneizar e integrar as informações provenientes de diversas fontes. Para o tema Macrobentos a realidade é bem semelhante com o adendo de que existem poucos estudos referentes a este assunto na região e a maioria deles foram desenvolvidos a partir do ano 2000, com exceção de Tommasi (1979).

Além disso, cada uma das fontes de dados selecionada nesta pesquisa foi construída com um propósito específico e pelo fato de não contemplar normas, determinações ou padrões estabelecidos para o levantamento dos dados, notou-se que há uma diferenciação dos dados, que vão desde a quantidade de volume utilizada na coleta a definição do sistema de coordenadas. É sabido que existem diversas técnicas neste levantamento de dados, mas o ponto a se destacar é que não é seguido um único padrão.

O objetivo aqui não é o criticar ou buscar falhas em cada uma das fontes, pois também faltaria subsidio técnico para tal – desejou-se aqui tentar levantar pontos considerados deficientes na perspectiva principal de se agregar diversas informações em um único banco de dados e seu posterior compartilhamento.

Este cenário faz com que exista uma enorme dificuldade para que o meio científico, entidades, etc., por exemplo, consigam analisar todas as estas fontes de forma conjunta sem que não haja algum tipo de perda de informação ou talvez, alguma interpretação equivocada dos dados. Para exemplificar esta situação, a Tabela 4, mostra a listagem de todas as fontes de dados e algumas informações suplementares que ajudam a contextualizar a fonte.

Tabela 4. Lista de fontes de dados utilizadas.

Fonte da Informação	Formato	Local de Estudo	Fonte da Informação	Sistema de Coordenadas
1. Tommasi (1979)	Impresso	Estuário e baía de Santos	Biblioteca IO-USP	Carta Náutica nº 1701 *datum* Córrego Alegre
2. Heitor (2002)	Impresso	Mar aberto da Praia Grande, Baía de Santos e Enseada do Guarujá	Biblioteca Digital USP	Sistema Geográfico *datum* SAD 69.
3. Embraport 2003	PDF	Estuário de Santos, próximo à Ilha Diana	Parceria IBAMA	Sistema Geográfico *datum* WGS-84
4. CODESP (2010)	Excel	Baía de Santos – área de descarte de sedimentos	Parceria CODESP e DTA	Sistema Geográfico *datum* WGS-84
5. CODESP (2008)	Digital - Excel	Estuário e Baía de Santos	Parceria CODESP	Universal Transverse de Mercator (Hemisfério Sul, Fuso 23) *datum* South American 1969 (SAD 69)
6. Ferreira (2008)	Impresso	Baía de Santos e Plataforma Continental Adjacente	Biblioteca IO-USP	Sistema Geográfico *datum* WGS-84

Conforme se pode notar na mesma Tabela 4, há pelo menos três tipos de formatos para as fontes de dados, que são: Impresso, Pdf e Planilha *Excel*, cada qual com suas características e atributos específicos. O local de estudo está intrinsecamente relacionado ao Sistema de Coordenadas fornecido na pesquisa que, conforme se observa, também são bastante distintos.

Já a Tabela 5 ilustra a disposição original dos dados provenientes do levantamento de Macrobentos para uma das fontes utilizadas nesta pesquisa:

Tabela 5. Tabela original ilustrando os pontos de coletas e alguns parâmetros: coordenadas, volume, sedimento e profundidade (Codesp, 2008).

Estação	Coordenadas		Volumes de sedimento	Tipo de sedimento	Profundidade
	Eastings (mE)	Northings (mN)	(L)		(m)
1	364243	7342570	4	Matéria orgânica	14,80
1.1	364243	7342570	5	Matéria orgânica	14,80
1.2	364243	7342570	4	Matéria orgânica	14,80
1.3	364243	7342570	3	Matéria orgânica	14,80
2	364587	7343656	2	Matéria orgânica e cascalho	12,80
2.1	364587	7343656	2	Matéria orgânica e cascalho	12,80
2.2	364587	7343656	2	Matéria orgânica e cascalho	12,80
2.3	364587	7343656	3	Matéria orgânica e cascalho	12,80
3	365514	7345615	4	Cascalho e conchas	7,80
3.1	365514	7345615	4	Cascalho e conchas	7,80
3.2	365514	7345615	3	Cascalho e conchas	7,80
3.3	365514	7345615	4	Cascalho e conchas	7,80
4	365053	7346057	4	Matéria orgânica	9,80
4.1	365053	7346057	5	Matéria orgânica	9,80
4.2	365053	7346057	4	Matéria orgânica	9,80
4.3	365053	7346057	4	Matéria orgânica	9,80
5	368881	7346654	3	Cascalho e conchas	14,80

Nota-se que, um estudo em específico por si só, mesmo quando não comparado à outras fontes de dados, contempla uma heterogeneidade de metodologias e informações como, por exemplo, no volume do sedimento coletado na amostra para cada um dos pontos de coleta (estação) e suas respectivas subestações. Além disso, foi aplicada uma metodologia específica realizando três levantamentos para o mesmo ponto de coleta (estação), como é possível observar na repetição das coordenadas.

Por fim, a Tabela 6, também proveniente do mesmo estudo supracitado, destaca alguns outros parâmetros como é mostrado a seguir:

Tabela 6. Tabela original ilustrando os subpontos de coletas e o Táxon por ponto amostral (Codesp, 2008).

Táxon/Ponto amostral	1.1	1.2	1.3	2.1	2.2	2.3	3.1	3.2	3.3	4.1	4.2	4.3
Volume (llts)	5,00	4,00	3,00	2,00	1,50	3,00	4,00	3,00	4,00	5,00	3,50	4,00
Porifera												
Cnidaria							1,00	2,00				
Nemertea		1,00				1,00	1,00	1,00	1,00			4,00
Bivalvia	1,00		1,00	3,00	2,00	4,00	9,00	12,00	10,00	1,00	4,00	1,00
Gastropoda											1,00	
Polychaeta	4,00	52,00	45,00	13,00	19,00	28,00	88,00	85,00	49,00	21,00	29,00	81,00
Oligochaeta							3,00					
Sipuncula		1,00	1,00				2,00	3,00	1,00		3,00	
Decapoda		1,00		2,00	2,00	1,00		1,00	1,00			
Peracarida		7,00	11,00	6,00		3,00	367,00	777,00	277,00	2,00		
Ostracoda		10,00										
Outros Crustacea									1,00			
Pycnogonida												
Nematoda		1,00					6,00	5,00	2,00		2,00	
Ophiuroidea			1,00				2,00	18,00				1,00
Holothuroidea							1,00	8,00	2,00		2,00	
Pisces												

Este exemplo é uma continuação da tabela anterior e para a leitura correta, é necessário agregá-las. Neste caso específico é exemplificada a quantidade de organismos vivos (Táxon) para cada Ponto Amostral. Há ainda uma grande quantidade de parâmetros levantados no momento da coleta como: condutividade, turbidez, oxigênio dissolvido, temperatura, salinidade, entre outros.

Com o objetivo de ilustrar as diferenças entre as diversas fontes, a Tabela 7 apresentada a seguir é resultado do estudo Heitor (2002). Analisando com mais detalhes, pode-se observar que primeiramente, como o propósito do estudo era distinto de Codesp (2008), foram levantadas apenas duas amostras para cada ponto de coleta, uma no verão (v) e outra no inverno (i). O sistema de coordenadas também difere quando comparado ao primeiro estudo apresentado e não há uma clara menção do volume utilizado.

Tabela 7. Tabela original ilustrando os pontos de coletas no inverno (i) e verão (v). (Heitor, 2002).

Estação	Latitude (S)	Longitude (W)	Profundidade (m)	Temperatura °C	Salinidade (UPS)
1i	-24,032	-46,407	9	23,31	34,11
1v	-24,032	-46,407	12	29,16	32,02
2i	-24,060	-46,394	15	22,16	34,37
2v	-24,060	-46,394	17	26,46	34,81
3i	-24,086	-46,383	19	21,98	34,65
3v	-24,086	-46,383	20	25,55	35,18
4i	-24,132	-46,363	25	21,37	35,46
4v	-24,132	-46,363	26	24,42	35,29
5i	-24,147	-46,408	24	21,32	35,51
5v	-24,147	-46,408	26	23,89	35,54
6i	-24,106	-46,425	17	22,12	34,54
6v	-24,106	-46,425	20	25,55	35,14
7i	-24,077	-46,439	15	21,95	34,44
7v	-24,077	-46,439	18	25,08	35,21

Aplicando este mesmo processo comparativo a cada uma das fontes de dados, nota-se que os mesmos tipos de situação já apresentados são percebidos novamente, ou seja, ausência de parâmetros ou métodos comuns dificultando a comparação e agregação dos estudos. Para que isso seja possível, algumas etapas foram seguidas:

- Padronização do Sistema de Coordenadas
- Padronização do Volume colhido no momento da coleta
- Padronização de nomenclaturas e termos

Mais detalhes sobre as dificuldades em se padronizar os pontos levantados podem ser encontrados no subitem 3.4.2 Dificuldades Práticas.

É importante ressaltar que não foi o objetivo desta pesquisa equiparar cada um dos parâmetros levantados em cada uma das fontes, pois isto envolveria abordar um tema muito mais específico às ciências da Biologia e Fauna Marinha que não é o foco atual. O objetivo aqui é testar o modelo de dados apresentado utilizando alguns parâmetros para isto, como é o caso das etapas supracitadas.

A ideia é que esta proposta possa ser seguida não só para futuros levantamentos sobre o tema (Macrobentos), como também sirva de referência a estudos semelhantes não necessariamente vinculados ao tema.

3.4.2 Dificuldades Práticas

O intuito de se apresentar as dificuldades práticas na elaboração de uma pesquisa é de compartilhar as situações adversas para que futuros estudos possam usar como referência e estas situações sejam diminuídas. Uma árdua etapa na construção de um Sistema de Informações Geográficas é o de adequar bases cartográficas e alfanuméricas. Isto se dá, pois, na maioria das vezes, o pesquisador não tem como objetivo final a disponibilização da pesquisa em formato espacial.

Praticamente para todas as fontes de dados foi necessário realizar uma adequação antes de inseri-las no Banco de Dados. Como já citado, algumas das fontes se encontravam no formato *Adobe Pdf*, que não permite a edição direta. Para estas situações foi necessário primeiramente exportar as informações para o *Microsoft Excel*, padronizar alguns parâmetros, e depois importá-los para o Banco de Dados.

A questão da ausência do Sistema de Coordenadas no momento da realização da coleta também foi algo recorrente. Sem dúvida, este não era o foco principal do estudo, mas esta pesquisa visa iniciar esta discussão de como é importante que o pesquisador tenha consigo uma equipe multidisciplinar que possa apoiá-lo, por exemplo, na definição correta do Sistema de Coordenadas que será utilizado.

Alguns dos estudos não abordaram em nenhum momento qual foi o Sistema de Coordenadas utilizado no momento da coleta; para estes, analisou-se os pares de coordenadas disponibilizados no estudo, normalmente com a unidade grau, minuto e segundo (d:m:s) e deduziu-se que o *datum* utilizado foi o *World Geodetic System 1984* (WGS-84), por ser o padrão da maioria dos hardwares de *Global Positioning System* (GPS). Por fim, os pontos foram plotados no mapa para verificar se coincidiam com a área geográfica em questão.

Por diversas vezes nota-se a confusão em informar apenas o *datum* (elipsoide de referência) utilizado, entretanto, não é informado o Sistema de Coordenadas em si. Por exemplo, informar apenas que o levantamento foi feito usando o *datum South American 1969* (SAD69), porém é necessário conhecer se foi utilizado algum tipo de projeção cartográfica como Policônica, *Universal Transversal de Mercator* (UTM) com seus respectivos parâmetros, ou mesmo coordenadas geodésicas, entre outros.

Uma vez definido o sistema de coordenadas para cada uma das fontes de dados, utilizou-se do software *GeoMedia Professional* para realizar a transformação para o padrão definido nesta pesquisa: Coordenadas Geográficas e *datum WGS-84*.

Um segundo ponto também já abordado é referente a quantidade de volume por litros levantados no momento da coleta. A Tabela 8 a seguir mostra a distribuição deste parâmetro:

Tabela 8. Volume coletado para as amostras nas fontes de dados.

Autor	**Tommasi (1979)**	**Heitor (2002)**	**Embraport (2003)**	**Codesp (2010)**	**Codesp (2008)**	**Ferreira (2008)**
Equip. Coleta	van Veen	van Veen modificado	van-Veen	van-Veen	van-Veen modificado (Petit Ponat)	van Veen e box-corer na Baía; box corer na Plataforma
Volume coletado	10-12 L	-	entre 1,5 e 6L	entre 1,8 e 6,6L	volume padronizado de 5L. Coletas variaram entre 2 e 5 L.	-
Estações de Coleta	42 pontos	45 pontos	12 pontos	09 pontos	40 pontos	14 estações
Nº Amostras	Tréplicas = 126 amostras	Réplicas = 90 amostras	2 réplicas = 24 amostras	triplicatas = 27 amostras	tréplica = 120 amostragens	Duplicatas somadas
Data de coleta	dezembro de 1974 a julho de 1975	outubro de 1997 e março de 1998	Setembro / 02 e fevereiro / 03	dias 11 e 12/11/2009	outubro de 2006	Nov/2004, mar-jul-nov/2005 (Baía); inverno/2005 e verão/2006 (Plataforma).

Observa-se que o volume não foi padronizado nos diversos estudos, com uma variação razoável quando comparado. Porém, o equipamento/método de coleta é semelhante o que se torna possível realizar a homogeneização dos dados. Para homogeneizar os dados, o número e a densidade de animais coletados foram corrigidos para o volume de 5 litros, em função de ser esse o volume mais utilizado nos estudos de referência.

A homogeneização dos termos e definições também é uma etapa primordial na construção de um Modelo de Dados que prevê o seu compartilhamento. Para isso, é importante que se conheça o que determinado parâmetro representa em uma fonte específica para posteriormente poder relacionado. O trabalho neste caso também é totalmente manual e precisa ser feito para cada fonte de dado. A Tabela 9 mostra os dados organizados para a fonte Codesp (2008).

Tabela 9. Ilustração de termos padronizados (Codesp, 2008).

PONTO_AMOSTRAL	COORD_E	COORD_N	DATA_LEVANTAMENTO	VOL_MEDIO	TIPO_SEDIMENTO
1	364243	7342570	09/10/2006	5	Matéria orgânica
2	364587	7343656	09/10/2006	5	Matéria orgânica e cascalho
3	365514	7345615	09/10/2006	5	Cascalho e conchas
4	365053	7346057	09/10/2006	5	Matéria orgânica
5	368881	7346654	09/10/2006	5	Cascalho e conchas
6	368576	7346812	09/10/2006	5	Matéria orgânica
7	368827	7348443	09/10/2006	5	Matéria orgânica e cascalho

A princípio, foi planejado também nesta pesquisa o uso de tecnologia ETL para automatizar a importação de dados para o Banco de Dados modelado. A ferramenta escolhida para testes é chamada de *FME Server*, desenvolvida pela *Safe Software*, que possibilita, entre outras coisas, as seguintes funcionalidades (FME, 2013):

- Validação de Dados e *Upload*: validação de dados espaciais, como problemas de topologia e inserção em alguma Base de Dados Espaciais.
- Conversão de Dados: Permite a realização da conversão de dados entre diversos formatos.
- Distribuição de Dados: Possibilita o compartilhamento de dados por meio de geoserviços espaciais.

Após diversos testes, chegou-se na conclusão que o uso de ferramentas ETL para estes dados específicos de Macrobentos seria inviável, uma vez que não existia um padrão entre os dados, conforme supracitado, e para cada fonte deveria ser feita a adequação manual. Como o grande objetivo de uma ferramenta ETL é de possibilitar a automatização de processos, entre eles a importação de dados, e que neste caso praticamente toda a adequação deveria ser feita manualmente, optou-se pela padronização de tabelas e importação manual.

3.5 DEFINIÇÃO E PREENCHIMENTO DE METADADOS

3.5.1 Criação de Catálogo de Metadados

O catálogo de metadados é, portanto, o repositório que armazena as informações das fontes de dados previamente importadas. Tais informações contidas no catálogo são organizadas conforme a norma ISO 19115 e Perfil MGB, garantindo assim, a interoperabilidade informações espaciais, via *Web*, através de operações de busca, visualização e pesquisa e *download* de dados.

Para a criação do Catálogo de Metadados, foi utilizado o *software* denominado *GeoNetwork*, que é baseado em princípios de Software Livre e de Código Aberto, *Free and Open Source Software (FOSS*) contendo as principais normas internacionais de serviços *Web* e protocolos (ISO e OGC).

Tal solução tem sido amplamente utilizada por diversos órgãos governamentais brasileiros, inclusive sendo adaptada à realidade brasileira pelo IBGE com atividades vinculadas a traduções e alterações na interface, correção de problemas (*bugs*), entre outros. Alguns dos principais portais baseados neste catálogo de metadados são: INDE, MMA, Fundação Nacional do Índio (Funai), Instituto Chico Mendes de Conservação da Biodiversidade (ICMBIO), Emplasa, etc. A Figura 25 ilustra o portal de metadados desenvolvido pela INDE:

Figura 25. Catálogo de Metadados da INDE.

Não é o intuito desta pesquisa abordar com detalhes o ambiente de administração e compartilhamento de dados usando o *software*. Para isso, Ožana e Horáková (2008) discutem com mais detalhes a arquitetura do *GeoNetwork*, ilustrando o funcionamento do ambiente administrativo, de pesquisa, etc. Shvaiko et al. (2010) discutem a criação de uma IDE para a região de Trento na Itália, inclusive adicionando uma camada de pesquisa semântica à solução.

Para a construção do Catálogo de Metadados deste projeto foi utilizada uma máquina servidora com as seguintes características:

- Sistema Operacional*: Windows Server 2008 R2 Enterprise, 64 bits*
- Banco de Dados: *McKoiDB*
- Versão do *software*: 2.6.4-0

Especificamente sobre a utilização do Banco de Dados, como padrão, a solução armazena as informações no banco supracitado, mas há ainda a possibilidade de se optar por outras soluções que também são certificadas, como: *MySQL*, *Postgresql* e *Oracle*.

Durante a construção do projeto foi feito um teste de sucesso utilizando SGBD Oracle como repositório, entretanto, optou-se por utilizar o padrão (*McKoiDB*), principalmente para aumentar o desempenho da aplicação. Entretanto, aconselha-se a utilizar SGBD mais robusto, contendo mais funcionalidades inerentes ao gerenciamento, controle de acesso, segurança, etc., quando utilizado em soluções reais e/ou corporativas.

Uma vez instalada a aplicação, é possível se definir, por meio de interface intuitiva, dentre outras opções, grupos de usuários para garantir ou restringir acesso a determinado registro geoespacial. A Figura 26, mostra em destaque, os *menus* em que se permite importar novos registros de metadados e também adicionar perfis de usuários administradores, editores, visualizadores, entre outros.

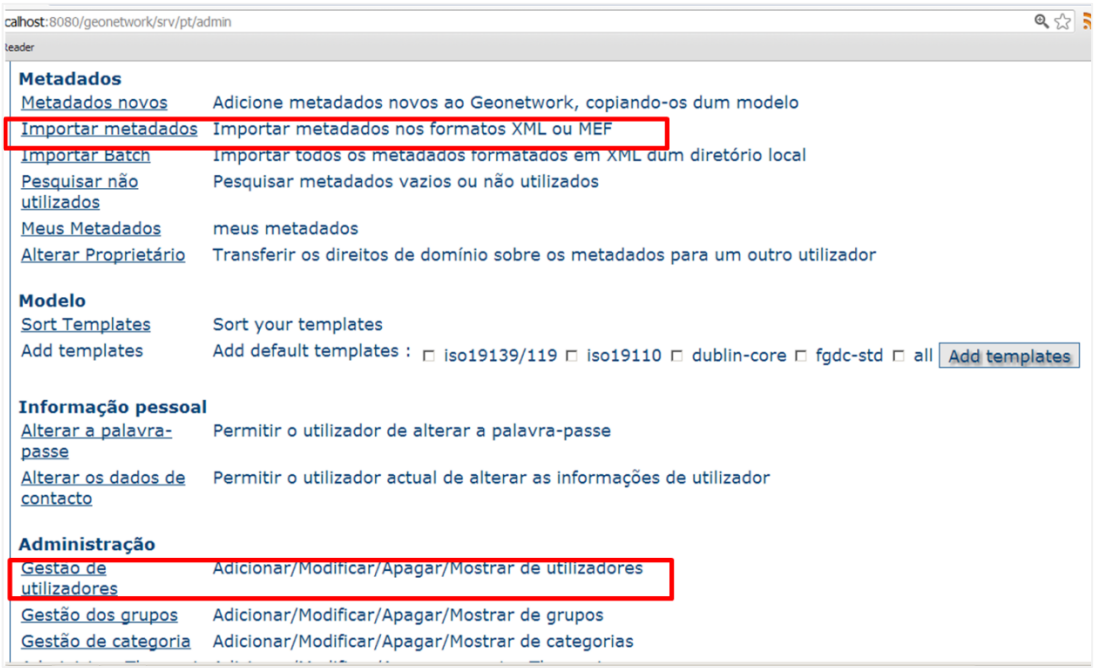


Figura 26. Interface Administrativa da ferramenta *GeoNetwork*.

Ainda é possível realizar a importação do Perfil MGB e configurar o serviço CSW para que os dados possam ser acessados por aplicações terceiras, que é exatamente como o Geoportal (Atlas Ambiental) acessa as informações contidas no Catálogo de Metadados.

3.5.2 Preenchimento de Metadados

Para o preenchimento de metadados, utilizou-se da própria interface do *GeoNetwork*. Para tanto, definiu-se um único padrão seguindo as especificações da ISO 19115 e Perfil MGB para ser usado como referência em todas as catalogações; especificações estas que são seguidas pelo *software* escolhido. A Figura 27 ilustra quais campos são obrigatórios e opcionais no preenchimento de metadados – isto é para garantir que todos os registros tenham o mínimo de informações que permitam o seu acesso por outras organizações:

Entidades e elementos do Núcleo de Metadados do Perfil MGB Sumarizado			
Entidade / Elemento	**Obrigatoriedade**	**Entidade / Elemento**	**Obrigatoriedade**
1. Título	obrigatório	12.Tipo de Representação Espacial	opcional
2. Data	obrigatório	13.Sistema de Referência	obrigatório
3. Responsável	obrigatório	14.Linhagem	opcional
4. Extensão Geográfica	condicional	15.Acesso Online	opcional
5. Idioma	obrigatório	16.Identificador Metadados	opcional
6. Código de Caracteres do CDG	condicional	17.Nome Padrão de Metadados	opcional
7. Categoria Temática	obrigatório	18.Versão da Norma de Metadados	opcional
8. Resolução Espacial	opcional	19.Idioma dos Metadados	condicional
9. Resumo	obrigatório	20.Código de Caracteres dos Metadados	condicional
10.Formato de Distribuição	obrigatório	21.Contato para Metadados	obrigatório
11.Extensão Temporal e Altimétrica	opcional	22.Data dos Metadados	obrigatório
		23.Status	obrigatório

Figura 27. Perfil MGB sumarizado (CONCAR, 2009).

O preenchimento de metadados normalmente é feito por aqueles usuários que não produziram o dado e é exatamente por isso que se definem campos obrigatórios em cada registro para tentar evitar a falha, já que, na maioria das vezes, é um procedimento bastante manual.

De uma maneira resumida, busca-se primeiramente entender os dados e o padrão de metadados a serem utilizados, depois quais campos fazem sentido para determinada catalogação, revisão e por fim, a forma de como o dado deve ser disponibilizado.

A seguir pode-se observar a descrição detalhada de metadados para a fonte de dados da macrofauna bentônica (Heitor, 2002), após o preenchimento no *software* e exportação das informações:

Title:

Composição e distribuição da macrofauna bentônica em áreas sob a influência da disposição oceânica de esgotos municipais da Baixada Santista e no Canal de São Sebastião

Abstract:

O presente estudo teve como objetivo conhecer a composição e distribuição da macrofauna bentônica da região da praia Grande, Baía de Santos, Enseada do Guarujá e do Canal de São Sebastião, áreas onde ocorre disposição oceânica de esgotos municipais pela SABESP, e analisar as associações de espécies e suas possíveis relações com as variáveis ambientais.

ISO 19115 - Metadata

Metadata information:

File identifier: cd8c4 189-2057-4ee6-a910-8fc43618ad97

Metadata language: portuguese

Metadata character set: utf8

Hierarchy level: dataset

Metadata point of contact:

Organization name: Universidade de Sao Paulo

Address (location of the responsible individual or organization):

Electronic mail address: gabrielncarvalho@gmail.com

Role: author

Metadata standard name: ISO 19115:2003/19139

Metadata standard version: 1.0

Identification information:

Citation:

Title: Composição e distribuição da macrofauna bentônica em áreas sob a influência da disposição oceânica de esgotos municipais da Baixada Santista e no Canal de São Sebastião

date: type of date: publication

Resource identifier: 123

Party responsible for the resource:

Abstract:

O presente estudo teve como objetivo conhecer a composição e distribuição da macrofauna bentônica da região da praia Grande, Baía de Santos, Enseada do Guarujá e do Canal de

São Sebastião, áreas onde ocorre disposição oceânica de esgotos municipais pela SABESP, e analisar as associações de espécies e suas possíveis relações com as variáveis ambientais.

Status of the dataset: completed

Point of contact:

Organization name: Universidade de Sao Paulo - Instituto Oceanográfico

Phone: -

Address (location of the responsible individual or organization):

Electronic mail address: -

Role: point of contact

Maintenance and update frequency information: not planned

keywords:

Macrobentos

Heitor

Baixada Santista

Biomonitoramento

type of keywords: theme

Method used to represent geographic information (spatial representation type):

vector

Resource character set: utf8

Topic category:

Resource extent (Extent information):

Geographic bounding box (extent):

West bound longitude: -49.8046875

East bound longitude: -42.24609375

South bound latitude: -26.575927734375

North bound latitude: -20.775146484375

Data quality information:

Scope of quality information: dataset

lineage: Dados foram construídos a partir de arquivo pdf (tese de doutorado), transformado para planilha e posteriormente para vetor.

Report:

Reference system information:

Reference system identifier:

Code: SAD69 (EPSG: 4618) code space: EPSG

Information about distribution:

Distributor Transfer Options:

URL: http ://geoserver/W MS_ HEITOR/service .svc/get

3.5.3 Compartilhamento de Informações

Uma das etapas cruciais durante a construção de uma IDE é garantir a facilidade e agilidade de acesso as informações por aqueles que a necessitam para tomar decisão. As informações contidas no Catálogo de Metadados (Servidor) podem ser acessadas de duas maneiras distintas:

a) Utilizando o próprio portal da tecnologia *GeoNetwork*, conforme mostra a Figura 28. É permitido ao usuário realizar pesquisa específica para as seguintes perguntas: O quê (título, resumo, palavra-chave), Onde (coordenadas, região) e Quando (determinado período de tempo). Os resultados são apresentados diretamente na interface do *software* possibilitando ao usuário navegar sobre os registros, clicar naquele de houver o interesse e obter informações mais específicas.

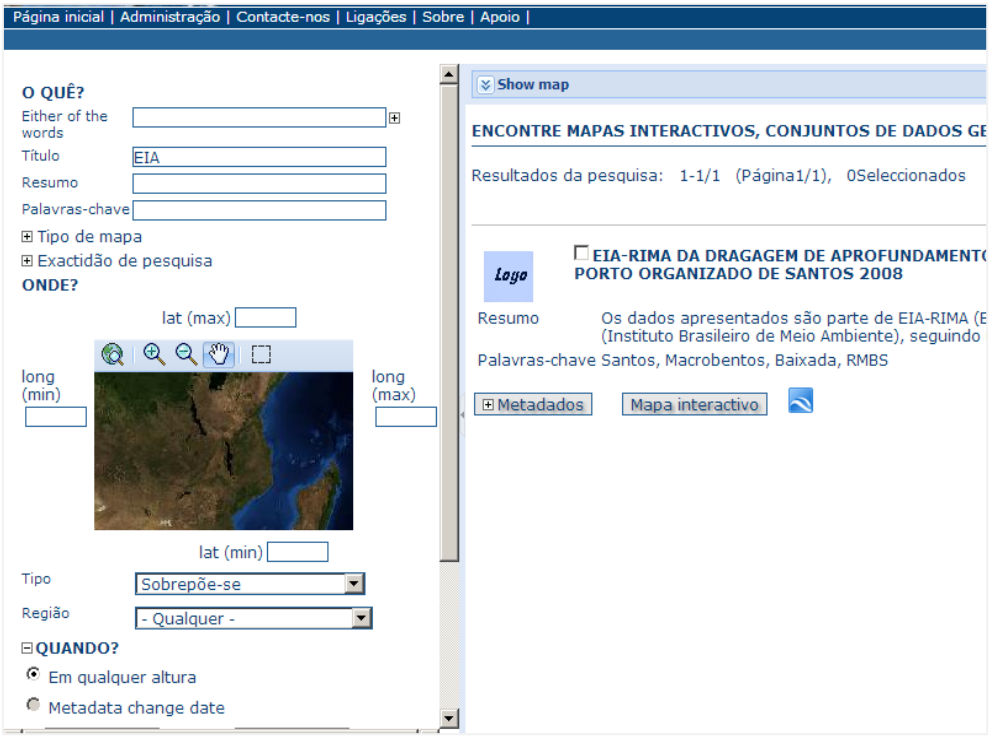


Figura 28. Interface de pesquisa no *GeoNetwork*.

b) Compartilhar as informações, por meio de *Web Services*, para serem visualizadas em ambiente terceiro, como é o caso do Geoportal. Para isso, é necessário ter acesso a URL específica do servidor de metadados. A URL a seguir mostra uma requisição sendo feita ao servidor de Catálogo de Metadados com o método *GetCapabilities*. Esta URL pode ser adicionada diretamente no Geoportal permitindo ao usuário realizar exatamente a mesma pesquisa (O quê, Onde e Quando) e obter os mesmos resultados. A Figura 29 ilustra um trecho do resultado da requisição do *Web service* ao Catálogo de Metadados, onde pode ser visualizada as palavras-chave (*keyword*), o tipo do serviço (CSW) e sua versão (2.0.2).

http://localhost:8080/geonetwork/srv/pt/csw?SERVICE=CSW&VERSION=2.0.2&REQUEST=GetCapabilities

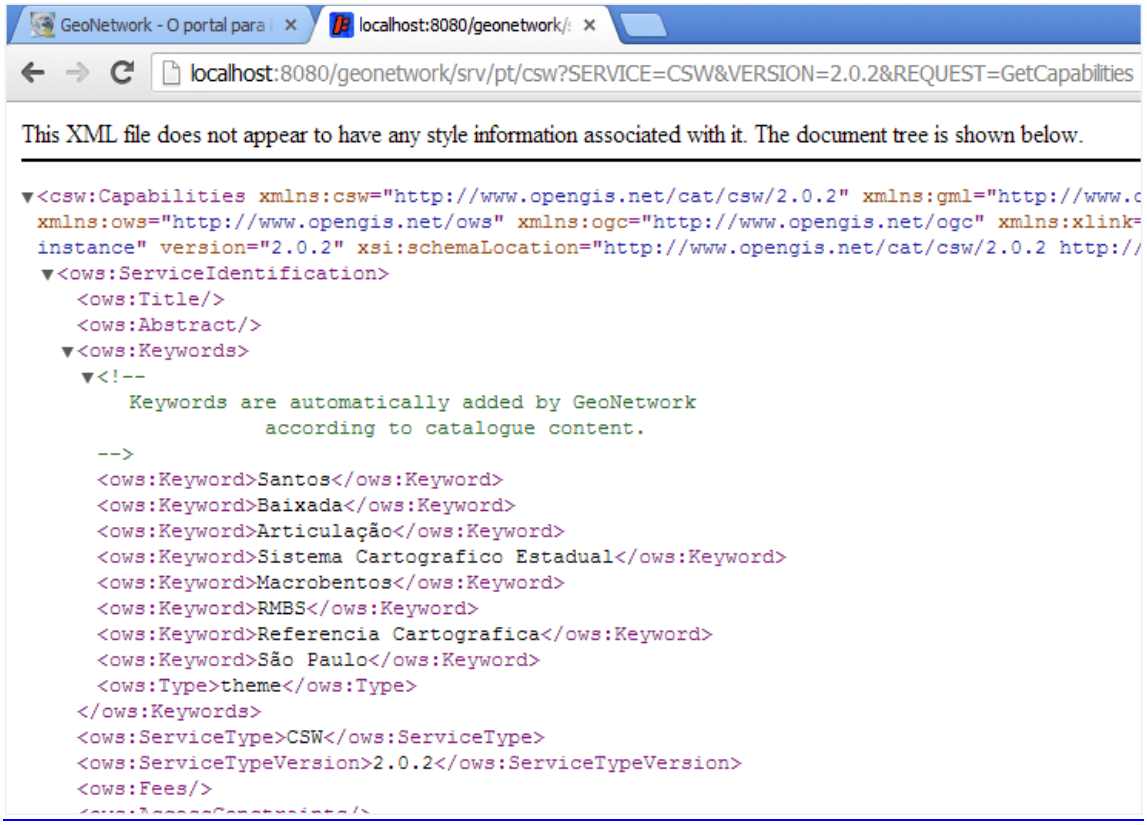


Figura 29. Resultado de requisição URL feita ao Catálogo de Metadados *GeoNetwork*.

3.6 CONSTRUÇÃO DO ATLAS AMBIENTAL (GEOPORTAL)

3.6.1 Características Gerais

Na instalação do *software Geonetwork*, para a presente versão, já está incluso o módulo denominado *Geoserver*, que é responsável pela geração de serviços geoespaciais no padrão OGC, como: WMS, WFS e WCS. Todavia, trata-se de uma ferramenta relativamente simples, com a ausência de algumas funcionalidades importantes como:

- Não contem interface integrada para pesquisa de metadados: esta deve ser feita primeiramente na interface do *GeoNetwork* e depois integrá-la ao portal
- Não permite a conexão com distintos padrões de serviços OGC como WFS-G, WMTS, etc.

Para a construção deste trabalho, foi utilizada uma das ferramentas *case* do mercado geoespacial, chamada *GeoMedia SDI Portal* (GEOMEDIA SDI, 2012), desenvolvida pela empresa *Intergraph Corporation*. A tecnologia *Intergraph* é uma das que estão sendo utilizada na construção de diversas IDEs ao redor do mundo. Algumas soluções se destacam como: (Poland Geoportal, 2013; SIGNA, 2013; IDEG, 2013; GIS4EU, 2013; GeoBilbao, 2013) entre outros.

A Figura 30 exemplifica a arquitetura do *framework* desenvolvido para a ferramenta:

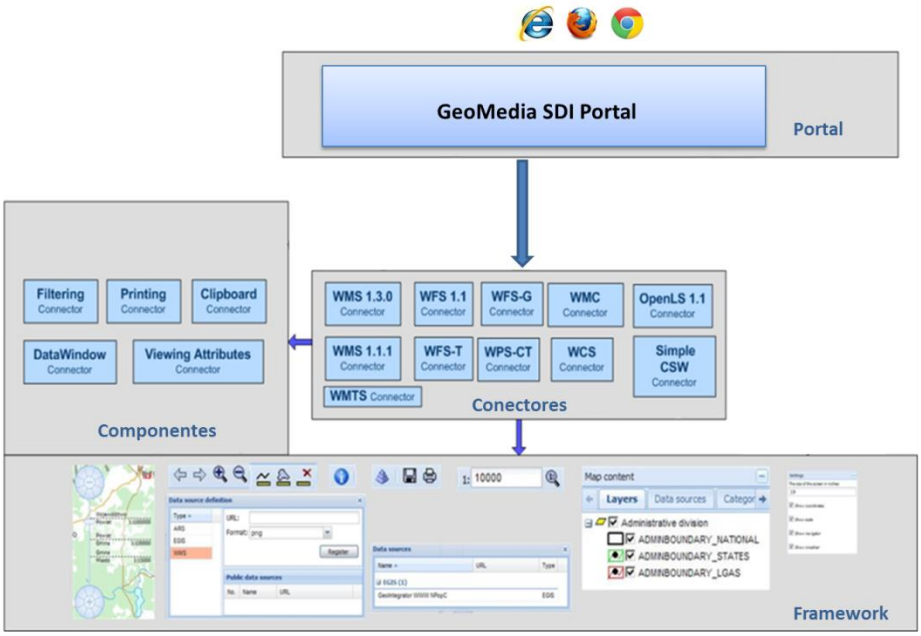


Figura 30. Arquitetura do Geoportal e suas principais funcionalidades (GeoMedia SDI, 2012).

A arquitetura exposta proporciona a integração com diversos padrões de serviços geoespaciais, por meio dos conectores seguindo a especificação OGC, permitindo a realização de operações intrínsecas aos geoportais como, a consulta alfanumérica de dados espaciais e de metadados, a impressão, medições e cálculos, a exportação de dados para formatos terceiros como *Microsoft Excel*, entre outros.

A estrutura física utilizada para a instalação da ferramenta é a mesma da apresentada anteriormente para a aplicação *GeoNetwork*, entretanto, durante uma implementação em ambiente de produção, recomenda-se a previsão de duas máquinas servidoras, cada qual para armazenar as aplicações separadamente. A seguir, destaca-se as características gerais do *GeoMedia SDI Portal*:

- Banco de Dados: *Microsoft Access*
- Versão do software: 6.1.1.26 (r7418)

- .NET 3.5

O Banco de Dados *Microsoft Access* selecionado na arquitetura tem como propósito principal armazenar as informações administrativas do sistema como configurações de usuários, definições das ferramentas do portal, conectores de serviços geoespaciais, tela inicial site, entre outros.

Já a Figura 31 evidencia a forma como os conectores geoespaciais atuam concomitantemente a partir de uma requisição do usuário, que neste caso, é ilustrado realizando a operação de zoom à uma área específica (retângulo vermelho). Cada serviço é requisitado simultaneamente, mas de forma independente, pois, caso algum dos provedores esteja enfrentando algum tipo de limitação (sem conectividade, por exemplo), o geoportal não será afetado de uma maneira geral e o usuário será avisado que determinado serviço não respondeu a requisição.

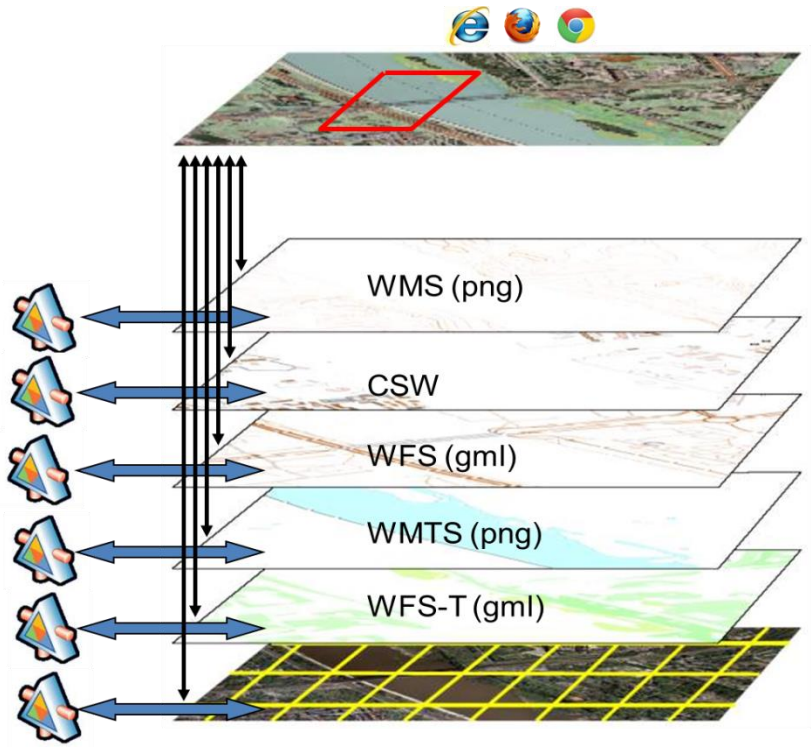


Figura 31. Consultas simultâneas da ferramenta (GeoMedia SDI, 2012).

3.6.2 Criação de Serviços Geoespaciais

A etapa de criação de serviços geoespaciais consiste na conexão com os dados provenientes da importação apresentada no item “3.4 Importação de Dados Geoespaciais”. A interface *desktop*, representada pela ferramenta *Intergraph GeoMedia Professional*, responsável em estabelecer a conexão com o Banco de Dados *Microsoft Access* e permitir que alguns parâmetros sejam definidos como: simbologia, ordem de visualização das feições com base na escala, entre outros. A segunda etapa consiste na criação em si dos serviços geoespaciais. Por meio de interface administrativa é possível definir primeiramente qual é o tipo (WMS, WFS, etc.) e o nome do serviço que deseja ser criado. Como já previamente discutido, o armazenamento dos metadados dos serviços e demais parâmetros estão localizados em diversas pastas e no Banco de Dados *Microsoft Access* do servidor.
Dependendo do tipo de aplicação que se deseja criar, é possível construir apenas um serviço geoespacial de metadados e associá-lo a todas as feições cartográficas presentes no projeto. Desta forma, quando determinado usuário requisitar o serviço em sua aplicação, o resultado listará todas as informações (camadas cartográficas) presentes no serviço, conforme mostra a Figura 32. Além disso, o metadado criado para descrever este geoserviço deverá contemplar de uma maneira genérica as informações de todas as feições presentes no serviço. Normalmente tal modelo é utilizado por empresas e órgãos que produzem dados geoespaciais como Limite de Estado, Curva de Nível, Limite de Municípios, etc., que possuem a mesma característica (data de produção, escala, entre outros) e desejam compartilha-los para a comunidade.

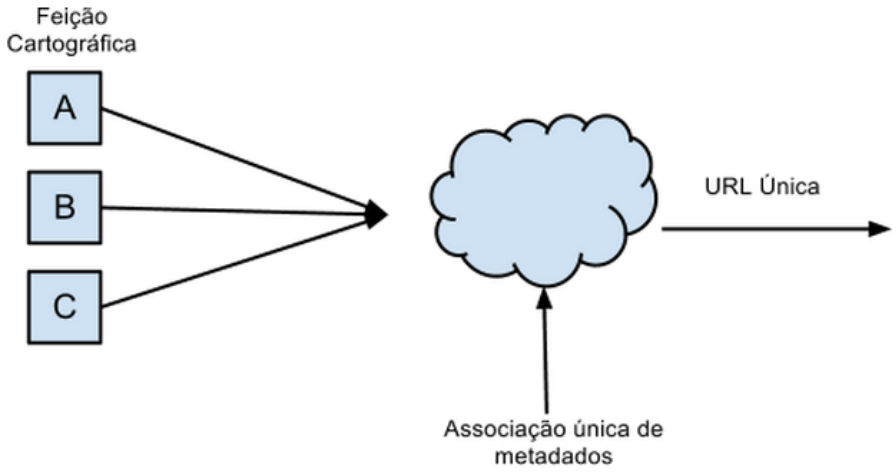


Figura 32. Associação de feições espaciais a um único registro de metadados.

Porém, esta estrutura não atende as necessidades da presente pesquisa, já que as fontes de dados levantadas para o tema Macrobentos foram realizadas em momentos e por pesquisadores e/ou empresas distintos, necessitando uma distinção entre as fontes. Como resultado, foi criado um serviço no formato OGC WMS para cada fonte existente (feição cartográfica). A associação entre o serviço criado e a fonte (feição cartográfica) é feita diretamente na interface *desktop*. Há neste caso um botão responsável em publicar as informações presentes no ambiente *desktop* para o geoserviço escolhido.

A integração entre o geoserviço com seu respectivo registo de metadados deve ser feito diretamente na interface do *GeoNetwork* inserindo a URL que representa o WMS criado, conforme destaca o retângulo da Figura 33: define-se no campo Ligação e especificamente URL o serviço espacial, além da versão, nome e breve descrição da informação. Esta associação permite que o usuário dentro do Geoportal possa pesquisar determinado metadado e visualizar a feição geográfica relacionada.

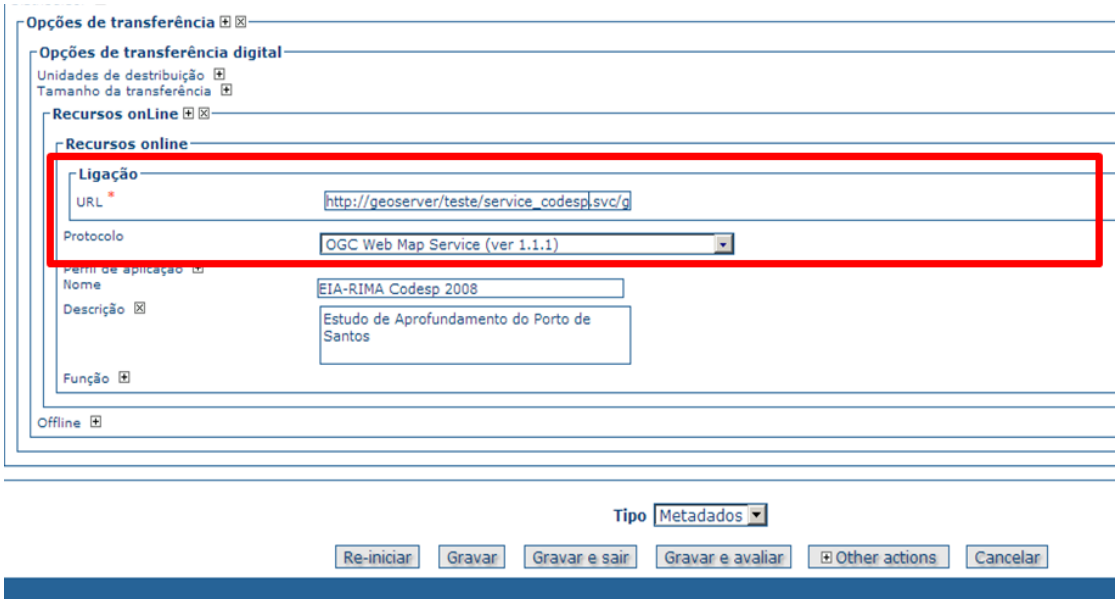


Figura 33. *Menu* acesso online integrando o registro de metadados com o dado espacial via URL OGC WMS.

3.6.3 Navegação pelo Portal

A construção do Geoportal é essencial para que as informações sejam devidamente compartilhadas por uma interface intuitiva. Uma das principais características destas tecnologias é minimizar cada vez mais a necessidade do usuário final possuir, como pré-requisito, um conhecimento avançado de geotecnologias de um modo geral. A Figura 34 representa exatamente este objetivo, mostrando do lado direito do mapa, os temas (legenda) contendo as feições cartográficas referentes a Macrobentos (Codesp, 2008; Heitor, 2002), além da base cartográfica de limites de Municípios e Estado.

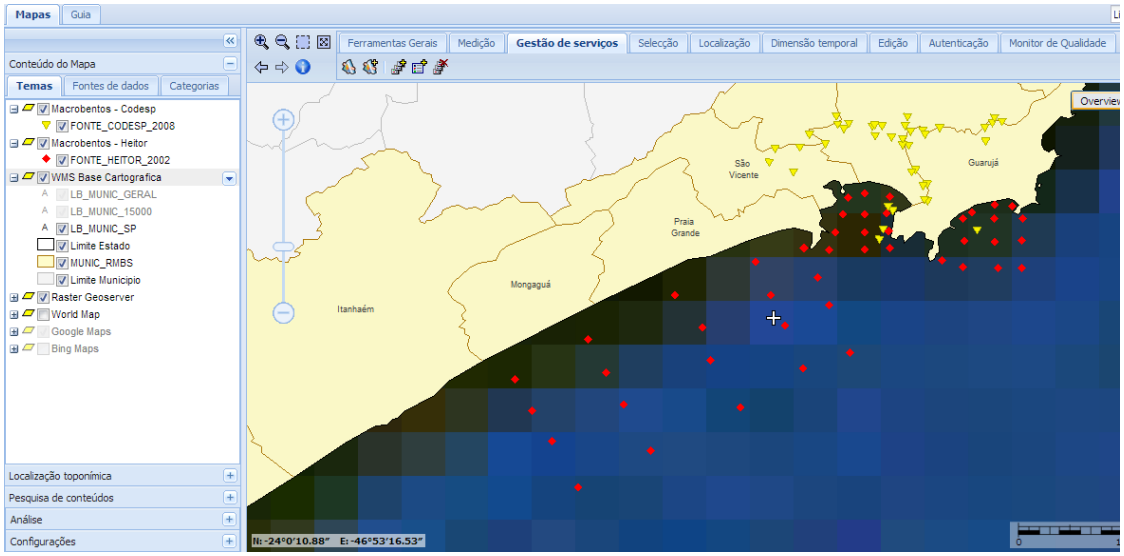


Figura 34. Visão Geral do Geoportal.

As informações associadas a uma determinada feição também pode ser obtida por meio da consulta de metadados. A Figura 35 destaca primeiramente com um retângulo vermelho o campo para busca dos metadados, que pode ser feito via palavra-chave, título, autor, limite espacial, entre outros; neste caso, conforme é possível observar, a palavra-chave utilizada foi Macrobentos e o resultado é mostrado na caixa logo abaixo do mesmo retângulo vermelho. Uma vez que o usuário clique em um determinado registro, uma nova janela é aberta detalhando o conteúdo para aquela informação – isto é exatamente o que a parte do lado direito da figura ilustra.

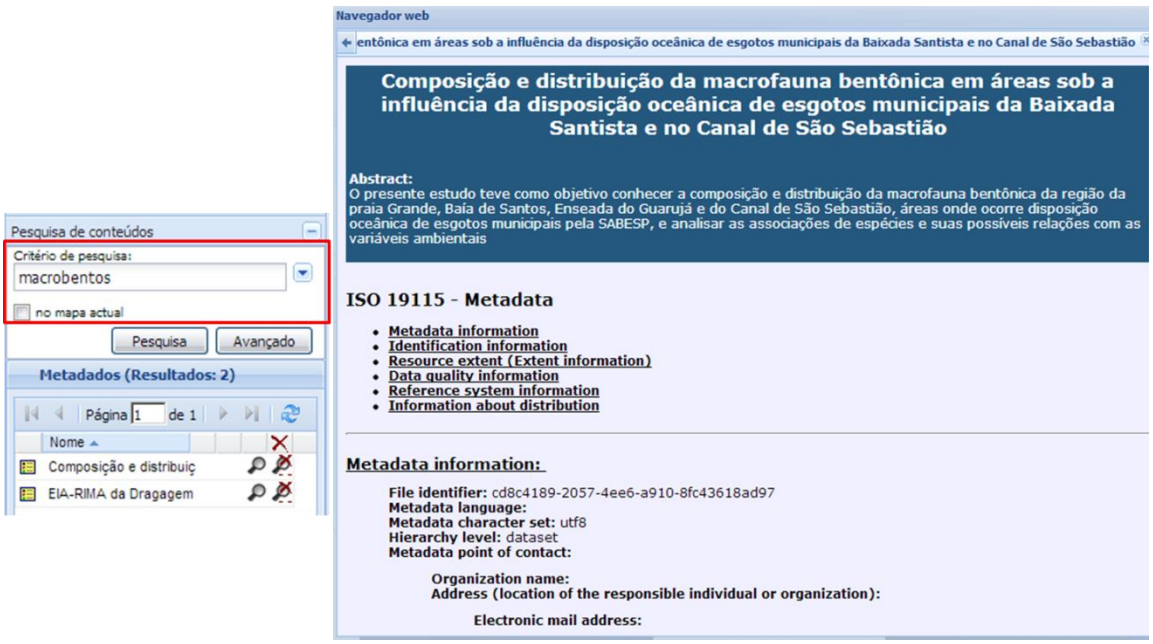


Figura 35. Interface de busca de metadados e janela com descrição detalhada.

O usuário também pode adicionar novos serviços WMS, WFS, WMTS, etc., diretamente no portal, proveniente de fontes externas. Neste caso em específico, a Figura 36 mostra a conexão com dois serviços providos pelo Instituto Brasileiro de Geografia e Estatística (IBGE) a partir das seguintes URLs:

http://www.geoservicos.ibge.gov.br/geoserver/gn/wms?service=WMS&version=1.1.0 = contendo informações de Limites Políticos dos Países (em amarelo no mapa).

http://www.geoservicos.ibge.gov.br/geoserver/IBGE/wms?service=WMS&version=1.1.0 = contendo as Regiões Metropolitanas no Brasil (em vermelho no mapa).

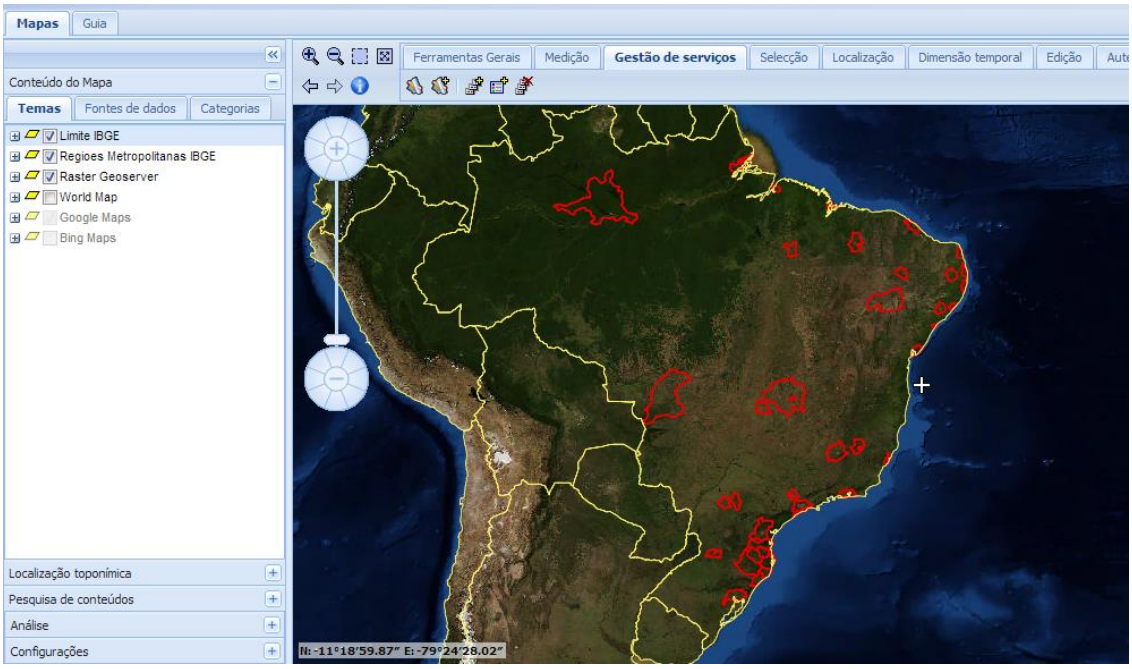


Figura 36. Adição de serviços externos do IBGE ao Geoportal.

Além de serviços externos, é permitido ao usuário alterar o mapa de fundo existente na aplicação para as tecnologias *Google Maps* ou *Bing Maps*. Isto pode, sem dúvida, auxiliar a tomada de decisão na análise de imagens de diferentes épocas, por exemplo, conforme mostra a Figura 37:

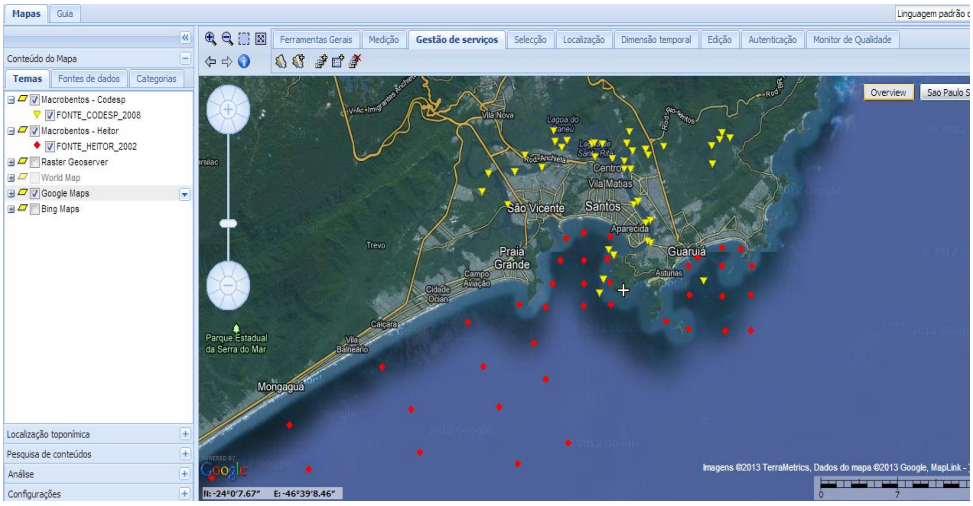


Figura 37. Imagens do *Google Maps* como plano de fundo somado aos pontos de coleta de Macrobentos.

Por meio da interface do Geoportal é possível também construir diversos mapas para atender uma demanda em específico e salvá-los para posterior uso ou então, a realização de uma impressão do conteúdo. A Figura 38 mostra a caixa de propriedades da impressão, na qual se pode definir a impressora, formato da imagem, modelo (croqui específico), dimensão da página (A4, A3, A2, etc.), orientação (Retrato ou Paisagem) e escala; a figura do mapa é resultante da impressão.

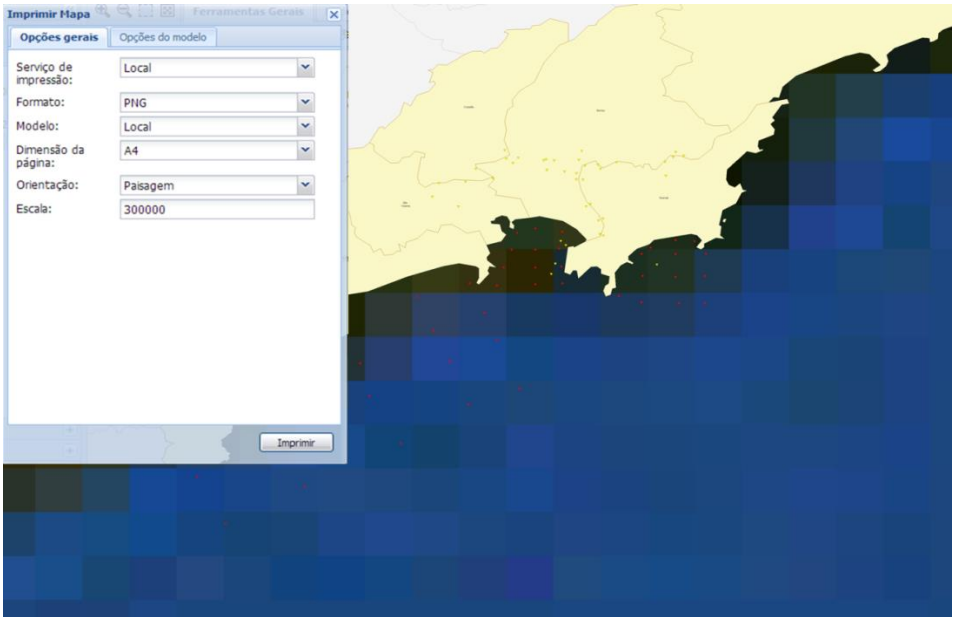


Figura 38. Imagens do *Google Maps* como plano de fundo somado aos pontos de coleta de Macrobentos.

Por fim, é possível também construir algumas análises, como por exemplo, medições de coordenadas, de distâncias lineares e polígonos. A Figura 39 mostra

com detalhes a coordenada de alguns pontos inseridos manualmente como a distância entre três pontos de coleta em quilômetros (*km*).

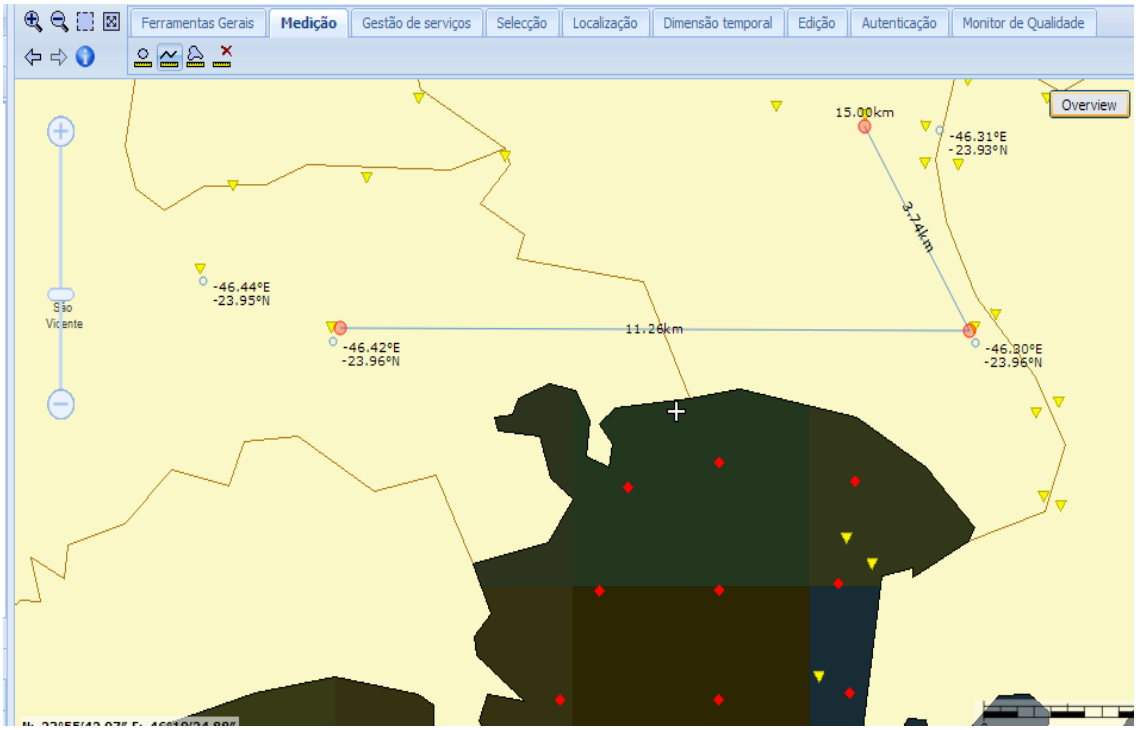


Figura 39. Medições de distância entre determinados pontos de coleta.

4 RESULTADOS

4.1 INTRODUÇÃO

O objetivo deste capítulo é discutir sobre os resultados obtidos desde a modelagem de dados até o compartilhamento de informações por interface intuitiva. A Figura 40 ajuda a explicar a arquitetura resultante do projeto para a construção do Atlas Ambiental (Geoportal) com informações de Macrobentos:

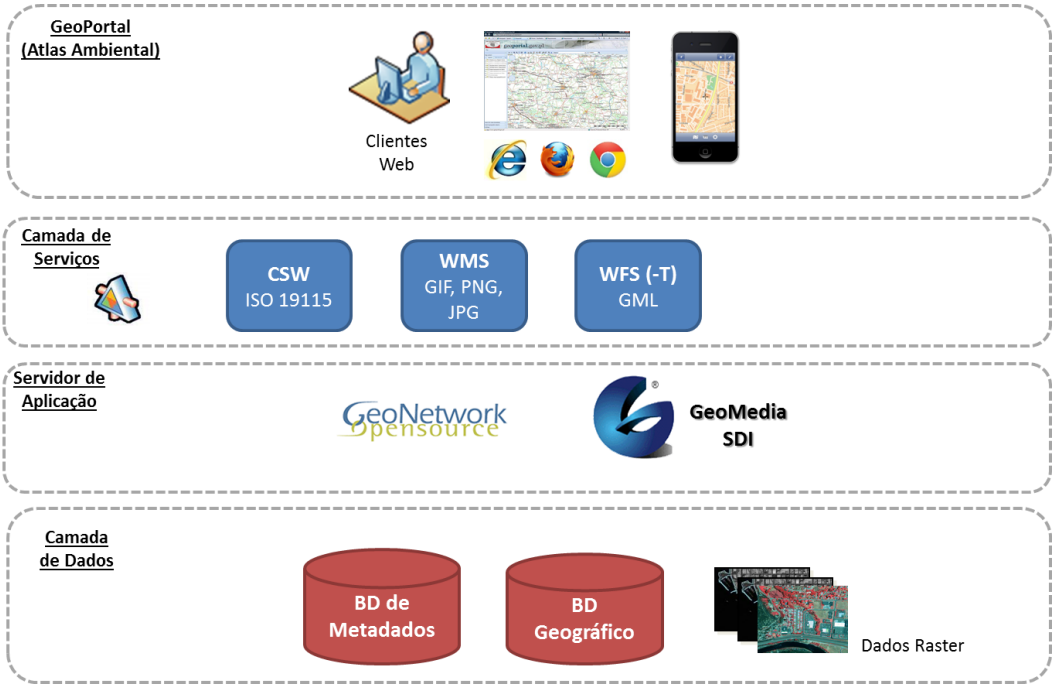


Figura 40. Arquitetura desenvolvida para o Geoportal.

A Camada de Dados foi composta por dois Bancos de Dados, sendo o primeiro responsável pelo armazenamento de informações de metadados e o segundo para dados espaciais. Ambos foram administrados por aplicações contidas no Servidor de Aplicação por meio dos seguintes *softwares*: *GeoNetwork* e *GeoMedia SDI Portal*.

A disponibilização dos dados ocorreu através da Camada de Serviço que, seguindo o padrão o OGC, compartilhou as informações por meio de geoserviços como: CSW (padrão ISO 19115) para metadados, WMS (formatos *.gif, *.png e *.jpg) para informações *raster* e WFS (formato *.gml) para dados vetoriais.

O usuário cliente do Geoportal teve a possibilidade de acessar as informações através do *browser*, aplicativo móvel (celular) e até mesmo utilizando um software

desktop que contemple funcionalidades de leitura de serviços no padrão OGC, como é o caso do *GeoMedia*, *ArcGIS*, *AutoCAD*, *Quantum GIS*, entre tantos outros.

4.2 CONSIDERAÇÕES

O entendimento das fontes de dados e suas respectivas tabelas foram, sem dúvida, um dos processos mais complexos nesta pesquisa, pois envolveu primeiramente, o entendimento do que aquele dado representava no estudo específico e depois, como esta fonte poderia ser adequada a um modelo mais conceitual e genérico. A construção em si do Modelo Conceitual para Macrobentos envolveu uma equipe interdisciplinar contemplando profissionais da área de Biologia, Geografia, Engenharia, Geologia, entre outros.

O Modelo Físico de Dados foi construído mediante análise dos atributos e outras informações presentes nas fontes de dados refletidas no Modelo Conceitual proposto. O objetivo inicial nesta pesquisa era de utilizar ferramenta ETL para facilitar e automatizar a importação de dados para o Modelo Físico, entretanto, devido a grande heterogeneidade das informações, esta etapa foi retirada do processo. Constatou-se, depois de seguidos testes, que o esforço para padronizar cada uma das fontes, para permitir que a ferramenta ETL pudesse ler, extrair e transformar para o Modelo Físico, era maior do que adequá-las manualmente. Vale ressaltar que este processo pode ser mais bem aplicado em situações onde os dados não sejam tão heterogêneos ou que tenham a mínima preocupação com a padronização de colunas, dados, etc. Devido a situação atual, optou-se por adequar e importar as fontes de dados manualmente ao modelo físico.

A etapa da elaboração dos mapas, utilizando ferramentas de geoprocessamento, ocorreu de forma satisfatória. Com os dados de Macrobentos representados como geometrias do tipo Ponto foi possível definir estilos para representação e algumas configurações básicas como a de visualização de atributos. O mesmo se deu para as camadas geográficas de Limite Estadual e Municipal provenientes do IBGE. Com as feições devidamente configuradas no ambiente desktop foi possível utilizar de interface intuitiva para a geração e publicação de serviços no padrão WMS.

Já a criação do ambiente de metadados, desde a instalação da ferramenta, a configuração e preenchimento de informações se deu também de forma satisfatória, conforme era esperado, dado a grande quantidade de pesquisas que estão sendo realizadas com a tecnologia *GeoNetwork*, como é o caso da própria INDE. A integração dos serviços geoespaciais, embora feita manualmente, com os registros de metadados também atendeu a contento, permitindo ao usuário final realizar buscas por palavras-chaves, descrição, tema, etc., e obter as informações requisitadas, inclusive o WMS associado.

A construção do Geoportal, Atlas Ambiental, foi construído baseado na tecnologia *GeoMedia SDI Portal* e atendeu as necessidades desta pesquisa que era de mostrar a importância de conseguir compartilhar informações, antes armazenadas apenas em formato analógico, de uma forma rápida e intuitiva através de uma interface intuitiva na qual o usuário pode, além de realizar pesquisas alfanuméricas e/ou espaciais, também funcionalidades como medições, impressões, etc.

5 CONCLUSÃO

Nesta pesquisa foi proposto um modelo conceitual para integração de dados, com o estudo de caso sobre Macrobentos, idealizado para ser utilizado em uma IDE, como uma possível solução para minimizar as dificuldades de se organizar dados espaciais e alfanuméricos sem padronização. Além do modelo em si, foram indicadas as etapas necessárias para a sistematização e modelagem de dados ambientais, a exemplo do objeto estudado.

Constatou-se que a modelagem de dados ambientais, além de ser uma alternativa para a representação do conhecimento de uma forma estruturada, exerce um papel facilitador na integração de dados provenientes de diversas fontes, para que possam ser compartilhados de maneira padronizada em um Atlas Ambiental Web (Geoportal). A sequência metodológica da construção de um modelo de dados espaciais é descrita, levando em consideração a integração de conhecimento de especialistas sobre os temas abordados.

Especificamente para a área ambiental e o tema Macrobentos, quando da aquisição e coleta dos dados, foi possível concluir que grande parte dos dados existentes para a Região Metropolitana da Baixada Santista não se encontram preparados para serem utilizados em SIG. Estão dispersos em vários documentos, depositados em diferentes instituições. Não há a preocupação com a questão espacial e com a padronização no processo de constituição da informação, dificultando o acesso aos dados por parte dos pesquisadores, dos tomadores de decisão, entre outros. Essa realidade é a mesma encontrada para as demais regiões do Brasil, o que implica em constante retrabalho quando há a necessidade de estudar uma região ou um determinado tema.

Somado a isso, notou-se também que, quando se trabalha com dados ambientais coletados em pesquisas de campo, diferentes metodologias e terminologias são utilizadas. Este fato dificulta ainda mais a integração dos dados, que podem apresentar incompatibilidade metodológica inviabilizando sua comparação ou integração. Verificou-se que um modelo conceitual para integração de dados, serve também como um sinalizador da importância dessas questões metodológicas de

coleta, e da necessidade em se registrar certas informações, sem as quais o processo de integração pode ser inviabilizado.

Dados provenientes de amostras coletadas em determinado estudo podem ser incompatíveis com outros dados coletados para a mesma área por diferenças metodológicas. Essa constatação alerta não apenas para as questões de padronização, como também, para a importância do levantamento dos metadados ou a sua criação para os dados já existentes.

A pesquisa mostra que, apesar das condições supracitadas, é possível conceber-se um modelo de dados que abrange um largo espectro de diferentes fontes de informação e que o uso das técnicas OMT-G e de outros procedimentos computacionais tidos como padrões em modelagem de dados, podem ajudar na elaboração de produtos mais úteis aos tomadores de decisão e aos demais usuários de dados ambientais.

A importação de dados para o modelo concebido foi uma das etapas mais complexas, já que foi necessário realizar o tratamento dos dados, na maioria das vezes, de forma manual, antes de sua disponibilização em uma IDE. Além disso, a pesquisa também ilustra as etapas práticas para a construção de um Geoportal desde a criação dos metadados, geoportal, geoserviços e a navegação via *web*.

Também foi a intenção da pesquisa fomentar a discussão sobre a necessidade de padronização no levantamento, organização e disponibilização de dados, sejam estes vinculado à área ambiental, social ou econômica, para uso em portais baseados em IDEs. Tal discussão deve ser aprofundada em futuros estudos, primeiramente, para mostrar a importância em se definir padrões/metodologias com o intuito de evitar retrabalhos manuais, mas principalmente em evidenciar o ganho que uma solução que agregue diversas fontes de dados pode proporcionar a entidades, comunidade, entre outros. Uma vez homogeneizado, permitirá a construção de análises geoestatísticas, temporais, mapas temáticos, etc.

O Brasil está cada vez mais envolvido nos conceitos, metodologias e padrões de IDEs conforme é possível observar nos diversos geoportais que estão sendo lançados nos últimos anos. A discussão cada vez mais abrangente, englobando temas não diretamente ligados a cartografia básica (hidrografia, limites políticos, infraestrutura, etc.), como é o caso ambiental, se faz necessária exatamente para conscientizar os produtores de dados da importância de se compartilhar as

informações por meio de serviços, metodologia de levantamentos, etc., como salienta a própria Infraestrutura Nacional de Dados Espaciais (CONCAR, 2011).

6 CONSIDERAÇÕES FINAIS

A heterogeneidade dos dados espaciais atualmente não é uma exceção para os dados ambientes e esta realidade pode ser encontrada nas mais diversas fontes de dados, além do que, pode ser classificada como sintática (formatos distintos de dados), estrutural (diferenças de esquemas) e semântica (diferença de termos em contextos específicos). Informações espaciais são provenientes de diversas fontes de dados, que não obstante, se encontram em distintos formatos (normalmente associados à *softwares* ou propósitos específicos), escalas e precisão cartográfica.

Para possibilitar a interoperabilidade de dados entre os diversos sistemas e organizações, deve-se buscar maneiras para adequar as incompatibilidades sintáticas e semânticas. As incompatibilidades sintáticas foram fruto desta pesquisa que tentou ilustrar as dificuldades em se agregar diversas fontes espaciais e não espaciais em um modelo de dados e sua posterior publicação em um Geoportal.

Todavia, a questão semântica do dado não foi abordada diretamente nesta pesquisa e, sem dúvida, pode ser um ponto de partida para uma futura discussão de como resolver as questões de significado para os dados ambientais, principalmente na disponibilização e acesso da informação via Geoportal por entidades, cientistas, etc., que não tiveram contato prévio com a informação. Como visto, cada uma das fontes de dados utilizou uma denominação própria para documentar e disponibilizar o resultado de seus estudos.

Este tipo de situação pode causar uma dificuldade grande na descoberta e entendimento dos dados, uma vez que determinados termos são usados apenas uma comunidade científica muito restrita, ou mesmo, pode haver problemas de sinônimos (várias maneiras de encontrar o mesmo dado), homônimo (diferentes sentidos para o mesmo nome), questões de escala, entre outros. Os elementos espaciais, por exemplo, quando modelados por entidades distintas, podem conter interpretações diferentes sobre o mesmo dado. Esta diferença pode causar problemas não só relacionados a interoperabilidade dos dados, como também, na interpretação equivocada, na perda de qualidade, entre outros.

Uma das propostas para se resolver as questões semânticas é a utilização do compartilhamento de conceitos comuns e terminologias entre as diversas fontes de dados permitindo que usuários e programas (*softwares*) interpretem adequadamente

o que cada objeto representa. Estes conceitos e termos podem ser descritos e armazenados em estruturas ontológicas, auxiliando o usuário a encontrar a informação adequada, a interpretar melhor os resultados e identificar comportamentos e padrões nos mesmos.

Para uma pesquisa futura é muito importante que a abordagem semântica seja contemplada e como ela pode ser aplicada nas questões espaciais dos dados. Sabe-se que existem diversas iniciativas que podem servir como ponto de partida, como é o caso dos padrões FGDC, ISO (19115 e 19139), OGC com o uso da linguagem *Web Ontology Language (*OWL) e *Resource Description Framework (*RDF) que, por meio do uso de metadados, técnicas de modelagem de dados e padrões de intercâmbio de dados, poderiam garantir a compatibilidade de significados.

Além destes serviços semânticos, esta pesquisa também não explorou o uso de outros serviços geoespaciais, conforme abordado previamente no item 2.2.2 Geo Serviços, por não haver a necessidade de implementação, todavia, vale o destaque que, a construção de cada Geoportal é algo bem específico e, portanto, é imprescindível que se faça um levantamento de requisitos técnicos abordando as questões de arquitetura e funcionalidades. Por exemplo, devido a questões de dificuldade de rede, recomenda-se utilizar o formato OGC WMTS para garantir o desempenho da aplicação, já que tal serviço é responsável em pré-processar (*cache*) os dados e distribuí-los mais rapidamente quando requisitado pelo usuário final. Já o formato WFS-T, permite a edição espacial e alfanumérica de dados via geoserviços, mas, especificamente para a construção do Atlas Ambiental não houve a necessidade de sua inserção, pois não agregaria valor neste primeiro momento, uma vez que o principal objetivo era de organizar os dados e distribuí-los.

Outra discussão também pertinente é de verificar quão aderente são as atuais interfaces de um Geoportal para acesso por usuários leigos em geotecnologias, que necessitem da informação espacial para sua tomada de decisão. Este é um pouco da realidade de um Geoportal voltado a área ambiental, pois muitos especialistas em Macrobentos, Sedimentos, etc., não necessariamente possuem conhecimento no tema para rapidamente obterem informação pertinente ao seu interesse.

Diversas são as iniciativas na elaboração de Infraestrutura de Dados Espaciais, conforme foi discutido nesta pesquisa, porém alguns destes Geoportais são bem

complexos dificultando a compreensão do conteúdo até mesmo por especialistas da área de geotecnologia. Um dos grandes desafios de empresas desenvolvedoras de *softwares* é de conseguirem construir uma interface que seja rica em informações e, ao mesmo tempo, simples e intuitiva.

REFERÊNCIAS BIBLIOGRÁFICAS

AHN, I. Y.; KANG, Y. C.; CHOI, J. The influence of industrial effluents on intertidal benthic communities in Panweol, Kyeonggi Bay (Yellow Sea) on the west coast of Korea. **Marine Pollution Bulletin**, v. 30, p. 200-206, 1995.

AMARAL, A. C. Z.; MIGOTTO, A. E. Importância dos anelídeos poliquetas na alimentação da macrofauna demersal e epibentônica da região de Ubatuba. **Bolm Inst. oceanogr.**, São Paulo, v. 29, n. 2, p. 31-35, 1980.

AMARAL, A. C. Z.; MORGADO, E. H.; PARDO, E. V.; REIS, M. O. Estrutura da comunidade de poliquetos da zona entre marés em praias da Ilha de São Sebastião (SP). **Publ. Esp. Inst. Ocean.**, São Paulo, v. 11, p. 229-237, 1995.

ARAUJO, S. I.; SILVA, G. H.; MUEHE, D. **Mapas de sensibilidade ambiental a derrames de óleo: ambientes costeiros, estuarinos e fluviais**. Rio de Janeiro: CENPES/Petrobras, Ed. 2ª, 2007. 134 p.

AUSTRALIAN MARINE SPATIAL INFORMATION SYSTEM. **AMBIS**. Disponível em: <http://www.ga.gov.au/imf-amsis2>. Acesso em: 07 de Set. 2011.

AUSTRALIAN OCEAN DATA CENTRE JOINT FACILITY. **AODCJF**. Disponível em: <http://www.aodc.gov.au/index.php?id=19>. Acesso em: 07 de Set. 2011.

BARTLETT, D.; SMITH, J. **GIS for Coastal Management**. Boca Raton: CRC Press LLC, 2005. 315 p.

BERNARD, L; KANELLOPOULOS, I.; ANNONI, A.; SMITS, P. The European geoportal - one step toward the establishment of a European Spatial Data Infrasctructure. **Elsevier Comput., Environ. and Urban System** v. 29, p. 15-31, 2005.

BICEGO, M. C.; TANIGUCHI, S.; YOGUI, G. T.; MONTONE, R. C.; SILVA, D. A. M.; LOURENÇO, R. A.; MARTINS, C. C.; SASAKI, S. T.; PELLIZARI, V. H.; WEBER, R. R. Assessment of contamination by polychlorinated biphenyls and aliphatic and aromatic hydrovarbons in sediments of the Santos and São Vicente Estuary System, São Paulo - Brazil. **Marine Pollution Bulletin**, v. 52, 2006, p. 1804-1815.

BILYARD, G. R. The value of benthic infauna in marine pollution monitoring studies. **Marine Pollution Bulletin**, v. 18, n. 11, p. 581-585, 1987.

BORGES, K. A. V.; DAVIS, C. A.; LAENDER, A. H. F. OMT-G: An Object-Oriented Data Model for Geographic Applications. **Geoinformatica**, v. 5, n. 3, p. 221-260, 2001.

BORJA, A. MUXIKA, I. Guidelines for the use of AMBI (AZTI's marine biotic index) in the assessment of the benthic ecological quality. ***Marine Pollution Bulletin***, v. 50, p. 787-789, 2005.

BROWN, A. C.; MCLACHLAN, A. **Ecology of sandy shores**. Amsterdam: Elsevier, 1990. 327 p.

BRUCE, C.; KROON, F. **Metadata catalogue of digital spatial datasets for the Tully-Murray Catchments**. Atherton: CSIRO Sustainable Ecosystems, Tropical Forest Research Centre, 2006.

CÂMARA, G.; JUNIOR, P. O. L. GeoBR: intercâmbio sintático e semântico de dados espaciais. **Informática Pública**, Belo Horizonte, v. 4, n. 2, p. 251-281, 2002.

CÂMARA, G.; DAVIS, C. A.; VINHAS, L.; QUEIROZ, G. R.; CASANOVA, M. **Bancos de Dados Geográficos**. Curitiba: MundoGeo, 2005.

CANADIAN GEOSPATIAL DATA INFRASTRUCTURE CSDI. **Geoconnections**. Disponível em: <http://www.geoconnections.org/en/index.html>. Acesso em 23 de março de 2011.

CARVALHO, G. N.; GIANOTTI, M. A.; SARTOR, S. M.; QUINTANILHA, J. A. Modelagem para integração do conhecimento sobre macrobentos em Infraestrutura de Dados Espaciais. **Ambi-Agua**, Taubaté, v. 7, n. 2, p. 195-213, 2012. (http://dx.doi.org/10.4136/ambi-agua.774)

CHAN, T. O.; FEENEY, M.; RAJABIFARD, A.; WILLIAMSON, I. The Dynamic Nature of Spatial Data Infrastructures: A Method of Descriptive Classification. **Geomatica**, v. 55, n. 1, p. 451-462, 2001.

CHOI, J. W.; KOH, C. H. A study on the polychaete community in Kwangyang Bay, southern coast of Korea. **Journ. Ocean. Soc. Korea.**, v. 15, n. 2, p. 153-162, 1984.

COMPANHIA DOCAS DO ESTADO DE SÃO PAULO - CODESP. **Relatório Técnico do Estudo de Impacto Ambiental do Aprofundamento do Canal e Bacias de Evolução do Porto Organizado de Santos**. Fundação Ricardo Franco (FRF), 2008.

COMPANHIA DOCAS DO ESTADO DE SÃO PAULO - CODESP. **Relatório sobre monitoramento das atividades de dragagem na área de disposição oceânica do material dragado pela CODESP e suas regiões adjacentes (Município de Santos/SP)**. 2005 até 2010.

COMPANHIA DE TECNOLOGIA DE SANEAMENTO BÁSICO - CETESB. **Sistema Estuarino de Santos e São Vicente**. São Paulo: CETESB, 2001.

COMPANHIA DE TECNOLOGIA DE SANEAMENTO BÁSICO - CETESB. **Relatório de Qualidade das Águas Interiores do Estado de São Paulo: Balneabilidade das Praias.** São Paulo: CETESB, 2004.

COMPANHIA DE TECNOLOGIA DE SANEAMENTO BÁSICO - CETESB. **Relatório de qualidade das águas litorâneas do estado de São Paulo: balneabilidade das praias.** São Paulo: CETESB, 2007.

COMPANHIA DE TECNOLOGIA DE SANEAMENTO BÁSICO - CETESB. **Relatório de qualidade das águas litorâneas do estado de São Paulo**. São Paulo: CETESB, 2008.

COMPANHIA DE TECNOLOGIA DE SANEAMENTO BÁSICO - CETESB. **Relatório de qualidade das águas litorâneas do estado de São Paulo**. São Paulo: CETESB, 2009.

COMISSÃO NACIONAL DE CARTOGRAFIA - CONCAR. **Perfil de Metadados Geoespaciais do Brasil (Perfil MGB), Nov. de 2009.** Disponível em: <http://www.concar.ibge.gov.br/perfil_mgb.aspx>. Acesso em 23 de Março de 2011.

CROMPVOETS, J., BREGT, A., RAJABIFARD, A. AND WILLIAMSON, I. Assessing the worldwide developments of national spatial data clearinghouses. **International Journal of Geographical Information Science**, v. 18, n. 7, p. 665-689, 2004.

CUNHA-LIGNON, M.; MENGHINI, R. P.; SANTOS, L. C. M.; NIEMEYER-DINÓLA, C.; SCHAEFFER-NOVELI, Y. Estudo de Caso nos Manguezais do Estado de São Paulo (Brasil): Aplicação de Ferramentas com Diferentes Espaciais Espaço-Temporais. **Revista de Gestão Costeira Integrada**, v.9, 2009, p.79-91.

DAVIS, C. A.; OLIVEIRA, P. A. SIG Interoperável e Distribuído para Administrações Municipais de Grande Porte. **Informática Pública**, Belo Horizonte, v. 4, n. 1, p. 121-141, 2002.

DAVIS, C. A.; ALVES, L. L. Local Spatial Data Infrastructures Based on a Service Oriented Architecture. In: VII Brazilian Symposium on Geoinformatics, 20-23 nov. 2005, Campos do Jordão, **Proceedings** GeoInfo, 2005. p. 30-45.

EMPRESA BRASILEIRA DE TERMINAIS PORTUARIOS – EMBRAPORT. **EIA-RIMA sobre a construção do terminal portuário da Embraport**, v. III, 2003.

EMPRESA PAULISTA DE PLANEJAMENTO METROPOLITANO SA – EMPLASA. Metrópole em Dados. Disponível em: <http://www.emplasa.sp.gov.br/portalemplasa>. Acesso em: 12 de Nov. 2011.

EXTENSIBLE MARKUP LANGUAGE - XML. **XML Essentials**. Disponível em: <http://www.w3.org/standards/xml/core>. Acesso em: 08 de Fev. 2013.

FARLEY, J. INSPIRing Effective Use of Metadata: Open Standards, ETL and Digital Cities. In: GLOBAL SPATIAL DATA INFRASCTUCTURE WORLD CONFERENCE, workshop 1.8, jun. 2009, Rotterdam. **Proceedings** Rotterdam: GSDI Association, 2009.

FEGRAUS, E.; ANDELMAN, S. J; JONES, M. B; SCHILDHAUER, M. Maximizing the value of ecological data with structured Metadata: An introduction to Ecological Metadata Language (EML) and principles for metadata creation. **Bulletin of the Ecological Society of America**, v. 86, p. 158–168, 2005.

FERREIRA, J. A. **Estudos das associações de anelídeos Polychaeta da Baía de Santos e plataforma Continental Adjacente (SP, Brasil) e suas interrelações com parâmetros físicos e geoquímicos estruturadores**. 2008, 204f. Dissertação (Doutorado no Instituto Oceanográfico) - Universidade de São Paulo, São Paulo, 2008.

FERREIRA, J.; MIRANDA, M.; ABELHA, A.; MACHADA, J. O Processo ETL em Sistemas Data Warehouse. In II Simpósio de Informática, 09-10 set., 2010, Braga. **Proceedings** INForum 2010, p. 757-765.

FEDERAL GEOGRAPHIC DATA COMMITTEE - FGDC. **Content Standard for Digital Geospatial Metadata Workbook, Version 2.0,** 2000. Disponível em: <http://www.fgdc.gov>. Acesso em: 22 de Fev. 2011.

FIDALGO, R. N. **Uma Infra-estrutura para Integração de Modelos, Esquemas e Serviços Multidimensionais e Geográficos**. 2005, 165f. Dissertação (Doutorado em Ciência da Computação) – Universidade Federal de Pernambuco, Recife, 2005.

FME. **Product Description – FME SERVER**. Disponível em: <http://www.safe.com/fme/fme-technology/fme-server/capabilities/>. Acesso em: 13 de Abr. 2013.

GEHARDI, D. F. M.; CABRAL, A. P.; KLEIN, A. H. F.; MUEHE, D. C. E. H.; NOERBERG, M. A.; TESSLER, M. G.; SARTOR, S.M. Mapping the Sensitivy to oil spill of the Santos basin coastline, Southeast Brazil. **Braz. J. Aquat. Sci. Technol**, Santos. 12 (2): p. 11-31. 2008. Disponível em: <https://www6.univali.br/seer/index.php/bjast/article/viewFile/659/533>. Acesso em: 20 de Ago. 2011.

GEOBILBAO. **Sistema de Informação Geográfica de Bilbao**. Disponível em: <http://www.geobilbao.net/>. Acesso em: 07 de Jun. 2013.

GEOMEDIA SDI. **Product Description – GEOMEDIA SDI**. [2012?]. Disponível em: <http://geospatial.intergraph.com/products/geospatial-portal/Details.aspx/>. Acesso em: 18 de Set. 2012.

GIS4EU. **Provision of Interoperable dataset to open GI to EU communities**. Disponível em: <http://gis4eu.intergraph.pl/gis4eu>. Acesso em: 07 de Jun. 2013.

GUTTING, R. H. An introduction to spatial database systems. **The VLDB Journal**, v. 3, n. 4, p. 357-399, 1994.

GLOBAL SPATIAL DATA INFRASTRUCTURE ASSOCIATION - GSDI. **The SDI Cookbook, version 2.0**, 2004. Disponível em: <http://www.gsdi.org/gsdicookbookindex>. Acesso em: 20 de Jan. 2011.

GLOBAL SPATIAL DATA INFRASTRUCTURE ASSOCIATION - GSDI. GSDI 11 World Conference, 2009. Disponível em: <http://www.gsdi.org/gsdicookbookindex>. Acesso em: 20 de Jan. 2011.

HEITOR, S. R. **Composição e distribuição da macrofauna bentônica em áreas sob a influência da disposição oceânica de esgotos municipais da Baixada Santista e no Canal de São Sebastião**, 2002, 245f. Dissertação (Doutorado no Instituto Oceanográfico) – Universidade de São Paulo, São Paulo, 2002.

HENTTU, H.; IZARET, J.M.; POTERE, D. **Geospatial Services: A $1.6 Trillion Growth - Engine for the U.S Economy**. Boston: The Boston Consulting Group (BCG), Junho de 2012.

HINTZ, D. Case Studies in Data Harmonization for SDI Initiatives. In: GLOBAL SPATIAL DATA INFRASCTUCTURE WORLD CONFERENCE, case study 340, jun. 2009, Rotterdam. **Proceedings** Rotterdam: GSDI Association, 2009.

HYPERTEXT TRANSFER PROTOCOL - HTTP. Disponível em: <http://www.w3.org/Protocols/rfc2616/rfc2616-sec1.html#sec1.1>. Acesso em: 08 de Fev. 2013.

HJELMAGER, J.; MOELLERING, H.; COPPER, A.; DELGADO, T.; RAJABIFARD, A.; RAPANT, P.; DANKO, D.; HUETY, M.; LAURENT, D.; AALDERSE, H.; IWANIAK, A.; ABADA, P.; DURENO, U.; MARTYNENKO, A. An initial formal model for spatial data infrastructures. **International Journal of Geographic Information Science**, v. 22, p. 1295–1309, 2008.

IDEG. **Infraestrutura de Datos Espaciais de Galicia**. Disponível em: <http://sitga.xunta.es/sitganet/index.aspx?lang=gl>. Acesso em 07: de Junho de 2013.

INTERNATIONAL COASTAL ATLAS NETWORK. **ICAN**. Disponível em: <http://ican.science.oregonstate.edu/home> Acesso em: 18 de Ago. 2011.

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION - ISO. **Geographic information – Metadata**. Disponível em: <http://www.iso.org/iso/catalogue_detail.htm?csnumber=26020>. Acesso em: 22 de Jul. 2011.

LISBOA FILHO, J.; IOCHPE, C. Um estudo sobre Modelos Conceituais de Dados para Projeto de Bancos de Dados Geográficos. **Informática Pública**, Belo Horizonte, v.2, p. 67-90, 1999.

LISBOA FILHO, J. **Estruturação e Modelagem de Banco de Dados**. Salvador: Fator GIS, 2000 (Apostila do curso apresentada no GIS Brasil 2000).

LUNARDI, O. A; ISSMAEL, L. S; ALVES, P. D. V; CARVALHO, L. H. M. Aquisição da Geometria de Dados Geoespaciais para a Infra-Estrutura Nacional de Dados Espaciais (INDE). In: XIV SIMPÓSIO BRASILEIRO DE SENSORIAMENTO REMOTO, Natal, Brasil, 25-30 abril 2009. **Anais…** Belo Horizonte: SBSR, 2009, p. 1795-1802.

KENNISH, M. J. **Practical handbook of Estuarine and Marine Pollution**. Boca Raton: CRC Press, 1997. 524 p.

MAGUIRE, D.J; LONGLEY, P.A. The emergence of geoportals and their role in spatial data infrastructures. **Elsevier Comput., Environ. and Urban System** v. 29, p. 3-14, 2005.

MARTINS, C. C; MAHIQUES, M. M.; BICEGO, M. C.; FUKUMOTO, M. M.; MONTONE, R. C. Comparasion between anthropogenic hydrocarbons and magnetic suscentibility in sediment cores from the Santos Estuary, Brazil. **Mar. Pollut. Bull**. v. 54, 2007, p. 226-246.

MARINE ENVIRONMENTAL DATA & INFORMATION NETWORK MEDIN. **Marine Geoportal**. Disponível em: <http://www.oceannet.org>. Acesso em: 02 de Set. de 2011.

MARINE IRISH DIGITAL ATLAS. **Digital Atlas**. Disponível em: <http://mida.ucc.ie/pages/about.htm>. Acesso em: 15 de Out. de 2011.

MASSER, I.; RAJABIFARD, A.; WILLIAMSON, I. Spatially enabling governments through SDI implementation. **International Journal of Geographical Information Science**, v. 22, n. 1, p. 5-20, 2008.

MASSER, I. (2010). **Building European spatial data infrastructures**. ESRI Press, 2010. 108 p.

MCLACHLAN, A. Physical factors in benthic ecology: effects of changing sand size on beach fauna. **Marine Ecology Progress Series**, v. 131, p. 205-217, 1996.

MENGHINI, R. P.; CUNHA-LIGNON, M.; JUNIOR, C. C.; SCHAEFFER-NOVELI, Y. Análise temportal dos impactos antrópicos e da regenaração natual em manguezais da ilha Barnabé (Baixada Santista, SP, Brasil) obtiva através de fotografias aéreas. In: XII SIMPÓSIO BRASILEIRO DE SENSORIAMENTO REMOTO, Florianópolis, Brasil, 21-26 abr. 2007. **Anais**... Florianópolis: SBRS, 2009, p. 4037-4044.

MILLARD, K.; ATKINSON, R.; WOOLF, A.; LOWRY, R.; HAARING, P.; HERNANDEZ, F.; HAMRE, T.; PILLICH, D.; BOTHIEN, H.; BERGHE, W.V.; JOHNSON, I.; MATTHEWS, B.; DAHLIN, H.; FLEMMING, N.; VLADYMYROV, V.; MECKENZIE, B.; EDWARDS, A.; HOEVEN, W.V.D.; HARPHEN, Q.; ROSS, G. **Using XML Technology for Marine Data Exchange. A Position Paper of the Marine XML Initiative**, 2005. Disponível em: <http://www.ukmarine.org/MarineXML_Position_Paper_R5r0.pdf>. Acesso em: 02 de Abr. 2011.

MOUNT, R.; BRICHER, P. **Estuarine, Coastal and Marine (ECM) National Habitat Mapping Project: ECM National Habitat Map Series User Guide**. Australia: Departament of Climate Change, version 1.1, 2008. 156 p.

MURRAY, C. **Oracle® Spatial Developer's Guide 11g**. v. 1 (11.1), 2009. Disponível em: <http://download.oracle.com/docs/cd/B28359_01/appdev.111/b28400.pdf>. Acesso em: 18 de Jun. 2011.

NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION NOAA. **Digital Coast**. Disponível em: <http://www.csc.noaa.gov/digitalcoast/data>. Acesso em: 10 de Out. 2011.

NATURE SDI PLUS. **SDI implementation for nature conservation issues**. Disponível em: <http://www.nature-sdi.eu/>. Acesso em: 08 de Fev. 2013.

NYBAKKEN, J. W. **Marine Biology – An Ecological Approach**. São Francisco: Fifth Ed., 1995. 516 p.

OCEAN DATA AND INFORMATION NETWORK FOR AFRICA ODINAFRICA. **African Marine Atlas**. Disponível em: <http://193.191.134.9/odinafrica3/data.html> Acesso em: 12 de Jul. 2011.

OPEN GEOSPATIAL CONSORTIUM OGC. **Reference Model**. Disponível em: <http://www.opengeospatial.org/standards/orm>. Acesso em: 04 de Fev. 2011.

PEARSON, T. H.; ROSENBERG, R. Macrobenthic succession in relation to organic enrichment and pollution of the marine environment. **Oceanogr. Mar. Biol. Ann.** Rev., v. 16, p. 229-311, 1978.

OZANA, R; HORAKOVA, B. Actual State in developing GeonetWork opensource and Metadata Network standardization. In: GIS Ostrava, 2008. **Proceedings** Ostrava: GIS Ostrava, 2008.

PESTANA, G.; SILVA, M. M. A New Architecture for Integrated OLAP/GIS analysis based on a Geographical Data Warehouse. In: INTERNATIONAL CONFERENCE E-SOCIETY 03-06 Jun., 2003, Lisboa. **Proceedings** Lisboa: IADIS International Conference, 2003.

POCKLINGTON, P.; WELLS, P. G. Polychaetes: key taxa for marine environmental quality monitoring. **Marine Pollution Bulletin**, v. 24, n.12, p. 593-598, 1992.

POLAND GEOPORTAL. **GIS information shared with the public**. Disponível em: <http://maps.geoportal.gov.pl/webclient/>. Acesso em: 08 de Fev. 2013.

RAJABIFARD, A.; BINNS, A.; MASSER, I.; WILLIANSON, I. The role of the sub-national government and the private sector in future spatial data infrastructure. **International Journal of Geographical Information Science**, v. 20, n. 7, p. 727-741, 2006.

RAJABIFARD, A.; VAEZ, S.S.; WILLIAMSON, I.P. Building Seamless SDI to Facilitate Land and Marine Environments. In: TENTH INTERNATIONAL CONFERENCE FOR SPATIAL DATA INFRASCTRUCTURE, 25-29 Feb. 2008, Trinidad. **Proceedings** Trinidad: GSDI Association, 2008.

RAMOS, A. J. A. Coastal and Marine Resource Information System (CMARIS): A Spatial Approach towards sustainable ecoregional management. In: XXth ISPRS Congress, 12-23 Jul. 2004 **Proceeding** Istanbul, Turkey, 2004.

RIVERO, M. S.; ELÍAS, R.; VALLARINO, E. A. First survey of macrofauna in the Mar del Plata Harbor (Argentina), and the use of polychaetes as pollution indicators. **Rev. Biol. Mar. Oceanogr**., Valparaiso, v. 40, n.2, p. 101-108, 2005.

RIZZO, A. E.; AMARAL, A. C. Z. Temporal variation of annelids in the intertidal zone of beaches of the São Sebastião Channel, southern Brazil. **Journal of the Marine Biological Association of the United Kingdom**, v. 80, n.6, p. 1007-1017, 2000.

RIZZO, A. E.; AMARAL, A. C. Z.. Spatial distribution of annelids in the intertidal zone in São Sebastião Channel, Brazil. **Scientia Marina**, v. 65, n. 4, p. 323-331, 2001a.

RIZZO, A. E.; AMARAL, A. C. Z. Environmental variables and intertidal beach annelids of São Sebastião Channel (State of São Paulo, Brazil). **Revista de Biologia Tropical**, Costa Rica, v. 49, n. 3, p. 849-857, 2001b.

RODRÍGUEZ, I.; MONTOYA, I.; SÁNCHEZ, M.J; CARRENO, F. Geographic Information Systems applied to Intergrated Coastal Zone Management. **Elsevier B.V**, v. 107, p. 100-105, 2008.

RODRÍGUEZ, J. J.; WINDEVOXHEL, N. J. Análisis Regional de la situación de la Zona Marina Costera Centroamericana, Banco Inter-Americano de Desenvolvimento (BID). Washington, D.C. N°. ENV – 121, 1998. Disponível em: <http://idbdocs.iadb.org/wsdocs/getdocument.aspx?docnum=822685>

SAKKOPOULOS, E.; MILDORF, T.; CHARVAT, K.; BENZINA, I.; KRAUSE, K. Plan4All GeoPortal: Web of Spatial Data. In: WWW 2012 - European Projects Track, April, 2012, Lyon. **Proceedings** Lyon, 2012.

SANTOS, L. C. M.; CUNHA-LIGNON, M.; SCHAEFFER-NOVELI, Y. Impacto de Petróleo no manguezal do rio iriri (Baixada Santista, São Paulo): Diagnóstico da Cobertura Vegetal com base em fotografias aéreas digitais (1962-2003). In: VIII Congresso de Ecologia do Brasil, 23-28 set., 2007. **Anais...**Caxambu: VIICEB, 2007.

SARTOR, S. M.; DEGASPARI, F. A. A Balneabilidade das Praias de Santos – Discussão dos critérios Oficiais de Avaliação. In: XXVII Congresso Interamericano de Engenharia Sanitária e Ambiental. **Anais...**Porto Alegre: ABES, 2000.

SARTOR, S.M.; HANS, M. F. P.; SARTOR, L.; LEAO, A. L. Coastal Marine Mapping as an Ecosystem Based Management - the case study of the Baixada Santista Region - São Paulo, Brazil. **Journal of Coastal Research**, Gold Coast, v. 50, p. 1178-1182, 2007.

SARTOR, S. M.; PALM, A. S., PALM, L. SHINOHARA, E. J.,QUINTANILHA, J. A. Tools For Coastal Marine Ecosystem-Based Management And Geographically Referred Data Integration – The Santos Region Case Study. In: GLOBAL SPATIAL DATA INFRASCTUCTURE WORLD CONFERENCE, case study 344, jun. 2009, Rotterdam. **Proceedings** Rotterdam: GSDI Association, 2009.

SEELEY, B.; PARR, J.; EVANS, J.; LEAR, D. Establishing best practice for the documentation and dissemination of marine biological data. London: The Marine Biological Association, COWRIE Ltd, 2008. 99 p.

SHVAIKO, P.; IVANYUKOVICH, A.; VACCARI, L.; MALTESE, V.; FARAZI, F. A **Semantic Geo-catalogue implementation for a regional SDI**. Trento, Italia, 2010. Parecer técnico apresentadona conferência INSPIRE 2010.

SIGNA. **Sistema de Informação Geográfica Nacional da Espanha**. Disponível em: <http://signa.ign.es/signa/>. Acesso em: 20 de Fev. 2013.

SIMITSIS, A.; SKOUTAS, D.; CASTELLANOS, M. Representation of conceptual ETL designs in natural language using Semantic Web technology. **Data & Knowledge Engineering**, v. 69, p. 96–115, 2010.

SNELGROVE, P. V. R.; BUTMAN, C. A. Animal-sediment relationships revisited: cause versus effect. **Oceanogr. Mar. Biol. Annu. Rev.**, v. 32, p. 111-177, 1994.

STRAIN, L.; RAJABIFARD, A.; WILLIAMSON, I. P. Spatial Data Infrastructure to Facilitate Coastal Zone Management. In: COASTAL ZONE ASIA PACIFIC CONFERENCE. **Proceedings** Brisbane: CZAP, 2004.

STRAIN, L. **An SDI Model to Include the Marine Environment**, 2008, 134f. Dissertação (Mestrado em Engenharia de Geomática) - Universidade de Melbourne, Melbourbe, 2008.

TAIT, M.G. Implementing geoportals: applications of distributed GIS. **Elsevier Comput., Environ. and Urban System** v. 29, p. 33-47, 2005.

TOMMASI, L. R. **Consideração ecológica sobre o Sistema Estuarino de Santos, Estado de São Paulo**, 1979, 489f. 2v. Dissertação (Livre Docência no Instituto Oceanográfico) - Universidade de São Paulo, São Paulo,1979.

TOLVANEN, H.; KALLIOLA, R. A structure approach to geographical information in coastal research and management. **Ocean & Coastal Management**, v.51, p. 485-494, 2008.

TRUJILLO, J.; LUJÁN-MORA, S. A UML Based Approach for Modeling ETL Processes in Data Warehouses. 22nd International Conference on Conceptual Modeling (ER 2003), p. 307-320. **Proceedings** Lecture Notes in Computer Science 2813, Chicago (USA), October 13-16 2003.

UN ATLAS, Universal Atlas of the Oceans. Desenvolvido pelo UN-Oceans, 2002-2011. Apresenta textos, mapas, bancos de dados, etc. referentes ao tema costeiro e oceânico. Disponível em <http://www.oceansatlas.org>. Acesso em: 25 de Out. 2012.

UNIVERSAL DESCRIPTION DISCOVERY AND INTEGRATION PROTOCOL - UDDI. **Web Services Policy 1.5 - Framework**. Julho de 2006. Disponível em: <http://www.w3.org/TR/2006/WD-ws-policy-20060731/>. Acesso em 08 de Fev 2013.

UNIFORM RESOURCE IDENTIFIER - URI. **Architecture of the World Wide Web, Volume One**. Dezembro de 2004. Disponível em: <http://www.w3.org/TR/2006/WD-ws-policy-20060731/>. Acesso em: 08 de Fev. 2013.

U.S. Commission on Ocean Policy.Developing a National Ocean Policy US CoOP. **Developing a National Ocean Policy: Mid-term Report of the U.S. Commission on Ocean Policy**, sep. 2002. Disponível em: <http://www.oceancommission.gov/documents/midterm_report/midterm_report.html>. Acesso em: 18 de Ago. 2011.

VACCARI, L.; SHVAIKO P.; MARCHESE, M. An emergent semantics approach to semantic integration of geo-services and geo-metadata in Spatial Data Infrastructures. **International Journal of Spatial Data Infrastructures**, v. 3, 2008.

VAN BUREN, T. Electronic Watershed: Semantic GIS, Ontologies, and Organizing Knowledge. In: 22nd ANNUAL ESRI INTERNATIONAL USER CONFERENCE, 08-12 Jul. 2002. **Proceedings** San Diego, 2002.

VASSILIADIS, P.; SIMITSIS, A.; SKIADOPOULOS, S. Conceptual Modeling for ETL Processes. **Proceedings** of the 5th ACM international workshop on Data Warehousing and OLAP. USA, 2002.

WARWICK, R. M. A new method for detecting pollution effects on marine macrobenthic communities. **Marine Biology**, v. 92, p. 557-562, 1986.

WEB SERVICES DEFINITION LANGUAGE - WSDL. **Web Services Description Language (WSDL) 1.1**. Março de 2001. Disponível em: <http://www.w3.org/TR/wsdl>. Acesso em: 08 de Fev. 2013.

APENDICE A – DESCRIÇÃO DE METADADOS PARA (CODESP, 2008)

Title:

EIA-RIMA da Dragagem de Aprofundamento do canal e Bacias de evolução do Porto Organizado de Santos 2008

Abstract:

Os dados apresentados são parte de EIA-RIMA (Estudo de Impacto Ambiental - Relatório) elaborado como exigência do IBAMA (Instituto Brasileiro de Meio Ambiente), seguindo um Termo de Referência, para solicitar o licenciamento da dragagem de aprofundamento do Porto de Santos.

ISO 19115 - Metadata

Metadata information:

File identifier: a99b369a-b218-4f6c-90b4-a07f2e2816e6

Metadata language: portuguese

Metadata character set: utf8

Hierarchy level: dataset

Metadata point of contact:

Organization name: Universidade de São Paulo

Address (location of the responsible individual or organization):

Electronic mail address: gabrielncarvalho@gmail.com

Role: author

Metadata standard name: ISO 19115:2003/19139

Metadata standard version: 1.0

Identification information:

Citation:

Title: EIA-RIMA da Dragagem de Aprofundamento do canal e Bacias de evolução do Porto Organizado de Santos 2008

date: type of date: publication

Resource identifier: 324

Party responsible for the resource:

Abstract:

Os dados apresentados são parte de EIA-RIMA (Estudo de Impacto Ambiental - Relatório) elaborado como exigência do IBAMA (Instituto Brasileiro de Meio Ambiente), seguindo um Termo de Referência, para solicitar o licenciamento da dragagem de aprofundamento do Porto de Santos.

Status of the dataset: completed

Point of contact:

Organization name: CODESP - Companhia Docas do Estado de São Paulo

Phone: +55 (13) 3202-6565

Address (location of the responsible individual or organization):

Electronic mail address: sic@portodesantos.com.br

Role: point of contact

Maintenance and update frequency information: not planned

keywords:

Macrobentos

Codesp

Baixada Santista

Porto de Santos

type of keywords: theme

Method used to represent geographic information (spatial representation type):

vector

Resource character set: utf8

Topic category:

Resource extent (Extent information):

Geographic bounding box (extent):

West bound longitude: -49.8046875

East bound longitude: -42.24609375

South bound latitude: -26.575927734375

North bound latitude: -20.775146484375

Data quality information:

Scope of quality information: dataset

lineage: Dados foram construídos a partir de arquivo Word, transformado para planilha e posteriormente para vetor.

Report:

Reference system information:

Reference system identifier:

Code: SAD69 (EPSG: 4618) code space: EPSG

Information about distribution:

Distributor Transfer Options:

URL: http ://geoserver/W MS_ CODESP/service .svc/get