

InforM@r

IDE de apoio ao Projeto de Extensão da Plataforma Continental de Portugal - um olhar sobre a estrutura subjacente

BRANDÃO, Filipe; NAVARRO, Ana; CAMPOS, Aldino

A Estrutura de Missão para a Extensão da Plataforma Continental (EMEPC) é responsável, desde a sua criação em 2004, pela aquisição de um vasto conjunto de dados geográficos para assegurar o Projeto de Extensão da Plataforma Continental de Portugal (PEPC). Apesar de haver desde o início um sistema de informação para os dados hidrográficos, o mesmo não sucede para os dados respeitantes a outras disciplinas. Esta realidade implica diversos problemas no acesso e manutenção de dados que, pela sua natureza, exigiram um grande esforço para a sua recolha. O InforM@r pretende colmatar esta situação, através do desenvolvimento de uma Infra-estrutura de Dados Espaciais (IDE) que armazene, disponibilize e mantenha todos os dados recolhidos no âmbito do PEPC. Adicionalmente, como forma de assegurar o disposto na directiva 2007/2/EC do Parlamento Europeu e do Conselho de 14 de Março de 2007 (INSPIRE - *Infrastructure for Spatial Information in the European Community*), todos os dados e conjuntos de dados geográficos serão devidamente catalogados e publicados no Sistema Nacional de Informação Geográfica (SNIG).

PALAVRAS-CHAVE

InforM@r, IDE, Data Warehouse, INSPIRE, conjuntos e serviços de dados geográficos.

1. INTRODUÇÃO

Cientes da dificuldade que representa a organização de uma campanha científica no mar, os responsáveis pela EMEPC tiveram sempre como objectivo prioritário a máxima rentabilização dos recursos. Com efeito, em todas as campanhas realizadas no mar, organizadas no decurso dos seus trabalhos, vastas equipas de cientistas e membros da comunidade académica foram convidados a participar, por forma a alargar ao máximo a participação de outros projetos científicos, com aquisição de diversos tipos de dados marinhos.

Desde a criação da EMEPC, foi adoptada uma filosofia “*file based*” no que diz respeito ao armazenamento dos dados. Este facto significa que os mesmos estão guardados de acordo com o modo como foram adquiridos, num sistema de gestão de ficheiros, de uma forma não integrada. Este tipo de realidade não faz sentido numa época em que proliferam os sistemas de gestão e armazenamento de dados, sendo os mesmos cada vez mais fáceis de obter e operar e, ainda menos sentido faz, quando atendemos aos custos e esforço que a sua aquisição representou relativamente aos ganhos que deles se podem obter. Tornou-se pois necessário proceder à organização de todos os dados coligidos ao longo dos já mais de oito anos de existência da EMEPC, num sistema de informação, denominado InforM@r, que permitisse:

- Conhecer a realidade relacionada com os dados existentes na EMEPC;
- Disponibilizar os dados de forma expedita e de acordo com critérios específicos;
- Implementar políticas de segurança no acesso a dados e sua manutenção; e
- Integrar a EMEPC com outros organismos produtores de informação geográfica.

De acordo com o disposto na directiva INSPIRE, Portugal está obrigado a promover a normalização dos conjuntos de dados e serviços de dados geográficos que sejam da responsabilidade de organismos públicos. Esta realidade implica a adopção de regras comunitárias, denominadas de Disposições de Execução (desenhadas e estabelecidas com base em normas internacionais já existentes, nomeadamente a série 19100 das normas ISO (*International Organization for Standardization*), por forma a garantir a compatibilidade e usabilidade entre as Infra-estrutura de Dados Espaciais (IDE) nacionais dos vários Estados Membros. A EMEPC, como um organismo público detentor de dados e

conjuntos de dados geográficos, está sujeita a todas estas normas. Tal facto implicou que o trabalho a desenvolver incluisse uma componente puramente técnica (IDE) e outra relacionada com a normalização dos conjuntos de dados e serviços de dados geográficos da EMEPC.

2. ESTADO DA ARTE

2.1. INFRA-ESTRUTURAS DE DADOS ESPACIAIS

A informação espacial é, em termos muito simples, apenas outra forma de informação. É aquela (também conhecida como informação geográfica) que pode ser geograficamente referenciada, ou seja, que descreve uma determinada localização ou informação ligada a uma localização específica [1]. Esta dimensão geográfica da informação tem vindo a ganhar cada vez maior importância. Com efeito, cerca de 80% da informação utilizada no processo de tomada de decisão é informação geográfica [2] [3].

As diversas abordagens levadas a cabo pelas organizações e instituições na gestão de dados espaciais resulta em inconsistências. Estas podem ter diferentes causas, como por exemplo, o não cumprimento de normas e de especificações técnicas. Este tipo de situação é crítico, uma vez que em muitas situações a utilidade da informação geográfica está ligada a contextos de gestão de emergências, que não se compadecem com este tipo de entraves e requerem acessos rápidos e integração dos dados em ambientes “multi-plataforma”. Pode pois dizer-se que, em última instância, este problema deve ser encarado como um problema de integração de informação espacial “multi-fonte” [4].

É precisamente para eliminar este tipo de situação, que surgem as IDE, as quais podem ser definidas como estruturas ou sistemas que facilitam a troca e partilha de dados espaciais entre utilizadores. Estas caracterizam-se como estruturas subjacentes, frequentemente sob a forma de políticas, normas e redes de acesso, que permitem que dados sejam partilhados inter e intra organizações, entre estados ou entre países [4]. De acordo com a literatura, e entre os profissionais ligados a este tema, é aceite que as IDE não sejam consideradas como sistemas bem definidos, mas sim como ambientes complexos, multi-facetados e dinâmicos [5] [4]. Este dinamismo está subjacente aos seus próprios componentes uma vez que se tratam de sistemas ou estruturas, não podendo esses mesmos componentes ser encarados de um modo independente dos demais [6]. Rajabifard e Williamson [7] definiram as IDE, como estruturas de 5 nós em que a interação entre os atores (utilizadores) e os dados espaciais é gerida pelos componentes dinâmicos de uma IDE, que incluem as normas, as políticas e as redes de acesso, tal como esquematizado na figura .

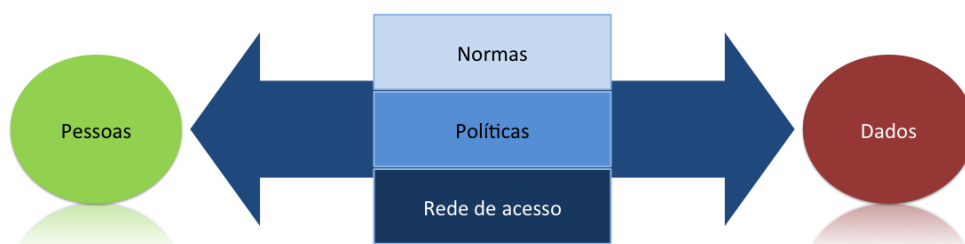


Figura 1 - Componentes de uma IDE (adaptado de [7])

2.2. VANTAGENS NA UTILIZAÇÃO DE IDE

A utilização deste tipo de sistemas (IDE) resulta em benefícios evidentes de diversas naturezas, sendo que os mais imediatos são o:

- **Conhecimento da realidade existente** - Agilização do processo de acesso a dados, quer a nível de tempo de resposta no acesso aos mesmos quer ao nível do conhecimento daqueles que se encontram disponíveis; e a

- **Otimização de recursos** - Diretamente relacionada com a anterior, a otimização de recursos permite a minimização dos esforços de aquisição de dados, uma vez que se pode sempre fazer um planeamento correto de modo a evitar a redundância de dados.

Estes dois aspectos são tanto mais evidentes, quanto maior atenção for dispensada à problemática da integração de dados multi-fonte, atrás referido. Mas para além destes benefícios que são imediatamente reconhecíveis, uma vez que são consequência direta da implementação de IDE, existem outras vantagens mais genéricas mas não menos importantes que devem ser mencionadas, tais como:

- Fomento do pensamento geográfico numa perspectiva transversal da sociedade;
- Estímulo à introdução da componente geográfica de dados já recolhidos;
- Introdução da componente geográfica em serviços já prestados;
- Benefícios económicos.

2.3. IDE DO AMBIENTE MARINHO

A necessidade de planeamento e gestão do ambiente marinho tem emergido nos últimos anos, surgindo a necessidade de inclusão da dimensão espacial dos dados associados a esse meio, uma vez que o planeamento espacial e a tomada de decisão não podem ficar limitados pela linha de costa [4]. Apesar de a utilização das IDE ter estado sempre mais relacionada com a resolução de problemas espaciais em ambiente terrestre [4], a adaptabilidade de cada um dos seus componentes (atrás referidos) ao ambiente marinho não parece ser de difícil resolução. De facto, as grandes diferenças que podem ser referidas são:

- **A natureza dos dados** - Os dados do ambiente marinho diferem dos dados terrestre sobretudo no que se refere ao processo de aquisição, uma vez que adquirir dados no mar acarreta dificuldades acrescidas. Esta realidade leva a que este processo seja, na maioria das vezes, realizado no âmbito de projectos nos quais todos os recursos disponíveis (pessoas e equipamentos) são explorados ao máximo, com o intuito de potenciar o tempo passado no mar; e
- **Redes de acesso** - Neste aspecto também algumas questões se levantam, uma vez que são as redes que permitem a troca de informação e, sendo as comunicações efectuadas através de redes, em terra e no mar as mesmas processam-se de maneira diferente. A comunicação de dados em alto mar apenas se pode processar via satélite, ao passo que em terra, as opções de comunicação são muito mais vastas.

Havendo IDE específicas para meios marinho e terrestre, a zona costeira acaba, não raras vezes, por ser agregada a um deles, originando fatalmente uma gestão menos eficiente da zona costeira. Estudos recentes sublinham a necessidade de haver uma abordagem holística e integrada da gestão destas áreas [8] [9], sendo este o principal objectivo da gestão integrada da zona costeira. Uma possível abordagem para a resolução deste tipo de situação seria a implementação de IDE únicas para terra e mar, que permitissem também a gestão integrada de todo o interface que separa os dois meios (figura 2).

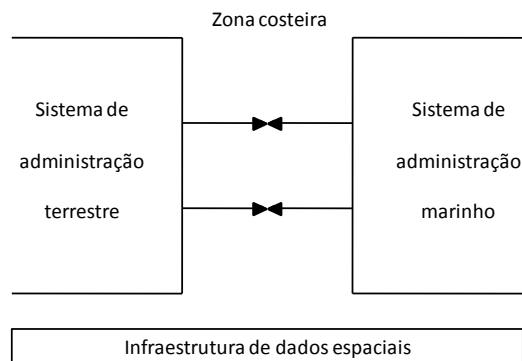


Figura 2 - Gestão integrada da zona costeira (adaptado de [10])

3. NORMALIZAÇÃO

Os custos inerentes à recolha de dados e o benefício que deles podem resultar, propiciam a procura de soluções que promovam o seu proveito e utilidade, investindo-se cada vez mais nas suas vertentes de pesquisa e disponibilização e em formatos que permitam que esses mesmos dados assumam o seu carácter transversal.

Portugal tem-se assumido, no domínio das IDE, como um país pioneiro, tendo criado há mais de duas décadas a sua própria infra-estrutura de informação geográfica, denominada de Sistema Nacional de Informação Geográfica (SNIG), pelo Decreto-Lei n.º 53/90, de 13 de Fevereiro. De acordo com o estabelecido na Directiva INSPIRE, a infra-estrutura nacional de informação geográfica portuguesa e as outras infra-estruturas regionais e locais ou temáticas existentes ou que venham a ser estabelecidas no país devem garantir:

- i) O armazenamento, a disponibilização e a manutenção de dados geográficos aos níveis mais adequados;
- ii) A combinação coerente de dados geográficos de diversas fontes no País e em toda a União Europeia, partilhados por diferentes utilizadores e aplicações;
- iii) A partilha de dados entre autoridades públicas, independentemente do seu nível de administração;
- iv) A disponibilização de dados geográficos em condições que não restrinjam de forma indevida a sua utilização generalizada; e
- v) A localização dos dados geográficos disponíveis, a avaliação da sua adequação para um determinado fim e o conhecimento das suas condições de utilização.

Houve pois a necessidade de se proceder à actualização do quadro legal do SNIG, adequando-o ao exposto na Directiva INSPIRE. Para tal foi estabelecido o Decreto - Lei n.º 180/2009, de 7 de Agosto [11], que cria o Registo Nacional de Informação Geográfica cujo âmbito de aplicação se refere, de acordo com o disposto no n.º 1, alínea b) do artigo 2.º, às autoridades públicas portuguesas com responsabilidades na produção e disponibilização de informação geográfica.

É neste contexto que a EMEPC, enquanto autoridade pública (artigo 3.º, n.º 1, alínea a) e i) do Decreto - Lei n.º 180/2009, de 7 de Agosto) detentora de conjuntos de dados geográficos que incidem sobre águas sob jurisdição nacional, em formato electrónico e que respeitam a temas enumerados nos anexos I, II e III da directiva INSPIRE, se vê na obrigação de implementar os

procedimentos de catalogação e disponibilização de informação geográfica, coligida ao longo dos anos da sua existência, sob a forma de metadados e de serviços de dados geográficos.

O Perfil Nacional de Metadados de Informação Geográfica (Perfil MIG) tem como objectivo principal clarificar aspectos ligados à implementação da produção, gestão e disseminação dos metadados em Portugal, de forma a assegurar a correta caracterização dos recursos geográficos e a sua harmonização com a infra-estrutura de informação geográfica portuguesa (SNIG) e europeia (INSPIRE). O Perfil MIG é composto por um subconjunto de metadados das normas ISO 19115 e ISO 19119 (extensão para serviços) cujos critérios de seleção se prendem com os requisitos da directiva INSPIRE, os elementos obrigatórios das normas ISO, a funcionalidade dos sistemas e a utilização frequente por parte dos técnicos [12]. Apesar de a utilização do Perfil MIG assegurar a conformidade não só com as normas ISO 19115 e 19119, mas também com a directiva INSPIRE, para a documentação dos dados do InforM@r optou-se pela utilização do perfil definido pela directiva INSPIRE, designado *INSPIRE Metadata Editor*. A opção pela utilização deste perfil de metadados prende-se com a obrigatoriedade de a EMEPC implementar o disposto na referida directiva, sendo que deste modo se assegura uma maior robustez para fazer face a eventuais alterações no próprio perfil.

4. DADOS DO INFORM@R

O InforM@r integra dados de várias naturezas que partilham o facto de todos terem sido adquiridos em campanhas oceanográficas. Estes dados, diretamente relacionados com o projecto de extensão da plataforma continental ou adquiridos numa lógica de aproveitamento dos recursos, constituem um valioso repositório uma vez que para além da quantidade e qualidade a eles associada, existe um custo de aquisição dos mesmos bastante elevado. Além do custo, existe ainda a dimensão temporal associada aos dados que também é por si só uma mais-valia inestimável porque permite a análise de séries temporais, importantes no estudo de vários fenómenos científicos. Entre estes dados incluem-se:

- Campanhas no mar - permitem ao utilizador saber informações acerca das campanhas realizadas, como por exemplo onde se realizou, qual o âmbito, plataforma, sistema de posicionamento, entre outros;
- ROV Luso - à semelhança dos dados das campanhas, fornecem uma série de informações respeitantes, neste caso, a mergulhos efectuados com o ROV Luso;
- Geologia - directamente usados para a justificação da proposta de extensão da plataforma continental, os dados de geologia são críticos no InforM@r. Podem ser recolhidos de diversos modos, sendo que a EMEPC possui amostras geológicas recolhidas com recurso a dragagens no fundo do oceano (por intermédio de dragas puxadas por um navio oceanográfico, a profundidades que em alguns casos excederam os 3000m) e através de mergulhos com o Luso;
- Oceanografia - incluem-se dados adquiridos com recurso a uma grande diversidade de equipamentos, que medem as propriedades físicas e químicas da água;
- Hidrografia - metadados de todos os levantamentos hidrográficos realizados no âmbito do PEPC, de modo permitir saber quais as áreas em que existem dados hidrográficos;
- Sísmica de reflexão - revestem-se de um carácter decisivo para o PEPC, uma vez que é a partir da sua interpretação que é possível conhecer a espessura da camada de sedimentos existente no leito do oceano, dependendo dessa informação um dos critério de extensão da plataforma continental; e

- Magnetismo e gravimetria - permitem o estudo das anomalias magnéticas e gravimétricas, através das quais é possível estabelecer a relação entre a origem de formações rochosas (como por exemplo, as ilhas do arquipélago dos Açores e os montes submarinos).

5. METODOLOGIA

O desenvolvimento de um sistema da natureza do InforM@r envolve o uso de tecnologias diversas sendo que neste trabalho se optou pela utilização exclusiva de *software* de código aberto (*open source*) e / ou gratuito. Entre estes, devem-se referir o *PostgreSQL*® como sistema de gestão de base de dados e o *PostGIS*® como suporte para dados geográficos. Para o processamento e manipulação de dados para carregamento, utiliza-se o pacote *Visual Studio Express 2012*®. No que se refere a *software* SIG *desktop* faz-se uso do *QuantumGIS*® ao passo que do lado do servidor utiliza-se o *Geoserver*®, para a componente de serviços de dados geográficos e a biblioteca *OpenLayers*® para o desenvolvimento da componente *webmap*.

5.1. DATA WAREHOUSE

A base de qualquer sistema de informação é um mecanismo de gestão e manutenção de dados, uma vez que são estes os seus pilares. Esta gestão é possível por intermédio de bases de dados que, como o próprio nome indica, “*são conjuntos partilhados de dados relacionados entre si de uma forma lógica e uma descrição desses próprios dados, desenhadas para corresponder às necessidades de informação e suportar as actividades de uma organização. Uma base de dados é implementada num sistema de gestão de bases de dados (SGBD), que é um programa informático que permite aos utilizadores a definição, criação, manipulação e administração de bases de dados*” [13].

Resultado da constante evolução e mudanças rápidas nos dias que correm, as organizações em todos os sectores têm a necessidade de efectuar análises de dados complexas de modo a suportar todos os processos de tomada de decisão. Contudo, as bases de dados tradicionais, denominadas de operacionais ou transaccionais, não satisfazem esses requisitos de análise de dados, uma vez que o seu principal objectivo é o suporte às operações diárias de uma organização, garantindo acessos simultâneos, manutenção de integridade dos dados e técnicas de recuperação desses mesmos dados. Este tipo de base de dados contém tipicamente dados detalhados e não dados históricos e, uma vez que são altamente normalizadas (sem redundância de dados), é comum terem problemas de desempenho na execução de interrogações complexas às bases de dados [13].

As *data warehouses* foram propostas por forma a melhor responder aos requisitos cada vez maiores dos utilizadores envolvidos nos processos de tomada de decisão [13]. Uma *data warehouse* é uma colecção de dados orientados, integrados, não voláteis e que variam no tempo, para suporte a decisões de gestão [14], sendo estas as características indicadas para descrever o tipo de dados recolhidos no decurso do PEPC, tendo estes ainda a particularidade de terem associada uma componente espacial. Os dados espaciais têm vindo a ser cada vez mais parte integrante de sistemas operacionais e analíticos em diversas áreas, tais como administração pública e redes de transporte, e a sua gestão e armazenamento têm vindo a ser feitas maioritariamente com recurso a bases de dados espaciais. A associação entre as bases de dados espaciais e *data warehouses* levou ao aparecimento das *data warehouses* espaciais, que permitem ao utilizador ter “o melhor dos dois mundos”. Por um lado as possibilidades associadas à análise de dados espaciais e, por outro, todos os mecanismos de segurança e integridade das bases de dados espaciais. Foram estas características que levaram ao desenvolvimento do InforM@r assente numa *data warehouse* espacial.

5.2. MODELO RELACIONAL

Uma das questões com que os profissionais ligados a *data warehouses* se têm de confrontar é o modelo segundo o qual o seu desenvolvimento é levado a cabo. Existem dois modelos básicos para o desenho de *data warehouses* que são vastamente considerados - o modelo relacional e o modelo

multidimensional. A grande diferença entre ambos está relacionada com modo como os próprios são desenhados. Por um lado, o modelo relacional é resultado de um modelo de dados, ou seja, a partir de um determinado nível de abstracção, ao invés do modelo multidimensional, que é desenhado de acordo com os requisitos dos utilizadores. Estas diferentes abordagens dão origem a diversas consequências no que toca às características dos modelos. A mais relevante dessas consequências está relacionada com o nível de adequabilidade do modelo. Sendo o modelo relacional desenhado em torno de um determinado nível de abstracção, o nível de flexibilidade que o mesmo oferece é muito elevado. Com esta flexibilidade perde-se ao nível da capacidade de resposta do modelo.

De um ponto de vista da arquitectura do sistema, estas diferenças significam que o modelo relacional é mais adequado como base para a implementação de *data warehouses*, uma vez que as mesmas devem suportar diferentes necessidades, de diferentes utilizadores que olham para os dados de maneiras distintas. Por outras palavras, uma *data warehouse* não é otimizada para o acesso de nenhum utilizador específico, mas sim para suportar um leque vasto de utilizadores [14]. É de referir ainda a este respeito que uma *data warehouse* implementada segundo o modelo relacional pode sem problema vir a ser utilizada como fonte de dados para alimentar um SGBD multidimensional [14], permitindo este facto encarar os dois paradigmas (relacional e multidimensional) como complementares ao invés de mutuamente exclusivos.

Acerca do já referido custo associado à velocidade de acesso a dados do modelo relacional, é de realçar o referido por [14] no que diz respeito ao grau de normalização. Por normalização entende-se o processo de avaliação e correção das estruturas das tabelas (por etapas a que se chamam formas normais) de modo a minimizar as redundâncias de dados e, como tal, reduzir os riscos de existência de anomalias nos dados [15]. Entre várias opções sugeridas para melhorar o desempenho de uma *data warehouse* implementada segundo o modelo relacional, é proposta a utilização deliberada de dados redundantes.

Na figura 3 é possível analisar o processo de acesso a dados, segundo uma determinada interrogação a *data warehouses* normalizadas (esquerda) e desnormalizadas (direita). Como é possível constatar, para o caso de sistemas normalizados, para uma mesma interrogação, é necessário o acesso a mais tabelas, quando comparado com um sistema desnormalizado.

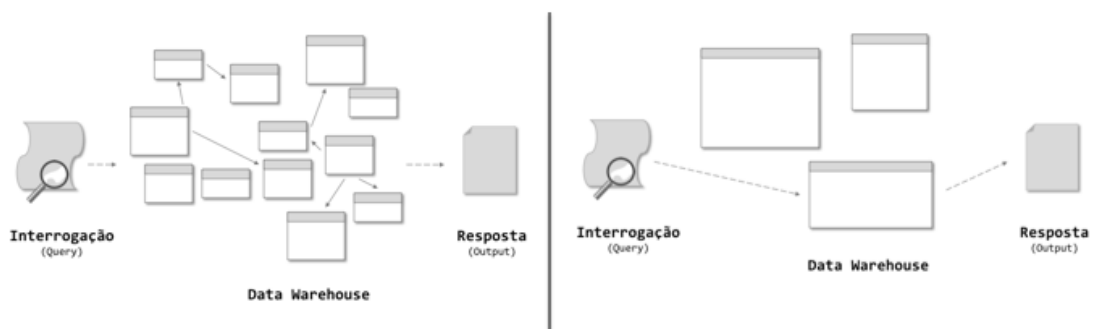


Figura 3 - Diferenças no acesso a dados entre *data warehouses* normalizadas e não normalizadas (adaptado de [14])

Esta abordagem (utilização deliberada de dados redundantes) foi utilizada no desenho do InforM@r para todos os atributos que, juntamente com os gestores de dados, foram identificados como altamente prováveis de serem frequentemente acedidos pelos diversos tipos de utilizadores a que o sistema se destina. Deste modo, aquilo que se poderia pôr em causa pela maior capacidade de armazenamento necessária, é largamente compensado com um aumento na velocidade de resposta do sistema.

Deve-se ainda referir que, durante os anos 80 foram levadas a cabo diversas pesquisas diretamente relacionadas com a intenção de se flexibilizar a assunção de que todas as tabelas de uma qualquer base de dados teriam obrigatoriamente de obedecer à primeira forma normal. Esta estabelece a obrigatoriedade de todos os atributos serem atômicos e mono-valorados. Toda a pesquisa efectuada resultou no estabelecimento do chamado modelo objecto/relacional, que preserva as fundações sobre as quais é construído o modelo relacional, acrescentando poder de modelação pela organização e armazenamento de dados em objetos. Este modelo permite a inclusão de tipos de dados complexos, resultando em maior expressividade e versatilidade do mesmo [13]. Sendo as geometrias um exemplo paradigmático, neste contexto de objecto, foi de uma forma natural que o desenvolvimento do InforM@r obedeceu ao princípios do modelo objecto/relacional.

5.3. DATA WAREHOUSE ESPACIAL

O facto de o InforM@r englobar dados espaciais implica a utilização de tecnologia específica para o efeito (*PostgreSQL® / PostGIS®*), que possui tipos de dados específicos de acordo com um dos padrões definidos pelo OGC (*Open Geospatial Consortium*) para o efeito - *Well Known Text (WKT)*. Para além desta característica essencial, as *data warehouses* espaciais permitem ainda atender às relações topológicas entre os vários elementos geográficos que as compõem. De acordo com [13], essas mesmas relações topológicas definem o modo como dois elementos espaciais se relacionam entre si, permitindo a realização de análises multidimensionais e espaciais complexas. Estando intimamente ligados aos conceitos de fronteira, exterior e interior, os operadores topológicos são variados, podendo-se referir, entre outros: a sobreposição, adjacência, a intersecção ou a inclusão.

6. RESULTADOS

Após o estudo dos vários paradigmas segundo os quais o InforM@r poderia ser desenvolvido, tornou-se possível iniciar a sua implementação. Este processo iniciou-se com a modelação de todo o sistema, processo descrito de seguida.

6.1. MODELAÇÃO DO INFORM@R

O primeiro passo no processo de modelação do InforM@r foi a definição do universo do discurso ou mini-mundo que, segundo [16], não é mais do que uma fracção de realidade representada por uma base de dados. Este processo iniciou-se com a convocação de reuniões com todos os gestores de dados das várias disciplinas envolvidas no PEPC, por forma a compreender em detalhe esses mesmos dados. Estes encontros permitiram perceber:

- Como se processa a aquisição de vários tipos de dados;
- Que tipo de parâmetros são medidos pelos vários equipamentos;
- Quais os parâmetros mais relevantes e que merecem maior atenção;
- Qual a natureza dos dados medidos (tipo de dados);
- Como pretendem os gestores que os dados sejam guardados (quais os atributos), uma vez que são eles que melhor percebem esses dados, que conhecem outros sistemas semelhantes e como neles os dados são guardados, estando este facto relacionado com o posterior processo de acesso aos mesmos; e
- Que tipo de interrogações são passíveis de ter maior ocorrência.

Foi ainda solicitado a todos os gestores de dados o seu contributo na elaboração de um dicionário de dados, que não é mais do que repositório de definições de todos os elementos e conceitos utilizados e manipulados pela organização e respectivos sistemas de informação e que incluem, entre outros,

dados, ficheiros, processos e entidades [17].

6.2. MODELO DE DADOS

Após a definição do mini-mundo em causa tornou-se possível definir o modelo de dados do InforM@r. Uma vez que é mais fácil examinar estruturas gráficas do que as suas descrições textuais, é aceite a preferência por parte de quem modela, de uma ferramenta na qual entidades e relacionamentos são representados graficamente. Como tal o Modelo de Entidades e Relacionamentos (MER) tornou-se largamente aceite como um padrão para a modelação de dados [15]. São conhecidas várias notações gráficas para a representação de sistemas utilizando o MER, entre as quais a notação original de *Chen* e as mais recentes “pé de corvo” e UML. Sendo todas válidas, pode-se referir a este respeito que:

- A notação de *Chen* favorece a modelação conceptual;
- A notação “pé de corvo” favorece uma abordagem virada para a implementação; e
- A notação UML (*Unified Modelling Language*) pode ser usada para modelação conceptual e para modelação mais virada para implementação [15].

Entre estas várias possibilidades, optou-se pela utilização da notação “pé de corvo”, pelo facto de ser mais orientada para a implementação. Esta característica é tanto mais importante pelo facto de se ter usado neste processo um *software*, que pode ser classificado como uma CASE (*Computer-Aided Software Engineering*) tool, que permite a geração de código DDL (*Data Definition Language*), para a implementação da base de dados propriamente dita.

Uma vez que o desenvolvimento do InforM@r foi feito de acordo com o modelo objeto/relacional, para a sua modelação utilizou-se o Modelo de Entidades e Relacionamentos Estendido (MERE) que, na prática, não é mais do que o convencional MER, com a capacidade de contemplação de tipos de dados complexos. O modelo integrou todas as informações necessárias para uma descrição da realidade tão detalhada quanto possível. Incluiu, forçosamente, também todos os elementos técnicos que permitiram a sua posterior implementação, nomeadamente:

- Qual o tipo de dados de cada atributo;
- A cardinalidade das relações entre tabelas;
- A identificação das chaves primárias e estrangeiras; e
- A explicitação de restrições que possam existir.

Com todos estes aspectos definidos, tornou-se possível o desenho do modelo propriamente dito, sendo que na figura 4 se apresenta, a título exemplificativo, a componente geológica do InforM@r.

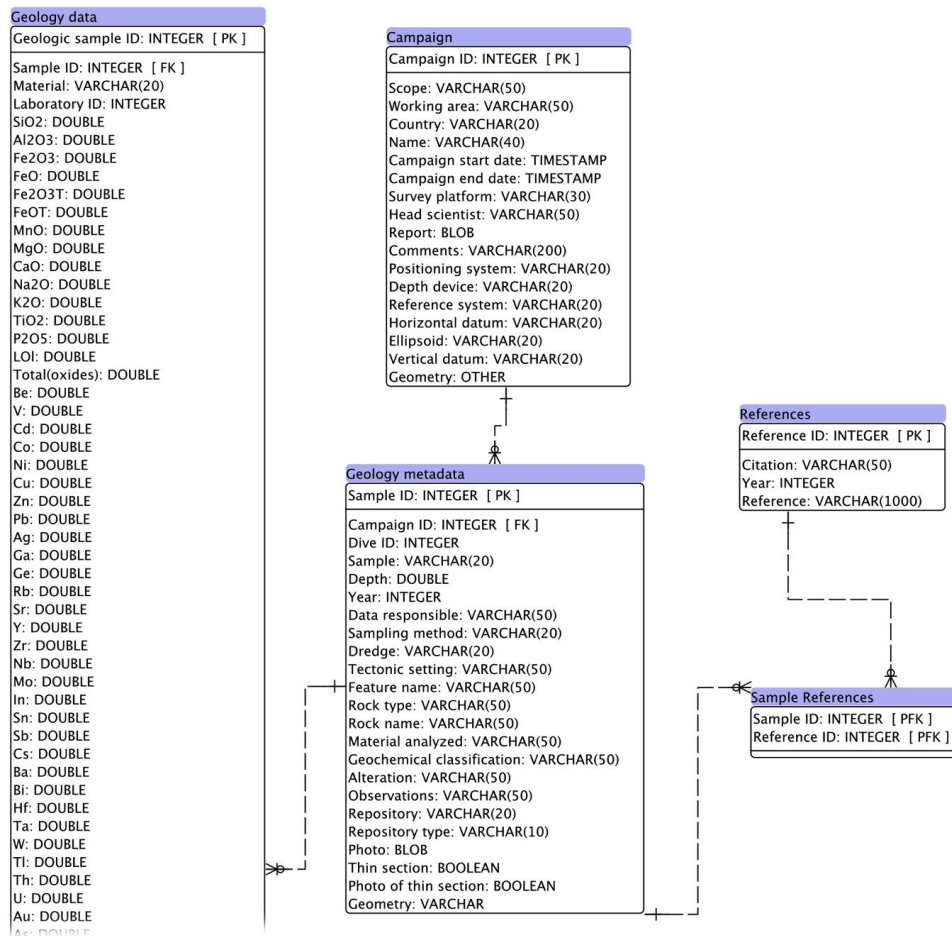


Figura 4 - Componente de geologia do InforM@r

Tendo-se optado por desenvolver o InforM@r segundo o modelo objeto/relacional foi desde logo aceite que em termos de desempenho poderia haver perdas relativamente a um desenvolvimento de acordo com o modelo multidimensional. Admitido este facto, foi importante pensar em métodos que pudessem minimizar esta situação que, tal como já foi referido, passaram sobretudo pela utilização deliberada de dados redundantes. Por exemplo, a informação relativa ao ano em que determinada amostra foi recolhida é essencial e poderia (e pode) ser derivada da tabela das campanhas, para uma qualquer amostra. Contudo, foi opção em muitas das tabelas de metadados incluir desde logo essa informação por forma restringir ao mínimo as tabelas envolvidas em cada interrogação ao InforM@r.

Outro aspecto relevante que deve ser mencionado foi a lógica de desenvolvimento subjacente a muitas das áreas científicas que compõem o InforM@r. Com efeito, em muitos casos foram definidas tabelas de metadados temáticos que tiveram dois objectivos primordiais:

- **Facilitar a pesquisa dos utilizadores** - cingindo os mesmos aos metadados acerca dos dados de uma determinada área temática do seu interesse, descartando de imediato todas as outras; e
- **Definição física de níveis de acesso** - pelo facto de funcionarem como nível máximo de acesso à maioria dos utilizadores do InforM@r, sendo que o acesso posterior a dados propriamente ditos se torna restrito.

6.3. MODELO FÍSICO

O modelo físico funcionou como o passo final na implementação do InforM@r. Nesta fase foram refinadas as informações do modelo lógico por via da definição de todos componentes que permitissem a construção da *data warehouse*. Esta fase de modelação utiliza já linguagens de modelação específicas, sendo que no caso do InforM@r recorreu-se à DDL, que é uma das categorias principais da SQL (*Structured Query Language*). O *software* utilizado para a modelação do InforM@r (*SQL Power Architect* ®) foi, como já referido, bastante útil nesta fase pela possibilidade de geração de todo o código necessário para a sua implementação, a partir do modelo desenhado (processo denominado de *forward engineering*), exemplificado na figura 5.

```
CREATE TABLE "informatv2".campaign (
  c_id INTEGER NOT NULL,
  scope VARCHAR(50) NOT NULL,
  working_area VARCHAR(50) NOT NULL,
  country VARCHAR(20) NOT NULL,
  name VARCHAR(40) NOT NULL,
  c_start TIMESTAMP NOT NULL,
  c_end TIMESTAMP NOT NULL,
  survey_platform VARCHAR(30) NOT NULL,
  head_scientist VARCHAR(50) NOT NULL,
  report BYTEA NOT NULL,
  comments VARCHAR(200) NOT NULL,
  pos_sys VARCHAR(20) NOT NULL,
  depth_dev VARCHAR(20) NOT NULL,
  ref_sys VARCHAR(20) NOT NULL,
  hdatum VARCHAR(20) NOT NULL,
  ellipsoid VARCHAR(20) NOT NULL,
  vdatum VARCHAR(20) NOT NULL,
  CONSTRAINT c_id PRIMARY KEY (c_id)
);

SELECT AddGeometryColumn('informatv2','campaign','geom', 4326, 'MULTILINESTRING', 2 );
```

Figura 5 - Exemplo de código DDL para a construção da tabela *campaign*

A implementação do modelo físico permitiu iniciar-se o carregamento da *data warehouse*. Neste processo foram desenvolvidos programas específicos para os diversos tipos de dados, de modo a garantir algumas tarefas de processamento necessárias, a sua conformidade com a estrutura das respectivas tabelas e a geração dos ficheiros finais, prontos para serem carregados.

7. DISCUSSÃO DE RESULTADOS E CONCLUSÕES

Todos os passos descritos neste trabalho permitiram, uma contextualização e tomada de consciência de todas as variáveis envolvidas num trabalho desta natureza. Com efeito, houve um processo intensivo e contínuo de estudo associado a temas como a normalização de dados e toda a legislação envolvida e, por outro lado, numa perspectiva mais técnica, as infra-estrutura de dados espaciais.

Após esta fase foi possível começar a tratar dos aspectos mais técnicos de todo o sistema começando-se, naturalmente, pelo seu alicerce, a base de dados. Neste aspecto foi essencial o conhecimento dos dados envolvidos e contacto com os especialistas responsáveis pela sua aquisição e processamento. Foi igualmente possível tomar a opção no que respeita à abordagem relacionada com o armazenamento dos dados propriamente dito, tendo-se optado, devido a diversos factores como a natureza dos dados e características dos mesmos, pelo estabelecimento de uma *data warehouse* espacial.

A *data warehouse* foi modelada tendo em conta as necessidades identificadas pelos especialistas de cada tipo de dados, tendo-se optado pela utilização do modelo relacional para o efeito. Esta opção está relacionada acima de tudo com a maior flexibilidade que esta abordagem apresenta face ao modelo multi-dimensional. Durante este processo foi igualmente identificada a necessidade de utilização deliberada de dados redundantes por forma a fazer face às eventuais perdas de rendimento do sistema.

Todo o caminho percorrido permitiu a implementação do modelo físico do InforM@r, estando-se

neste momento em pleno processo de ETL (*Extraction Transformation and Loading*) por forma a popular o sistema. Em simultâneo, iniciou-se já todo o processo de desenvolvimento da interface *webmap* do InforM@r, estando já em funcionamento na EMEPC um protótipo que permite aos seus elementos acederem a alguns serviços de mapas. Pretende-se também dar a possibilidade de os utilizadores realizarem algumas tarefas de geoprocessamento utilizando-se para este fim *Web Processing Services*. No que aos metadados diz respeito, está já em curso a criação das fichas de metadados que permitirão a ligação do InforM@r ao Sistema Nacional de Informação do Mar (SNIM@r) que, por sua vez, funcionará como o nó de ligação dos dados do ambiente marinho em Portugal ao SNIG.

8. REFERÊNCIAS

1. *Glossary*. [cited 2013 24 de Junho]; Available from: <http://www.anzlic.org.au/glossary#spatialinfo>.
2. Ryttersgaard, J. *Spatial Data Infrastructure - Developing Trends and Challenges*. in *International Conference on Spatial Information for Sustainable Development*. Nairobi, Kenya (2001).
3. Klinkenberg, B., *The true cost of spatial data in Canada*. *The Canadian Geographer* (2003).
4. Strain, L., A. Rajabifard, and I. Williamson, *Marine administration and spatial data infrastructure*. *Marine Policy*, **30**(4): p. 431-441 (2006).
5. L. Grus, J.C.A.K.B., *Spatial data infrastructures as complex adaptive systems*. *International Journal of Geographical Information Science*. **24**(3): p. 439-463 (2010).
6. Kok, B. and B. van Loenen, *How to assess the success of National Spatial Data Infrastructures?* *Computers, Environment and Urban Systems*. **29**(6): p. 699-717 (2005).
7. Rajabifard, A., and Williamson. I., *Spatial Data Infrastructures: Concept, SDI Hierarchy and Future directions*. (2001).
8. Wescott, G., *The Theory and Practice of Coastal Area Planning: Linking Strategic Planning to Local Communities*. *Coastal Management*, **32**(1): p. 95-100 (2010).
9. Darius Bartlett, R.L., Maria Carmen Garriga, *Marine and Coastal Data Infrastructures: a missing piece in the SDI puzzle?*, in *Global Spatial Data Infrastructure 7th Conference*: Bangalore, India (2004).
10. 3, U.-P.W.G., *Report on the international workshop on administrating the marine environment—the spatial dimension*. Kuala Lumpur (2004).
11. *Decreto-Lei n.º 180/2009 de 7 de Agosto*. MINISTÉRIO DO AMBIENTE, DO ORDENAMENTO DO TERRITÓRIO E DO DESENVOLVIMENTO REGIONAL: Diário da República, 1.ª série — N.º 152 — 7 de Agosto de 2009.
12. *Perfil Nacional de Metadados para Informação Geográfica (Perfil MIG)*. SNIG - Sistema Nacional de Informação Geográfica.
13. Elzbieta Malinowski, E.Z., *Advanced DataWarehouse Design. From Conventional to Spatial and Temporal Applications*: Springer-Verlag Berlin Heidelberg (2008).
14. Inmon, W.H., *Building the Data Warehouse*. Fourth ed. (2005).
15. Carlos Coronel, S.M., and Peter Rob, *Database Systems: Design, Implementation, and Management, Ninth Edition*. Course Technology, USA: Cengage Learning (2011).
16. Navathe, R.E.S.B., *Fundamentals of Database Systems*. 3rd ed.: Addison-Wesley (2000).
17. Videira, A.S.e.C., *UML, Metodologias e Ferramentas CASE*. 2nd ed. Vol. I : Centro Atlântico (2000).

AUTORES

Filipe BRANDÃO
filipe.brandao@emepc.mam.gov.pt
v.pt
 EMEPC
 Gabinete de geomática

Ana NAVARRO
acferreira@fc.ul.pt
 IDL - FCUL
 DEPARTAMENTO DE
 ENGENHARIA GEOGRÁFICA,
 GEOFÍSICA E ENERGIA

Aldino CAMPOS
Aldino.campos@emepc.mam.gov.pt
v.pt
 EMEPC
 Responsável

