INSTITUTO SUPERIOR DE ENGENHARIA DE LISBOA

Departamento de Engenharia de Eletrónica e Telecomunicações e de Computadores

Mestrado em Engenharia Informática e de Computadores

Análise estatística de informação georreferenciada

André Pereira de Matos
(Licenciado em Engenharia Informática e de Computadores)

Trabalho de projeto para obtenção do grau de Mestre em Engenharia Informática e de Computadores

Orientadores:

Mestre Lara Cristina de Paiva Lourenço Santos

Licenciado José Luís Falcão Cascalheira

Júri:

Presidente: Doutor Helder Jorge Pinheiro Pita

Vogais:

Mestre Fernando Manuel Gomes de Sousa

Mestre Nuno Miguel Soares Datia

Licenciado José Luís Falcão Cascalheira

Mestre Lara Cristina de Paiva Lourenço Santos

Novembro de 2012

INSTITUTO SUPERIOR DE ENGENHARIA DE LISBOA

Departamento de Engenharia de Eletrónica e Telecomunicações e de Computadores

Mestrado em Engenharia Informática e de Computadores

Análise estatística de informação georreferenciada

André Pereira de Matos
(Licenciado em Engenharia Informática e de Computadores)

Trabalho de projeto para obtenção do grau de Mestre em Engenharia Informática e de Computadores

Orientadores:

Mestre Lara Cristina de Paiva Lourenço Santos

Licenciado José Luís Falcão Cascalheira

Júri:

Presidente: Doutor Helder Jorge Pinheiro Pita

Vogais:

Mestre Fernando Manuel Gomes de Sousa

Mestre Nuno Miguel Soares Datia

Licenciado José Luís Falcão Cascalheira

Mestre Lara Cristina de Paiva Lourenço Santos

Novembro de 2012

[DECLARAÇÕES]

Declaro que este trabalho de projeto é o resultado da minha investigação pessoal e independente. O seu conteúdo é original e todas as fontes consultadas estão devidamente mencionadas no texto, nas notas e na bibliografia.

O candidato,

(André Pereira de Matos)

Lisboa, ... de ..... de .....

Declaro que esta dissertação /trabalho de projeto/relatório de estágio se encontra em condições de ser apresentada a provas públicas.

O(A) orientador(a),

(Lara Cristina de Paiva Lourenço dos Santos)

O(A) orientador(a),

(José Luís Falcão Cascalheira)

Lisboa, ... de ..... de .....

Agradecimentos

Tenho de agradecer aos Engenheiros Luís Falcão e Lara Santos pela orientação e disponibilidade ao longo deste último ano de projeto, tendo desempenhado um papel fundamental para a conclusão do mesmo. Ao SAPO pela colaboração e compreensão nas datas estabelecidas para as entregas, e ao Instituto Nacional de Estatística pela disponibilização dos dados.

Um agradecimento especial aos familiares e amigos pela compreensão e amizade manifestados apesar da falta de atenção e ausências.

Resumo

Com o volume de informação disponível para análise e tomada de decisões, a capacidade de sumarização da informação, facilitando o processo de análise e tomada de decisões em tempo útil, assume um papel fundamental nos dias correntes. Com o crescimento de sistemas de exploração geográfica dos dados, esta tese propõe e implementa uma solução para a exploração e sintetização da informação geográfica.

A exploração geográfica dos dados é efetuada com recurso a mapas temáticos, gráficos e tabelas *pivot*. Para a exploração geográfica dos dados, são estudadas várias técnicas de geração de mapas temáticos, tecnologias associadas e conceitos de sistemas de informação geográfica. O foco do projeto recai sobre sumarização de informação estatística georreferenciada, através de mapas coropletos com suporte em mapas dinâmicos na *Web*.

Como prova de conceito, são utilizados os dados disponibilizados pelo Instituto Nacional de Estatística (INE), através de um *Web Service*. Estes dados, são integrados no sistema possibilitando a representação de cada um dos indicadores através das ferramentas de sumarização da informação implementadas: Gráficos, Tabelas *pivot* e Mapas temáticos.

Palavras Chave

Análise estatística espacial, análise estatística, georreferenciação, mapas temáticos, sistemas de informação geográfica, gráficos, informação, sintetização, análise

Abstract

With the amount of information available for analysis and decision making, the ability of summarizing the information, making the process of analysis and decision making timely plays a key role in the current days. With the growth of geographic information systems, this thesis proposes and implements a solution to explore and summarize geographic information.

The geographical analysis of the data is done using thematic maps, graphs and pivot tables. For the analysis of geographic data are studied various techniques of thematic mapping, associated technologies and concepts of geographic information systems. The project focus is on summarizing spatial statistical information through choropheth maps, supported by dynamic maps in Web.

As a proof of concept are used data from Instituto Nacional de Estatística (INE) available from a Web Service. Indicators provided by the service are integrated into the systems and can be summarized through: charts, pivot tables and thematic maps.

Keywords

Spatial statistical analysis, statistical analysis, geocoding, thematic maps, geographic information systems, charts, information, synthesizing, analysis

Conteúdo

Lista de Figuras

Lista de Tabelas

Listagens de código

1

Introdução

O processo de tomada de decisões tem por base o conhecimento e a informação disponível. No entanto, nos últimos quarenta anos assistiu-se a um crescimento exponencial da informação disponível para análise, como suporte à tomada de decisões, tendo hoje em dia um ritmo de crescimento em que duplica a cada três anos. [1]

1.1 Motivação e contexto

O grande volume de informação disponível torna essencial a sua sumarização, tendo a análise estatística um papel fundamental no processo de tomada de decisões. [1][6]

O processo de tomada de decisões é ilustrado na Figura 1.1.
Numa primeira fase, o analista deverá identificar todas as alternativas à decisão que pretende tomar. Depois de identificadas as alternativas, deverá assinalar as implicações de cada uma das alternativas. A decisão resulta da comparação das implicações identificadas, considerando os fins, as pressões, as restrições e o objetivo final.[1]

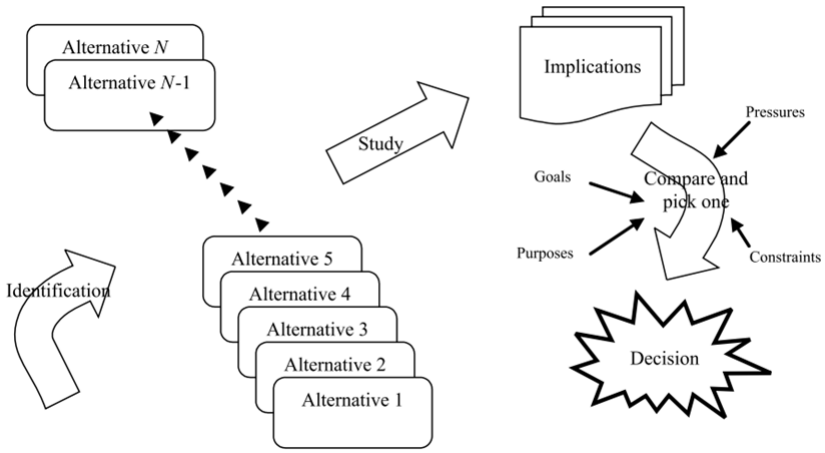

Figura 1.1. Processo de tomada de decisões (De Handbook on Decision Support System - Part1 [1])

No entanto, existem alguns obstáculos que dificultam o processo de tomada de decisões, nomeadamente na identificação das alternativas:

- Grande quantidade de informação;
- Informação dispersa;
- Informação desorganizada;
- Nem sempre se encontra em formato digital;
- Não se encontra acessível para todos.

Com o crescimento dos sistemas de informação geográfica (SIG), começou a explorar-se na análise estatística a componente geográfica dos dados, aparecendo um novo ramo do conhecimento que dá pelo nome de estatística geográfica. A análise estatística geográfica, apoiou-se nos sistemas SIG de forma a aproveitar as ferramentas de construção de mapas:

- Na geração de mapas temáticos;
- Na visualização da informação. [7]

De forma a facilitar o processo de tomada de decisões, é essencial a existência de ferramentas que permitam ultrapassar os obstáculos identificados [6], sendo introduzida a exploração geográfica dos dados. A exploração geográfica, possibilita ao analista ter uma visão abrangente dos dados e da forma como se encontram distribuídos sobre o território.

Em vários negócios, a análise geográfica da informação é fundamental para o processo de tomada de decisão. Por exemplo, uma entidade bancária poderá decidir não conceder um crédito para aquisição de um imóvel se, na zona, a taxa de incumprimento for elevada e a densidade populacional for baixa. O cruzamento destes dois indicadores poderá indiciar que a probabilidade do novo crédito entrar em incumprimento é elevada. Este é apenas um exemplo ilustrativo, na prática a concessão de crédito deverá obedecer a critérios mais rigorosos.

1.2 Enquadramento

Os gráficos de barras e de linhas foram concebidos em 1796 por William Playfair, tendo posteriormente concebido os gráficos circulares em 1801 [8]. Os mapas temáticos são um tipo de gráfico desenhados para apresentar um determinado tema interligado com uma área geográfica. As primeiras explorações de mapas como uma forma de representação estatística remontam ao século XVIII, tendo-se começado aí a tentar representar mais do que a posição geográfica sobre o mapa [9].

A representação gráfica da informação, sob a forma de gráficos e mapas, permite sintetizar a mensagem que se pretende transmitir e ajudam na visualização e compreensão de uma realidade que pode escapar à nossa compreensão imediata. [10]

Para perceção gráfica na análise estatística, destacam-se três áreas [11]:

- Computação - A existência de ferramentas que suportem a análise gráfica de informação assume um papel importante na rápida perceção e extração de conclusões;
- Métodos - Os métodos consistem na escolha da informação a representar de forma a ajudar o analista no estudo dos dados. Esta área poderá ser interativa, permitindo ao analista a exploração dos dados interagindo com o sistema de forma a selecionar a informação a visualizar.
- Construção - Decidida a informação a apresentar é necessário construir a representação gráfica. Nesta área deverão ser definidos: a forma de codificação da informação, as escalas, as cores e as texturas tendo sempre como principal objetivo facilitar a leitura e perceção do gráfico.

Para a análise geográfica dos dados, são usados mapas temáticos, sendo estes utilizados de forma a servir os seguintes objetivos [9]:

- Apresentar informação específica sobre determinados locais;
- Extração de padrões espaciais sobre a informação;
- Comparação relativa entre as várias regiões apresentadas de forma a extrair padrões.

A preparação dos dados, bem como a sua representação gráfica, deverá ser cuidadosamente tratada e estruturada, uma vez que existe a possibilidade da má interpretação dos dados. Este é um risco que será sempre inerente à análise estatística [10]. No entanto, esse risco será tanto menor quanto maior a qualidade da representação gráfica, devendo esta ser a mais adequada para cada um dos casos de análise.

1.3 Objetivo e visão

Este projeto visa o desenvolvimento de uma plataforma para análise estatística sobre informação georreferenciada, permitindo a integração e exploração de indicadores de diferentes fontes de dados.

Como principal objetivo, pretende-se resolver os obstáculos identificados na secção 1.1, sendo para isso necessária a integração de informação estatística proveniente de diferentes fontes de dados. A integração e centralização da informação, visa facilitar o trabalho do analista na construção de relatórios estatísticos e cruzamento de dados. Ou seja, pretende-se, por exemplo, que seja possível um analista construir um relatório com informação proveniente do Instituto Nacional de Estatística (INE) e um indicador fornecido pela Direção Geral de Energia e Geologia (DGEG), permitindo assim o cruzamento da informação.

É objetivo sintetizar a informação graficamente sob a forma de:

- Mapas temáticos, tirando partido da componente geográfica dos dados. Pretende-se apresentar uma visão global da distribuição dos dados sobre o território;
- Gráficos circulares e gráficos de barras, permitindo uma análise comparativa dos atributos da dimensão projetada, tendo em conta as condições de filtragem;
- Tabelas *pivot*, permitindo a visualização dos valores concretos associados às dimensões projetadas, sobre as condições de filtragem.

A Figura 1.2, apresenta a visão da ferramenta desenvolvida no âmbito do projeto para a construção de relatórios estatísticos recorrendo a indicadores de fontes de informação diferentes. Este relatório é com-

posto por diferentes componentes de sintetização de informação de forma a permitir ao analista obter respostas ou extrair padrões:

- Sobre o mapa, é apresentada uma camada temática de informação representando o ganho médio mensal em Portugal por regiões;
- Nos gráficos circulares são representados os indicadores do ganho médio mensal e a taxa de desemprego, estando a informação em ambos os casos filtrada para Lisboa e a projetar os atributos da dimensão género;
- No gráfico de barras é apresentado o indicador da densidade populacional, estando a informação filtrada em Lisboa e são projetados os atributos da dimensão género;
- Na tabela *pivot* é representado o indicador do ganho médio mensal, para Lisboa e Porto, possibilitando a comparação através de valores absolutos das duas regiões.

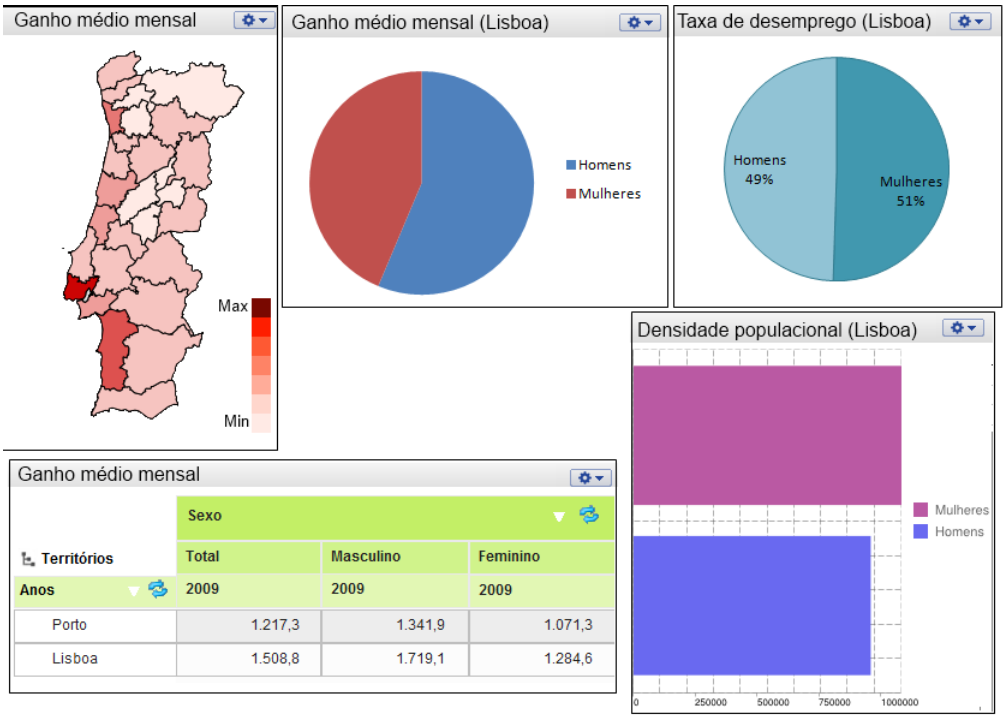

Figura 1.2. Visão da ferramenta de análise estatística de informação georreferenciada.

Com este relatório estatístico, é possível concluir através da visualização do mapa temático que em Lisboa os salários são mais elevados. Através do cruzamento dos dados dos diferentes indicadores é possível

concluir que as Mulheres têm um salário inferior aos Homens na região de Lisboa.

O relatório representado na Figura 1.2 contém indicadores das seguintes fontes:

- INE, com os indicadores do ganho médio mensal e densidade populacional;
- Ministério do trabalho e da segurança social, com o indicador da taxa de desemprego.

De forma a possibilitar a integração da ferramenta em aplicações empresariais que pretendam explorar informação estatística, será disponibilizada uma camada de serviços e uma *framework* cliente, facilitando a integração da plataforma em aplicações *Web*.

1.4 Trabalhos relacionados

Nos últimos anos têm aparecido alguns projetos de apoio à análise estatística, sendo que os seguintes exploram a componente geográfica da informação. Nesta secção, apresentam-se alguns desses projetos.

1.4.1 SAPO Mapas

O SAPO Mapas[1], disponibiliza uma ferramenta de análise estatística que explora a componente geográfica dos dados disponibilizados pelo INE. A informação é representada sobre o mapa. No entanto, esta ferramenta apresenta algumas limitações:

i) Apenas permite a visualização de dados sobre as regiões previamente estabelecidas, distrito, concelho, freguesia da CAOP (Carta Administrativa Oficial de Portugal) 2010 e dos NUTS (Nomenclatura de unidades territoriais) do INE;
ii) Apenas possibilita a exploração de indicadores do INE.

1.4.2 Pordata

O Pordata[2] é um repositório de indicadores sobre Portugal e Europa obtidos através de entidades oficiais com competências de produção de informação nas áreas respetivas [3]. Este portal pretende agregar

[1] http://mapas.sapo.pt

[2] http://www.pordata.pt

[3] http://www.pordata.pt/Sobre+a+Pordata

indicadores sobre diferentes temas e disponibilizá-los de forma gratuita para consulta, com o objetivo de facilitar o processo de tomada de decisões, ultrapassando os obstáculos identificados na secção 1.1.

Posteriormente ao início do desenvolvimento deste projeto, para além da exploração dos dados através de gráficos e tabelas *pivot*, o Pordata adicionou ao sistema a possibilidade de efetuar uma exploração geográfica dos dados através de mapas temáticos.

Devido à aposta na exploração geográfica dos dados, o Pordata é um sistema semelhante ao sistema implementado neste projeto.

1.4.3 ArcGIS

O ArcGIS[4] é uma ferramenta desenvolvida pela ESRI (*Economic and Social Research Institute*)[5] com o objetivo de facilitar o processo de análise de dados e tomada de decisões. A informação a analisar poderá estar armazenada numa base de dados local ou nos *Cloud Services* da ESRI, sendo os mapas temáticos um dos tipos de representação suportados.[12]

Os dados são disponibilizados através de um serviço na componente servidor do ArcGIS, sendo posteriormente consumidos pelas aplicações Móveis, *Desktop* e *Web* do ArcGIS.

Comparativamente ao projeto implementado, o ArcGIS apresenta algumas semelhanças na análise dos dados, nomeadamente na sua exploração geográfica. No entanto, o ArcGIS não integra dados de diferentes fontes de informação para a exploração dos dados, sendo apenas utilizada a informação armazenada numa base de dados local ou no sistema de armazenamento da ESRI (*Cloud Services*).

1.4.4 Excel

A Microsoft incluiu na ferramenta Excel a funcionalidade de análise estatística de informação, possibilitando ao utilizador análises de dados através de gráficos ou tabelas *pivot*.

De forma a obter dados para análise, é possível ligar o Excel às seguintes fontes de dados:

---

[4] http://www.esri.com/software/arcgis/features

[5] http://www.esri.com/

- SQL Server;
- Analysis Services (sistema OLAP);
- A um ficheiro XML;
- A uma tabela definida no ficheiro Excel.

Recentemente, através de uma parceria com a ESRI, foi anunciado o suporte no Office de análise estatística geográfica[6][7]. A informação presente numa folha de cálculo do Excel, é georreferenciada (através de uma morada) sendo depois representada sobre um mapa temático.

1.5 Organização do documento

O documento encontra-se organizado da seguinte forma:

- No Capítulo 2, são apresentados conceitos sobre análise estatística e análise estatística geográfica;
- No Capítulo 3, são abordados conceitos para representação geográfica da informação, sendo abordados sistemas de coordenadas, projeções e *standards* para a disponibilização da informação;
- No Capítulo 4, são abordadas as ferramentas de suporte à implementação do sistema;
- No Capítulo 5, é proposta e apresentada a implementação da solução, sendo debatidos alguns detalhes de implementação;
- No Capítulo 6 são feitas algumas considerações finais e são abordados alguns temas para trabalho futuro.

[6] http://geo.geek.nz/post/27876100534/esri-maps-for-office-launches-extends-location

[7] http://www.esri.com/software/esri-maps-for-office/features

2

Análise Estatística

A análise estatística engloba a recolha, organização, análise e interpretação dos dados, suportando a tomada de decisões [13].

Relativamente à análise de dados foram identificados dois tipos de análises de dados [14]:

- *Exploratory data analysis* (EDA);
- *Confirmatory data analysis* (CDA).

Estes dois métodos de análise de dados fornecem duas abordagens diferentes para a comparação dos dados observados. O método EDA, define uma abordagem de análise a um conjunto de dados que possibilita ao utilizador a exploração de dados sintetizados, tendo este uma visão geral sobre o conjunto. Recorre-se normalmente ao uso de ferramentas visuais, tais como gráficos, permitindo a identificação de padrões na informação. O método de análise EDA é normalmente usado para conhecer o conjunto de dados, sendo por isso uma análise que não necessita de um modelo estatístico ou de a formulação de uma hipótese.

O método de análise CDA, ao contrário do método EDA, possibilita uma análise de confirmação de uma hipótese anteriormente formulada. Sendo uma análise mais focada e centrada apenas num subconjunto dos dados, são utilizados valores numéricos para as comparações em detrimento de gráficos, usados na análise EDA.

Embora sejam abordagens diferentes, estes dois tipos de análise de dados podem complementar-se, sendo uma mais valia a existência de uma ferramenta que permita integrar as duas análises.

Quando se trata de análise espacial, existem extensões aos dois tipos de análises identificadas [15]:

- *Exploratory spatial data analysis* (ESDA);
- *Confirmatory spatial data analysis* (CSDA).

O método ESDA representa um processo preliminar de modelação e formulação de hipóteses, permitindo conhecer o conjunto de dados disponíveis. A representação destes dados, sendo aplicado a dados georreferenciados, é efetuada sobre diferentes pontos de vista, entre eles os mapas temáticos.

O método CSDA é usado em muitas ferramentas com algoritmos de estimativa, para a confirmação de uma formulação previamente identificada.

2.1 Natureza e características da informação

De acordo com as suas características, a informação poderá ser classificada nos seguintes tipos de medidas de atributos [16]:

- Nominais;
- Categóricos;
- Ordinais;
- Intervalos;
- Rácios.

Os atributos nominais são meramente identificativos, poderão ser utilizados como etiquetas (e.g. o nome de uma pessoa).

Os atributos categóricos são valores discretos que se encontram agrupados segundo uma categoria. No entanto, não existe qualquer ordem entre estes, nem contêm informação que os possa diferenciar dos membros do grupo (e.g. cor dos olhos).

Os atributos ordinais são um caso particular dos atributos categóricos, onde é possível estabelecer uma relação de ordem. No entanto, a informação existente apenas permite ordenar não permitindo quantificar a diferença entre os diversos valores. (e.g. As fases da vida: Jovem, Adulto e Idoso).

Os intervalos são medidas ordinais sendo possível estabelecer a diferença entre os diversos valores. São

medidas dimensionais normalmente expressas numericamente (e.g. Intervalos de idades: 10-20, 20-30).

Os rácios são medidas adimensionais, sendo assim possível estabelecer uma comparação quantitativa entre os atributos desta categoria.

2.1.1 Variáveis

Os valores adquiridos segundo cada uma das medidas anteriores, contêm informação acerca de uma determinada característica da dimensão de análise [16].

Para a pesquisa de informação, os valores dos atributos das dimensões passam a ser variáveis, possibilitando a pesquisa de padrões na informação.

De acordo com os tipos de medidas dos atributos, a Figura 2.1 representa o tipo de variáveis existentes, podendo estas ser discretas ou contínuas. As variáveis contínuas, podem assumir todos os valores de um intervalo - caso da idade em anos, meses, horas, etc. No entanto, para facilitar a análise de dados, é possível discretizar uma variável contínua, facilitando assim a identificação de padrões. Por exemplo, no caso do ano, este pode ser discretizado através do dia da semana, dia do mês ou até pelo próprio mês, passando a ser uma variável discreta com valores cíclicos.

As variáveis discretas representadas na Figura 2.1 (Constantes, Binárias, Valor múltiplo e Valor duplo), assumem apenas um conjunto de valores numerável.

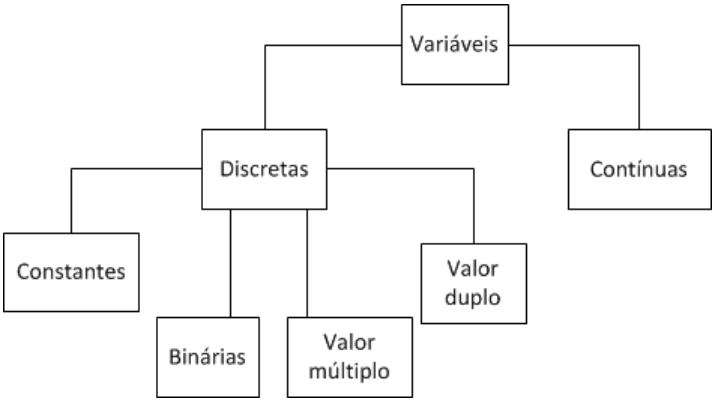

Figura 2.1. Classificação da informação das variáveis dos atributos.

Para a escolha acertada das variáveis que melhor permitem ilustrar um determinado tema, contribui o nível de conhecimento do assunto a analisar e dos dados em questão. Assim, a tarefa de composição de um gráfico ou de um mapa, de acordo com os objetivos pretendidos, é muitas vezes uma tarefa de tentativa erro [17].

2.2 Dimensões

As dimensões são o ponto de entrada nas tabelas de factos, contendo a sua descrição e a do negócio em questão. Os atributos das dimensões descrevem os factos tentando organizar a informação e possibilitando que esta seja explorada. [18]

As dimensões são usadas para filtrar e agregar os factos a apresentar numa determinada análise a um determinado nível de detalhe, sendo as agregações dos dados normalmente associadas às hierarquias dos atributos das dimensões [19].

De forma a facilitar a navegação e a síntese da informação, é possível indicar hierarquias nos atributos das dimensões. A Figura 2.2, apresenta a hierarquia da localização.

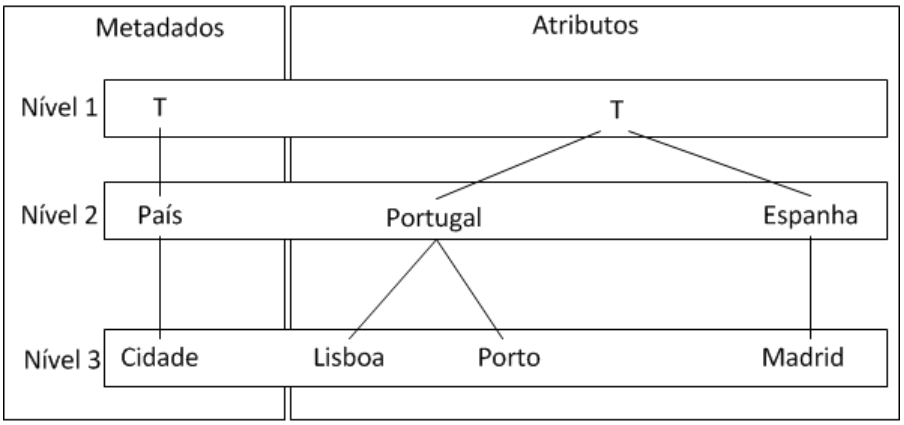

Figura 2.2. Representação da hierarquia da dimensão localização.

A hierarquia da localização, representada na Figura 2.2 apresenta três níveis. O primeiro nível de agregação, o total, um segundo nível com a informação agregada ao nível de País e o terceiro nível, com a informação de cidade. Sendo o terceiro nível, o mais granular da informação.

A definição das hierarquias é importante na medida em que indica algumas das agregações que se pretendem efetuar na análise da informação. Declarando as hierarquias existentes, é possível efetuar um pré processamento da informação, para o cálculo destas possíveis agregações, diminuindo o tempo de resposta do sistema.

2.3 Factos

Os factos são a medida do negócio que deverá ser analisada de forma a compreender um determinado comportamento ou padrão. [18][19]

Um facto, tem associado um conjunto de dimensões de análise e uma granularidade [19]. A granularidade do facto é determinada pela combinação das suas dimensões de análise. Considerando as seguintes dimensões:

- Produto;
- Cidade;
- Data (representada ao dia).

Se a tabela de factos em questão representa as vendas, a granularidade do facto diz respeito às vendas de um produto por dia e por cidade.

A integração da informação no exemplo anterior é sintetizada e posteriormente integrada na tabela de factos. Note-se que a granularidade do facto é a venda de produtos por dia e por cidade, o que implica que a informação seja agregada (por dia e por cidade) aquando da sua integração no sistema. [18]

2.3.1 Tipos de factos

Os factos mais relevantes são factos numéricos e aditivos, uma vez que é possível aplicar várias funções para sumariar a informação. É possível efetuar somas, médias, calcular mínimos e máximos sobre qualquer uma das dimensões de análise [18] [19].

No entanto, existem ainda os seguintes tipos de factos:

- Semi-aditivos: Estes factos apenas podem ser agregados ao longo de algumas dimensões;
- Não aditivos: Estes factos não podem ser sumariados, são factos já calculados, por exemplo, percentagem de desconto;
- Não numéricos: Tal como os não aditivos, estes factos não podem ser sumariados. Apenas se podem efetuar contagens sobre eles.

2.4 Mapas temáticos

Os mapas temáticos permitem a representação de informação quantitativa ou qualitativa, tendo sido concebidos com o propósito de representar um tema específico associando-o a uma área geográfica [20], sendo normalmente utilizados para a representação de indicadores demográficos, tais como a densidade populacional.

Na criação de mapas temáticos, são destacadas quatro técnicas [21]:

- Mapas coropletos;
- Mapas de símbolos proporcionais;
- Mapas isopletos;
- Mapas de pontos.

Os mapas coropletos representam dados quantitativos agregados sobre regiões predefinidas através de uma cor. A cor é a representação de uma frequência que possibilita a comparação com outras regiões, dando assim uma visão abrangente dos dados. A Figura 2.3, apresenta um exemplo de um mapa coropleto onde é sintetizado o número de nascimentos relativos a 2011. De acordo com a legenda, as áreas mais escuras representam os países onde a taxa de nascimentos por cada 1000 habitantes é maior. As áreas mais claras representam os países onde a taxa de nascimentos é menor.

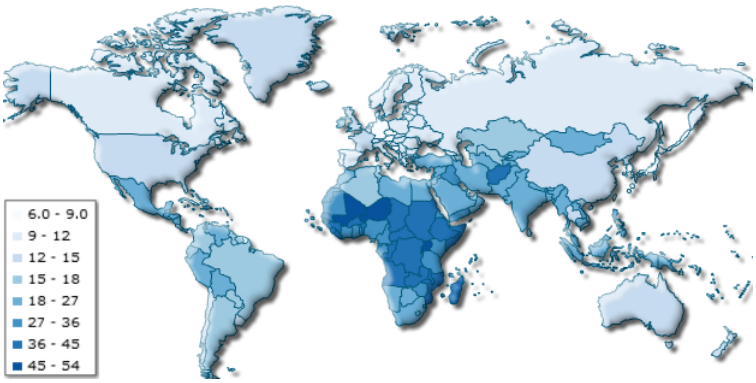

Figura 2.3. Exemplo de um mapa coropleto (Imagem adaptada de: http://www.indexmundi.com/map/?v=25).

Os mapas de símbolos proporcionais variam o tamanho do símbolo de acordo com o valor correspondente à área geográfica representada. Por exemplo, a Figura 2.4, apresenta o número de utilizadores de Internet referente ao ano de 2004, representando os círculos maiores nos locais onde existem mais utilizadores de Internet.

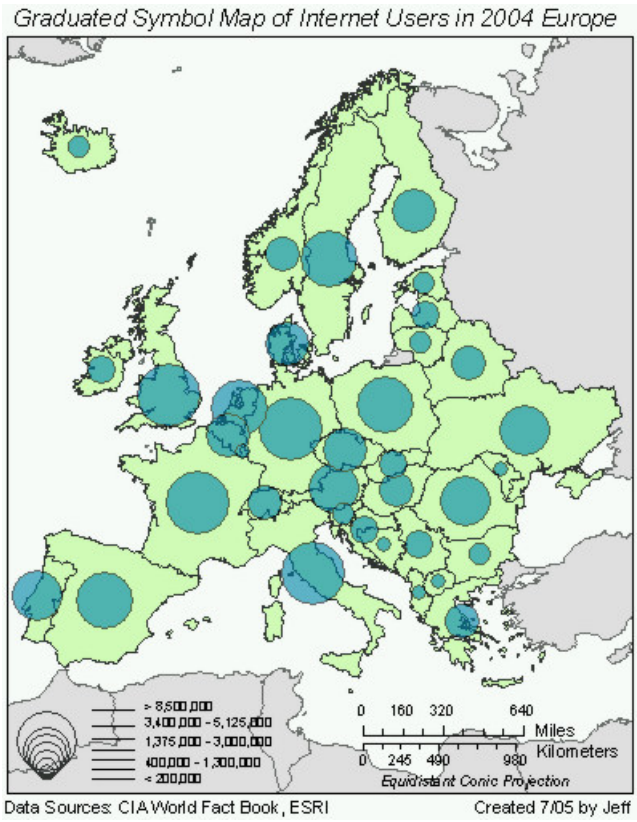

Figura 2.4. Exemplo de símbolos proporcionais (Imagem adaptada de: http://www.geog.ucsb.edu).

Os mapas isopletos, também conhecidos como mapas de contornos, representam fenómenos contínuos, como por exemplo a precipitação, ou a temperatura. A Figura 2.5 apresenta um exemplo da utilização de mapas isopletos, representando as temperaturas no dia 9 de Agosto de 2000.

Na observação do mapa, é possível verificar que não existe separação das cores de acordo com as regiões, existindo regiões com várias tonalidades de cor. Esta é a principal característica dos mapas isopletos contrastando assim com os mapas coropletos.

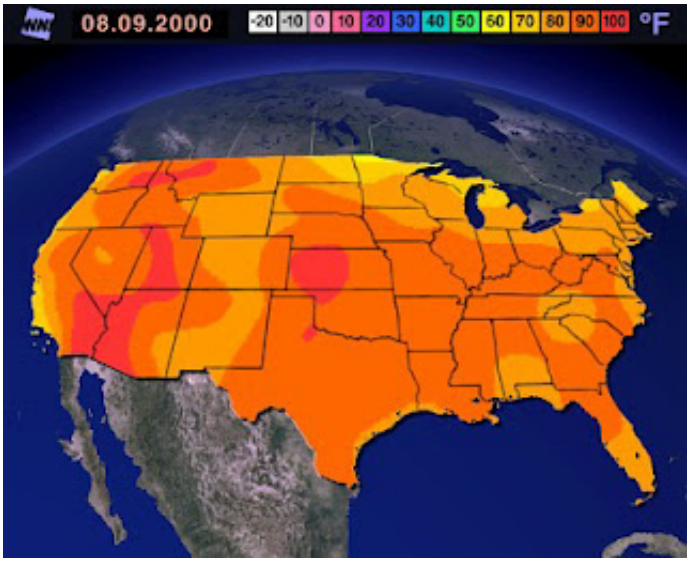

Figura 2.5. Exemplo de mapas isopletos (Imagem adaptada de: http://enb105-2012s-kst.blogspot.pt/).

Um mapa de pontos é usado para representar a presença de um tema associado a uma localização, podendo representar uma ou várias unidades. A Figura 2.6 apresenta um exemplo da utilização de um mapa de pontos, representando a distribuição de depósitos de minerais pelo mundo.

Figura 2.6. Exemplo de mapas de pontos (Imagem adaptada de: http://www.mapsofworld.com/world-mineral-map.htm).

2.5 Mapas coropletos

A representação da informação em mapas coropletos, implica a partição da informação em classes de valores assumidos pela variável que se pretende representar. A partição poderá variar de acordo com a forma como os dados são distribuídos e do método de classificação escolhido, podendo assim existir mapas diferentes para o mesmo conjunto de dados. [17]

O processo de classificação pressupõe duas questões que é necessário responder [17]:

1. Em quantas classes se agrupa o conjunto de dados?
2. Quais deverão ser os limites das classes?

Os valores absolutos não são adequados para a construção de gráficos coropletos, uma vez que as regiões não são homogéneas em termos de área. Por exemplo, em termos de população, é expectável que uma região com uma área maior tenha um maior número de habitantes. Assim, de forma a possibilitar a comparação entre regiões com dimensões diferentes, é necessário proceder à normalização dos dados, de forma a evitar uma má interpretação da representação. A normalização dos dados poderá recorrer ao tamanho das áreas, no caso de indicadores sobre a densidade populacional, ou ao número de habitantes, no caso de indicadores sobre natalidade. [17][20]

2.5.1 Número de classes

O número de classes não se baseia unicamente no estudo dos dados, devendo variar de acordo com o público alvo, a distribuição dos dados na variável a analisar, e de acordo com o número de unidades territoriais. [17][22]

O agrupamento dos dados num número reduzido de classes leva a uma maior generalização, uma vez que existe apenas um símbolo a que correspondem muitos dados. Se o número de classes for maior, existe uma maior aproximação à realidade, no entanto, temos as limitações do nosso sistema visual. Assim, para o público comum foi definido que poderão existir até sete classes para análise, podendo este número ser mais elevado para um público mais experiente, até um limite de 10 classes.[17]

Sendo que o número de classes também deverá depender do número de unidades territoriais visualizadas, [22] defende que o número de classes deverá ser calculado de acordo com a Fórmula 2.1 e [17] com a Fórmula 2.2, onde C representa o número de classes e n o número de unidades territoriais.

$$C = 1 + 3.3 \times ln(n) \tag{2.1}$$

$$C = 1 + 3.22 \times log(n) \tag{2.2}$$

2.5.2 Limites das classes

A seleção do método de classificação implica o conhecimento dos dados, assim como o objetivo da representação. Na seleção do método de classificação deverá ser tido em conta o público alvo da análise, sendo que para um público pouco experiente, os princípios de construção dos mapas deverão ser simples. [17]

Existem dois grupos para os métodos de classificação. Os métodos matemáticos, onde os cálculos se baseiam nos valores da variável, permitindo formar intervalos de amplitude regular e os métodos estatísticos que se baseiam no estudo das frequências dos valores.

Dos métodos de classificação existentes, destacam-se os seguintes métodos:

- Métodos matemáticos;
  - Intervalos iguais;
  - Intervalos em progressão aritmética;
- Métodos estatísticos;
  - Quantis;
  - Média e desvio-padrão.

Para um conjunto ordenado de valores, considere-se a Fórmula 2.3:

$$A = M - m \tag{2.3}$$

onde, *A* representa a amplitude dos dados, dada pela diferença entre o maior valor (*M*) e o menor valor (*m*).

Intervalos iguais

Neste método de classificação, a distribuição dos dados pelas diferentes classes é uniforme. No entanto, existe a possibilidade de existirem classes vazias se houver descontinuidade na distribuição dos dados.

Considere-se a Fórmula 2.4:

$$C = A/n \tag{2.4}$$

onde, *C* representa a amplitude dos intervalos e *n* o número de classes.

O cálculo dos limites superiores das classes é apresentado na Fórmula 2.5, sendo o limite inferior, o limite superior da classe precedente:

$$L_1 = m + 1 \times C, L_2 = m + 2 \times C, ..., L_n = m + n \times C \tag{2.5}$$

Intervalos em progressão aritmética

Este método de classificação é adequado para distribuições assimétricas, fornecendo classes mais detalhadas à parte da distribuição com mais valores.

Considere-se:

$$R = A/C \tag{2.6}$$

onde *R* representa a razão da progressão, *A* a amplitude e *C* a amplitude das classes calculada de acordo com a Fórmula 2.7.

$$C = (n + 1) \times \frac{n}{2} \tag{2.7}$$

onde, *n* representa o número de classes.

O cálculo dos limites superiores das classes é apresentado na Fórmula 2.8, sendo o limite inferior, o limite superior da classe precedente:

$$L_1 = m + R, L_2 = m + 2 \times R, ..., L_n = m + n \times R \tag{2.8}$$

Quantis

Este método de classificação permite eliminar o peso dos valores extremos. No entanto, pode dar origem a uma escolha incorreta dos limites na existência de descontinuidades, uma vez que ignora as particularidades da distribuição dando a mesma importância a cada classe.

O cálculo dos limites das classes é apresentado na Fórmula 2.9:

$$V = N/n \tag{2.9}$$

onde, *N* é o número total de observações e *n* o número total de classes. A primeira classe contém os primeiros *V* valores, a segunda, os segundos *V* valores e a última, os últimos *V* valores.

Média e desvio-padrão

Este método de classificação permite a comparação entre mapas diferentes, sendo pouco adequado para a sua utilização com um grande número de classes.

Sendo X_T a média e S_T o desvio-padrão do conjunto, para um número de classes ímpar, a classe intermédia terá como limite inferior: $X_T - S_T$ e como limite superior $X_T + S_T$.
A Fórmula 2.10, apresenta o cálculo dos restantes intervalos.

$$L_{inf(i-1)} = X_T - 2 \times S_T, L_{sup(i-1)} = X_T + 2 \times S_T \tag{2.10}$$

3

Georreferenciação de conteúdos

Um sistema de informação geográfica (SIG), tem como principal objetivo o armazenamento, consulta e disponibilização digital de objetos espaciais com foco na sua localização terrestre. [23]

A representação da informação em sistemas SIG, é efetuada através de dois tipos de dados [2]:

- Dados *raster*;
- Dados vetoriais.

Os dados *raster* são fotografias aéreas georreferencias da terra e contêm informação de disposição e georreferenciação. Os dados vetoriais, também se encontram georreferenciados, sendo associados a uma localização, e podem ser de três tipos:

- Linhas;
- Pontos;
- Polígonos.

3.1 *Feature*

Uma *feature*[1], também designada por objeto geográfico ou entidade, tem como objetivo a representação geográfica de um objeto do mundo real, por exemplo: um edifício, uma ponte ou uma estrada. Cada *feature* é composta por duas componentes [23]:

[1] Uma vez que não foi encontrado um termo equivalente em Português, é usado o termo *feature* para descrever um objeto geográfico.

- Uma componente descritiva que contém um conjunto de atributos que a caraterizam. Por exemplo, o nome do edifício, o ano de construção e o proprietário;
- Uma componente geográfica permitindo a sua localização no espaço.

A representação de uma *feature* é efetuada através dos três tipos de dados vetoriais apresentados anteriormente (Linhas, pontos e polígonos), sendo que devido à complexidade dos objetos do mundo real e existindo a possibilidade de composição, foram introduzidos os conceitos de [23]:

- *Feature* simples;
- *Feature* complexa.

Uma *feature* simples contém apenas a modelação de um único objeto do mundo real, por exemplo um edifício. Uma *feature* complexa baseia-se na composição de outras *features* para a sua modelação. Por exemplo, a representação de uma cidade e dos seus edifícios. A Tabela 3.1, apresenta a definição de *features* simples e complexas através da sua composição. [23]

feature = (descrição, geografia) – feature simples
| (descrição, {feature}) – feature complexa

Tabela 3.1. Definição de *feature*

3.2 Temas

Nos sistemas SIG a informação espacial que modela o mesmo tipo de objetos, é agrupada num tema. Por exemplo, Rios, Prédios ou Cidades, são diferentes temas cada um com *features* que retratam este tipo de objetos. Assim, um tema é um conjunto de *features* que pretendem modelar objetos do mesmo tipo. [23]

3.3 Sistema de coordenadas

Para a localização de uma *feature* no espaço, é necessária a utilização de um sistema de coordenadas de forma a expressar um ponto (X, Y). [2]

Existem muitos sistemas de coordenadas disponíveis, sendo os mais conhecidos e usados o *World Geodetic System 1984* (WGS84), identificado pelo SRID (*Spatial Reference System Identifier*) 4326 e o sistema métrico, identificado pelo SRID 900913. [2]

O sistema de coordenadas WGS84, pode ser representado com as coordenadas em graus, minutos e segundos, ou em alternativa num modelo decimal - existindo a possibilidade de conversão entre os dois formatos[2]. Este sistema de coordenadas é muito popular, sendo usado o modelo graus, minutos e segundos em muitos GPS e o modelo decimal como interface dos maiores fornecedores de mapas na *Web*, entre eles Google[3] e Bing[4].

O sistema métrico apresenta como grande vantagem a possibilidade do cálculo de áreas e distâncias entre dois pontos com uma maior precisão. O sistema de coordenadas WGS84 define as coordenadas em graus, sendo que um grau de latitude pode variar mais de 21 Kms entre o equador e os pólos.

3.4 Projeções

De forma a possibilitar a representação das *features* sobre um mapa, é necessário a utilização de uma projeção.

Uma projeção baseia-se em modelos matemáticos que permitem transformar um modelo representado numa superfície tridimensional curvada, num plano bidimensional. No entanto, é necessário ter em consideração que a utilização de uma projeção irá sempre criar distorção na representação, tendo sido identificados quatro tipos de distorções possíveis: [2]

- Distância;
- Direção;
- Forma;
- Área.

---

[2] http://www.jeepreviews.com/wireless-gps-coordinates/

[3] http://maps.google.com

[4] http://maps.bing.com

Existem muitas projeções disponíveis, sendo estas agrupadas em três grupos: [24]

- Projeções cilíndricas;
- Projeções cónicas;
- Projeções para um plano (*Azimuthal*).

Destes três grupos de projeções identificadas, serão destacadas de seguida duas projeções cilíndricas: Mercator e *Universal Transverse Mercator* (UTM).

3.4.1 Mercator

Na sua representação num plano cartesiano, a longitude corresponde ao eixo do x e a latitude ao eixo do y, sendo as linhas de latitude e longitude paralelas entre si. A Figura 3.1, apresenta a informação através da projeção Mercator num plano cartesiano. [2]

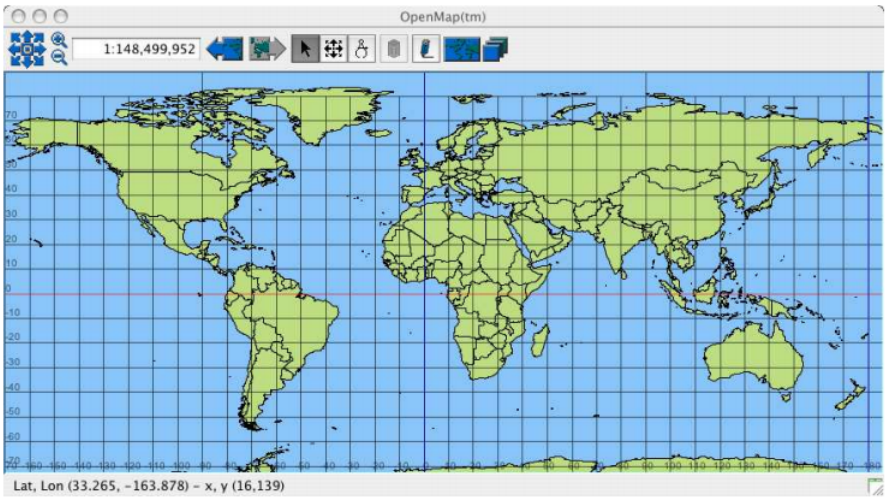

Figura 3.1. Representação do mapa Mundo num plano cartesiano. (Imagem adaptada de [2]).

Os problemas de precisão devem-se ao facto das linhas de longitude serem paralelas, uma vez que num modelo esférico, estas convergem para um único ponto nos pólos. A Figura 3.2 apresenta o modelo esférico da representação num plano cartesiano.

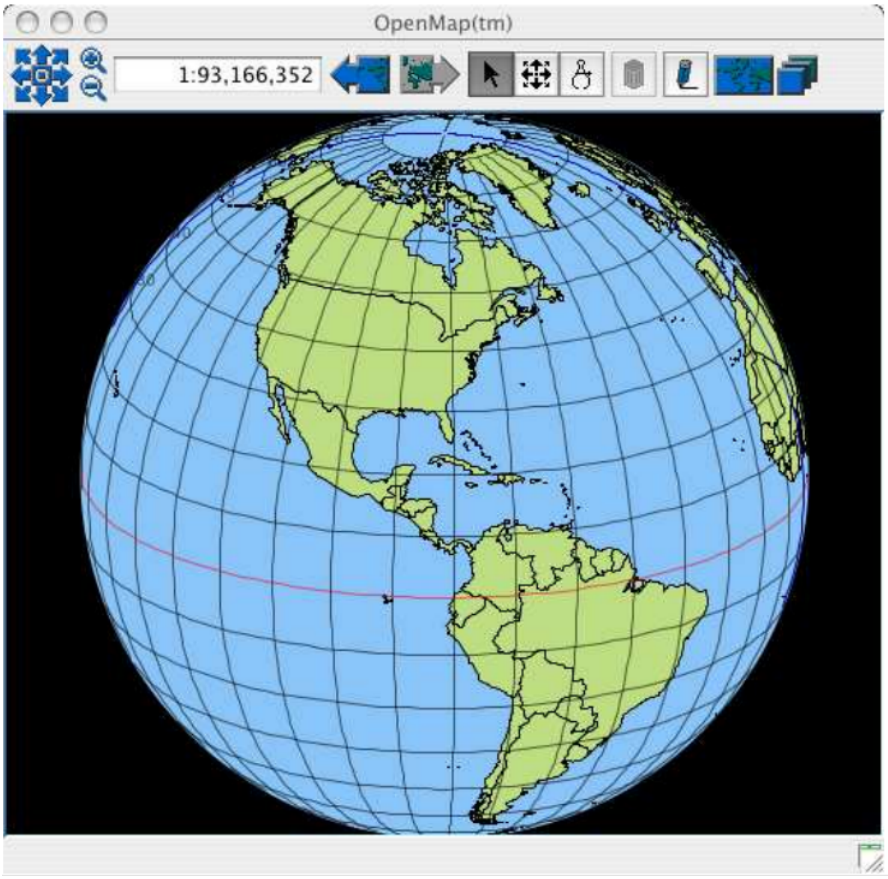

Figura 3.2. Representação do mapa Mundo numa esfera. (Imagem adaptada de [2]).

À medida que nos aproximamos dos pólos a imprecisão é maior. Pela observação da Figura 3.2, podemos concluir que perto dos pólos cada célula da grelha não é representada por um quadrado, sendo no modelo esférico um trapézio e transformando-se num triângulo nos pólos.

Como consequência, podemos observar na Figura 3.1 o tamanho desproporcional da Gronelândia comparativamente ao tamanho da África do Sul. Assim, esta projeção apresenta distorções ao nível da área, da forma e da distância, preservando apenas a direção.

3.4.2 *Universal Transverse Mercator*

A projeção *Universal Transverse Mercator* (UTM) tem como objetivo preservar a distância, usando as suas coordenas em metros em vez de graus, sendo assim possível a utilização de uma régua para medir

distâncias sobre o mapa. Esta projeção preserva a distância, a área e a forma, distorcendo a direção. Por vezes nem sempre representa a norte o que deveria estar representado a norte.[2]

De forma a minimizar o número de distorções, são usadas diferentes projeções na representação de diferentes locais do globo. Assim, a projeção UTM não é uma projeção mas sim um conjunto de projeções, cada uma delas usada na projeção de diferentes áreas geográficas.

A projeção UTM divide o globo em 60 zonas, cada uma com 6º de longitude, a que correspondem 60 zonas de longitude a norte e 60 zonas de longitude a sul. Com a divisão do globo em 120 zonas, a área a projetar é mais pequena, podendo ser usada em cada uma das zonas uma projeção diferente de forma a minimizar as distorções provocadas. A utilização desta projeção implica um compromisso entre a precisão e a quantidade de informação representada. [2]

A distorção causada por uma projeção é influenciada por vários fatores, entre eles, a projeção escolhida e o sistema de coordenadas no qual se encontra representada a informação [2].

3.5 *OGC Web Services*

Open Geospatial Consortium (OGC)[5], é uma entidade composta por empresas, agências governamentais e universidades com o objetivo de definição de *standards* para os sistemas SIG.

Com o objetivo de possibilitar o consumo da informação dos sistemas SIG na *Web*, são definidos pelo OGC dois *standards*:

- *Web Map Service* (WMS) [5];
- *Web Feature Service* (WFS) [25].

3.5.1 *Web Map Service*

O serviço WMS tem como principal objetivo a produção de imagens de mapa através de informação geográfica, que poderão ser dados *raster* ou vetoriais, como identificados anteriormente.

[5] http://www.opengeospatial.org/

Nesta secção serão analisadas as operações disponíveis neste serviço, sendo posteriormente utilizado para a integração das imagens temáticas na ferramenta de análise estatística.

O serviço WMS é disponibilizado através de uma interface REST e define três operações:

- GetMap: Usada para obter a imagem de mapa com a informação geográfica;
- GetCapabilities: Permite obter os meta-dados da informação disponível;
- GetFeatureInfo: Permite obter informação sobre determinadas *features* representadas.

GetCapabilities

A operação GetCapabilites existe em todas as implementações do standard WMS, possibilitando a consulta dos temas disponíveis e dos parâmetros para cada um dos temas.

GetMap

A operação GetMap devolve um mapa com informação de um ou vários temas, representando as suas *features* abordadas na secção 3.1, de acordo uma folha de estilo designada por *Style Layer Descriptor* (SLD), abordada na secção 3.5.3.

A Tabela 3.2, apresenta os parâmetros obrigatórios para um pedido WMS.

Adicionalmente como parâmetros opcionais, é possível indicar se a imagem deverá ser transparente nos locais onde não se encontram representadas *features* (parâmetro TRANSPARENT), ou se esse espaço deverá ser preenchido com uma determinada cor (parâmetro BGCOLOR).

Parâmetro	Exemplo	Descrição
VERSION=1.3.0	VERSION=1.3.0	Identifica a versão do protocolo a usar.
REQUEST=GetMap	REQUEST=GetMap	Indica o tipo de pedido.
LAYERS=layer_list	LAYERS=edificios	Indica o tema ou temas que se pretende representar. Deverão ser separados por vírgula, caso sejam vários.
STYLES=style_list	STYLES=edifícios	Indica o nome dos estilos a aplicar à *layer*. Os estilos estão definidos numa folha de estilo abordada na secção 3.5.3.
CRS=namespace:identifier	CRS=EPSG:4326	Define o sistema de coordenadas em que as *features* deverão ser projetadas sobre a imagem.
BBOX=minx,miny,maxx,maxy	BBOX=-170 0,-50 90	Define a área para a qual se pretende visualizar as *features*. Este parâmetro permite filtrar o conjunto de dados, excluindo as *features* que se encontram fora da área a representar.
WIDTH=image_width	WIDTH=256	Indica a largura da imagem gerada.
HEIGHT=image_height	HEIGHT=256	Indica a altura da imagem gerada.
FORMAT=output_format	FORMAT=image/png	Indica o formato da imagem gerada, podendo ser jpeg, png ou gif.

Tabela 3.2. Parâmetros para o pedido WMS GetMap (Adaptada de [5])

A listagem 3.1 apresenta um exemplo de um pedido WMS GetMap para a obtenção de uma imagem onde é representado o distrito de Lisboa.

```
http://localhost:8080/geoserver/wms?
  LAYERS=district
  &STYLES=
  &FORMAT=image/png
  &SERVICE=WMS
  &VERSION=1.3.0
  &REQUEST=GetMap
  &CRS=EPSG:4326
  &BBOX=-9.8056841594026,38.606861032385,-8.761896622794,39.111919517841
  &WIDTH=682
```

```
11 &HEIGHT=330
```

Listagem 3.1. Exemplo de um pedido WMS GetMap

GetFeatureInfo

A operação GetFeatureInfo possibilita a pesquisa de *features* que se encontram numa determinada localização da imagem, sendo necessário indicar o ponto (X,Y), o tamanho da imagem e a área visível de forma a ser possível calcular a coordenada do ponto selecionado. Um caso de uso comum para esta operação é o clique no mapa. O utilizador efetua o clique sobre o mapa, e são devolvidas as *features* correspondentes ao ponto clicado.

A Tabela 3.3, apresenta os parâmetros obrigatórios para um pedido WMS à operação GetFeatureInfo, ao qual deverão ser adicionados os parâmetros obrigatórios identificados na operação GetMap.

Parâmetro	Exemplo	Descrição
VERSION=1.3.0	VERSION=1.3.0	Identifica a versão do protocolo a usar.
REQUEST=GetFeatureInfo	REQUEST=GetFeatureInfo	Indica o tipo de pedido.
QUERY_LAYERS=layer_list	LAYERS=edificios	Indica o tema ou temas que se pretende representar. Deverão ser separados por vírgula, caso sejam vários.
INFO_FORMAT=output_format	INFO_FORMAT=text/xml	Indica o formato da resposta ao pedido.
X=pixel_column	X=200	Coordenada X da imagem clicada.
Y=pixel_column	Y=100	Coordenada Y da imagem clicada.

Tabela 3.3. Parâmetros para o pedido WMS GetFeatureInfo (Adaptada de [5])

A Listagem 3.2 apresenta um exemplo de um pedido GetFeatureInfo de forma a obter informação sobre o distrito de Lisboa.

```
1 http://localhost:8080/geoserver/wms?
  REQUEST=GetFeatureInfo
3 &BBOX=-10.153103%2C38.583904%2C-8.065528%2C39.594021
  &SERVICE=WMS                    &INFO_FORMAT=text/xml
```

```
&QUERY_LAYERS=district          &FEATURE_COUNT=50
&LAYERS=district                &WIDTH=682
&HEIGHT=330                     &STYLES=
&CRS=EPSG:4326                  &VERSION=1.3.0
&Y=269                          &Y=255
```

Listagem 3.2. Exemplo de um pedido WMS GetFeatureInfo

3.5.2 *Web Feature Service*

O serviço WFS tem como objetivo a gestão e a pesquisa de *features* sobre um conjunto de dados. Ao contrário do serviço WMS que devolve uma imagem, este serviço devolve os dados/*features* no formato GML (*Geography Markup Language*), abordado na secção 3.6.2.

Para a gestão de *features* sobre um repositório de dados, este serviço disponibiliza as seguintes operações:

- Criar uma nova *feature*;
- Apagar uma *feature* existente;
- Atualizar uma *feature*;
- Bloquear uma *feature*. Esta operação é normalmente utilizada com o objetivo de atualização, existindo o conceito transacional e de tempo limite de operação associados;
- Pesquisa de *features*.

3.5.3 *Style Layer Descriptor*

Style Layer Descriptor (SLD) é um documento de estilos que permite manipular a representação das *features* desenhadas. O serviço WMS, apresentado na secção 3.5.1, devolve uma imagem com a representação de *features*, cuja visualização poderá ser manipulada através de um documento de estilos. [3]

Um documento SLD possibilita aplicar estilo a:

- Todas as *features* de um tema;
- Uma *feature* de acordo com as suas características ou atributos;
- Todas as *features* dos vários temas representados.

Assim, é possível representar numa imagem os rios, as estradas e os edifícios com cores e representações diferentes, sendo ainda possível variar a largura da estrada de acordo com as suas características, podendo por exemplo, destacar os auto-estradas e os itinerários principais. A Figura 3.3, apresenta uma imagem em que são representados rios, estradas e edifícios, tendo sido aplicado um estilo diferente às *features* de cada tema.

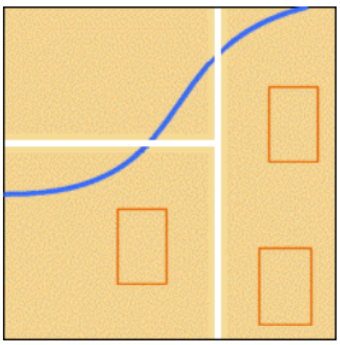

Figura 3.3. Representação de vários temas com diferentes estilos. (Imagem adaptada de [3]).

Um documento SLD é um documento XML com o *schema* apresentado na Listagem 3.3. O Elemento NamedLayer permite aplicar um estilo pré-definido a todas as *features* de um determinado tema. O elemento UserLayer contém mais opções permitindo alterar o estilo aplicado a cada *feature* de acordo com as suas características ou atributos.

```
<xs:element name="StyledLayerDescriptor">
  <xs:complexType>
    <xs:choice minOccurs="0" maxOccurs="unbounded">
      <xs:element ref="sld:NamedLayer"/>
      <xs:element ref="sld:UserLayer"/>
    </xs:choice>
    <xs:attribute name="version" type="xs:string" use="required"/>
  </xs:complexType>
</xs:element>
```

Listagem 3.3. *Schema* de um documento SLD.

A Listagem 3.4 apresenta o estilo utilizado para a geração da imagem da Figura 3.3. Para cada um dos temas (rios, estradas e edifícios), são usados estilos pré-definidos, sendo aplicado o mesmo estilo a todas

as *features* de cada tema.

```
<StyledLayerDescriptor version="1.0.0">
    <NamedLayer>
        <Name>Rivers</Name>
        <NamedStyle>
            <Name>CenterLine</Name>
        </NamedStyle>
    </NamedLayer>
    <NamedLayer>
        <Name>Roads</Name>
        <NamedStyle>
            <Name>CenterLine</Name>
        </NamedStyle>
    </NamedLayer>
    <NamedLayer>
        <Name>Houses</Name>
        <NamedStyle>
            <Name>Outline</Name>
        </NamedStyle>
    </NamedLayer>
</StyledLayerDescriptor>
```

Listagem 3.4. Exemplo de um documento SLD.

3.6 Armazenamento e transporte da informação

O Armazenamento e transporte de *features*, pode ser efetuado através de várias alternativas, sendo destacados os seguintes formatos:

- Shapefile [26];
- KML [27];
- GML [28].

Outra alternativa a considerar para o armazenamento da informação, são as bases de dados geográficas, sendo o transporte assegurado por um dos formatos anteriores ou através do serviço WFS, apresentado na secção 3.5.2.

3.6.1 *Shapefile*

O *Shapefile* é um formato para armazenamento e transporte de dados vetoriais definido pela ESRI, contendo *features* acerca de um tema. No seu conteúdo, cada *feature* contém um conjunto de atributos que a caracterizam, sendo um desses um atributo espacial que permite localizar a *feature* no espaço. [26]

Um *Shapefile* é constituído por três ficheiros distintos:

- .shp, é um ficheiro que apenas contém os dados geográficos. Os pontos, as linhas e as geometrias;
- .shx, é um ficheiro usado para indexação da informação geográfica;
- .dbf, é um ficheiro que contém dados não geográficos, associados aos dados vetoriais. Podem ser descrições, nomes ou moradas.

Opcionalmente poderá existir um ficheiro .prj, usado para indicar o sistema de coordenadas e o sistema de projeção utilizados.

3.6.2 GML

GML (*Geography Markup Language*) é um formato para armazenamento e transporte de dados vetoriais definido pelo OGC. O ficheiro de dados é definido em XML, onde se destacam as seguintes primitivas [28]:

- Definição de uma *feature* (*Feature*);
- Modelação de dados vetoriais (*Geometry*);
- Indicação do sistema de coordenadas;
- Observações ou atributos, para indicar as características de cada uma das *features*;
- Definição de *features* dinâmicas, possibilitando adicionar campos que não existem no esquema de definição da *feature*;
- Definição de regras de representação e estilo.

Ao contrário dos *Shapefiles* que definem vários ficheiros para armazenar a informação, o formato GML utiliza apenas um ficheiro onde contém toda a informação das *features* de um tema, sendo este o formato de dados normalmente usado nas respostas a pedidos WFS (ver secção 3.5.2).

3.7 Desenvolvimento de aplicações SIG

Na construção de aplicações SIG são usados os vários temas de informação disponíveis. A cada um dos temas corresponde uma camada de informação, contendo as *features* do tema a representar. Na camada base são normalmente usados dados *raster*, com fotografias aéreas, sendo posteriormente adicionadas as camadas com informação vetorial: ruas, rios, estradas e pontos de interesse. [17]

Na geração dos mapas estas camadas são sobrepostas, sendo necessário para garantir a correta sobreposição dos dados que em cada uma das camadas de informação seja usado o mesmo sistema de coordenadas e a mesma projeção (ver secção 3.3 e 3.4). [2]

A Figura 3.4, apresenta a construção de um mapa recorrendo a várias camadas de informação. A camada base é uma camada de dados *raster*, com fotografias aéreas, sendo posteriormente adicionadas duas camadas com dados vetoriais, com informação de rios e pontos de interesse.

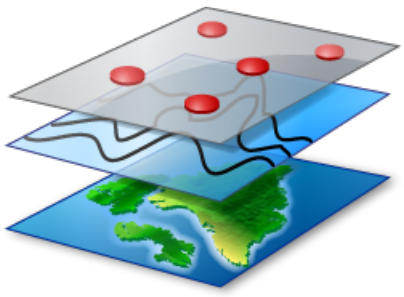

Figura 3.4. Camadas de informação para a construção de mapas (Imagem obtida de: http://www.iconarchive.com/show/gis-gps-map-icons-by-icons-land/Layers-icon.html).

4

Tecnologias de suporte

De forma a viabilizar o desenvolvimento do projeto, foi necessário identificar e estudar tecnologias que permitissem a geração de imagens temáticas e a sua visualização em mapa, dando prioridade a tecnologias que permitam a visualização num *browser*.

4.1 GeoServer

O GeoServer [1], é uma aplicação *Web*, *open-source*, que implementa os serviços *standard* WMS e WFS definidos pelo OGC (ver secção 3.5 do Capítulo 3) para o desenvolvimento de aplicações SIG na *Web*. [4]

De forma a tratar da leitura e armazenamento dos objetos geográficos, o GeoServer foi desenvolvido como uma camada de abstração sobre o GeoTools (ver secção 4.2), sendo este o responsável pela leitura das *features* de várias fontes de dados e aplicação de funções geográficas de filtragem sobre os objetos. Assim, o GeoServer trata da implementação dos serviços definidos pelo OGC e da renderização da informação.[4]

De forma a facilitar o processo de disponibilização da informação como um serviço, o GeoServer disponibiliza uma aplicação *Web* para a administração das fontes de dados e da informação disponibilizada.

A Figura 4.1 apresenta a arquitetura geral do GeoServer. A arquitetura do GeoServer está separada em vários módulos interligados por um injetor de dependências, sendo assim possível estender funcionali-

[1] http://geoserver.org

dade e desenvolver novos módulos (ver secção 4.2.1).

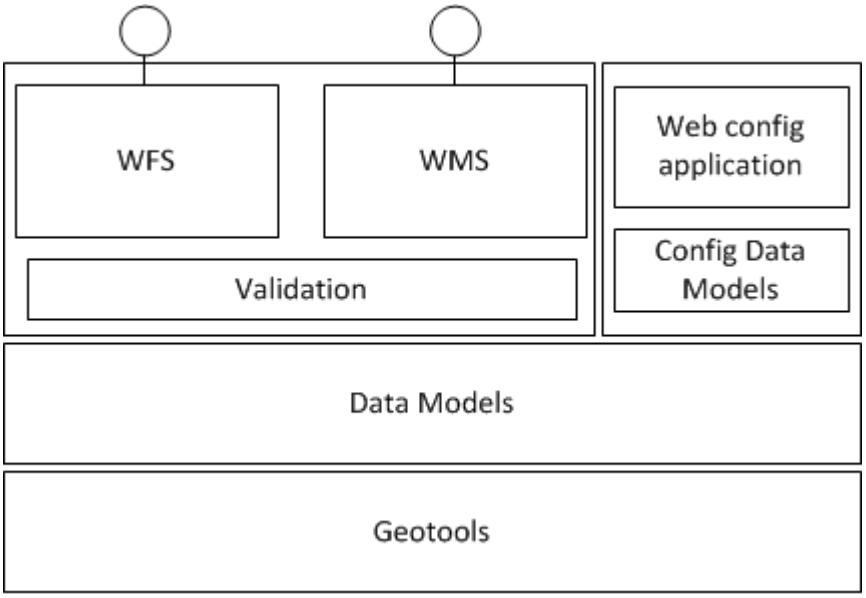

Figura 4.1. Arquitetura do GeoServer (imagem adaptada de http://www.jeevanchaaya.com/2009/03/12/geoserver-a-migration-story-—part-2-talking-the-architecture/)

Na camada de apresentação, é disponibilizada a implementação dos serviços WFS e WMS e a aplicação *Web* para a administração das fontes de dados e da informação disponibilizada, sendo os modelos responsáveis pela leitura e armazenamento da configuração.

A camada de dados é responsável pela filtragem e leitura de *features* de várias fontes de dados. A implementação desta camada é da responsabilidade do GeoTools (ver secção 4.2), que disponibiliza as *features* para o GeoServer devolver em cada um dos serviços WMS ou WFS.

4.1.1 Filtros

De forma a filtrar e limitar a informação apresentada, o GeoServer introduziu nos serviços WMS e WFS um parâmetro para filtrar a informação, de acordo com o standard *Common Query Language* (CQL), criado pelo OGC [29]. [4]

CQL é uma linguagem de interrogação que permite filtrar o conjunto de informação. Na Listagem 4.1 é apresentado um exemplo da utilização de um filtro geográfico:

```
BBOX(the_geom, -90, 40, -60, 45)
```

Listagem 4.1. Exemplo de um filtro CQL

A Listagem 4.1, apresenta a utilização de um filtro geográfico sobre a informação. O resultado da aplicação do filtro pode ser visualizado na Figura 4.2, onde são apenas representadas as regiões que se encontram dentro da área filtrada.

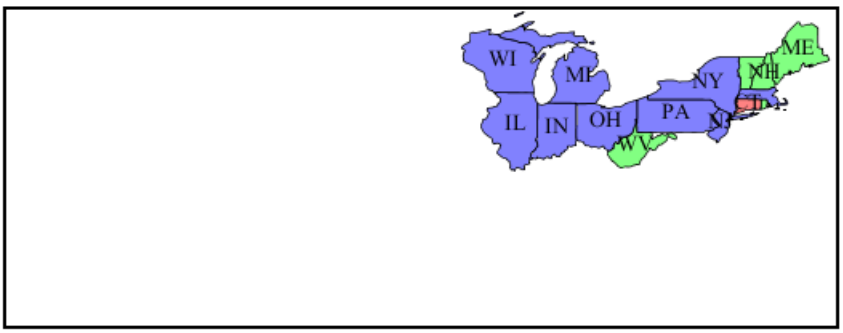

Figura 4.2. Exemplo da aplicação de um filtro geográfico (imagem de [4])

Os filtros deverão ser aplicados sobre uma das características das *features*, existindo vários tipos de operadores. Para comparações simples, estão disponíveis os seguintes operadores:

- >, >=, <, <=, <>, =;
- <caracteristica> BETWEEN <minval> AND <maxval>;
- <caracteristica> LIKE '<valor>'.

Existem ainda filtros de função e filtros geográficos, sendo possível compor os filtros recorrendo às expressões AND e OR. Como exemplo ilustrativo, a Listagem 4.2, apresenta um filtro para obter apenas as regiões onde existem mais de um milhão de habitantes e o número de Homens é superior ao número de Mulheres, considerando apenas uma área geográfica.

```
PERSONS > 1000000 AND MALE > FEMALE AND BBOX(the_geom, -90, 40, -60, 45)
```

Listagem 4.2. Exemplo de composição de um filtro CQL

4.1.2 *Cache*

Para diminuir a carga na geração de imagens e o tempo de resposta dos serviços WMS (ver secção 3.5.1 do Capítulo 3), para dados estáticos, o GeoServer disponibiliza uma estratégia de *cache* através do serviço GeoWebCache [4].

A Figura 4.3, apresenta a arquitetura geral do servidor de *cache*, GeoWebCache.

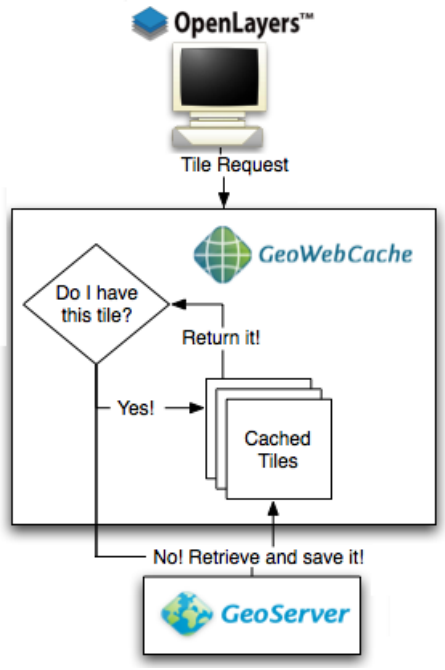

Figura 4.3. Arquitetura do servidor de *cache* (imagem adaptada de http://opengeo.org/publications/opengeo-architecture/)

O GeoWebCache é um servidor de *cache* que se encontra disponível entre o cliente e o GeoServer, funcionando como *proxy* transparente na comunicação de pedidos WMS. Para a obtenção de imagens, o cliente efetua um pedido ao servidor de *cache* que vai validar se a imagem já se encontra disponível em *cache*. Caso a imagem não se encontre ainda em *cache*, o GeoWebCache, efetua um pedido ao GeoServer, guarda a reposta devolvida em *cache*, para futuras utilizações, e devolve a imagem ao cliente.

4.2 Geotools

O GeoTools [2], é uma *framework open-source* para o desenvolvimento de aplicações SIG, sendo usada no GeoServer como o motor SIG para gestão das *features* [30]. A Figura 4.4, apresenta a arquitetura do GeoTools.

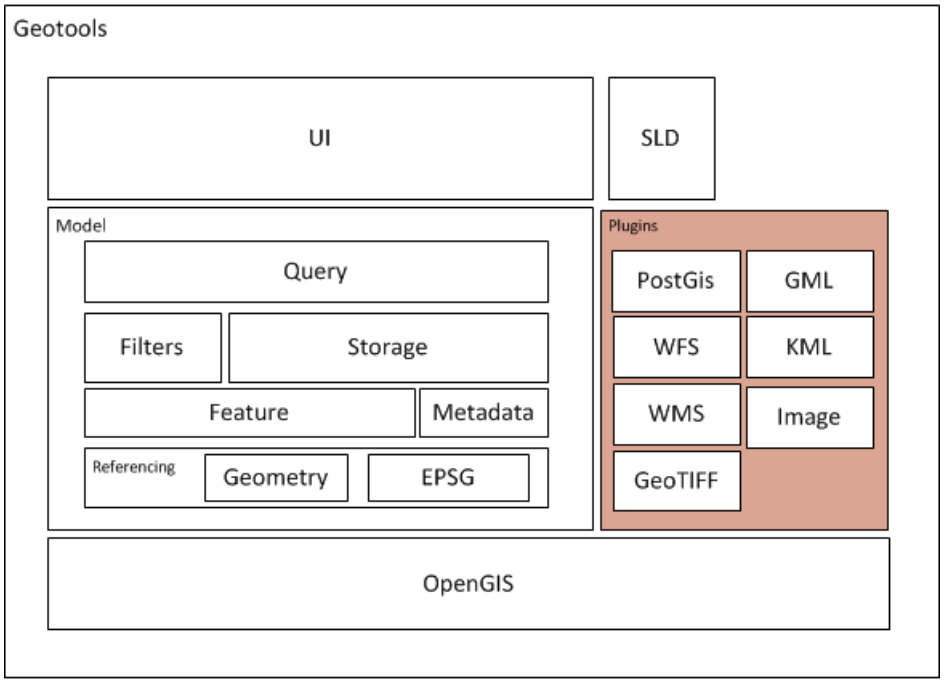

Figura 4.4. Arquitetura do GeoTools (imagem adaptada de http://docs.geotools.org/latest/userguide/welcome/architecture.html)

A camada de apresentação, composta pelos blocos UI e SLD, integram componentes para a visualização da informação. O bloco UI integra componentes para a visualização e edição das *features*, de forma espacial. O bloco SLD condiciona a forma como a informação é representada (ver secção 3.5.3), adicionando estilo à representação.

Os componentes de representação utilizam as *features* provenientes de uma fonte de dados, sobre a qual podem efetuar interrogações. Os componentes do bloco Model, permitem abstrair a fonte de dados e for-

[2] http://www.geotools.org/

necem *features* (com um sistema de coordenadas) permitindo reprojetar a informação.

O GeoTools permite que sejam adicionadas novas fontes de dados através de um sistema de *plugins*. O utilizador tem a possibilidade de desenvolver uma fonte de dados e adicioná-la ao sistema, ficando disponível no GeoTools e GeoServer. As fontes de dados, estão representadas no bloco *plugins*, e existem já alguns *drivers* implementados, entre eles *drivers* para acesso à informação de:

- Base de dados PostgreSQL com a extensão PostGIS;
- Serviço WFS e WMS;
- Ficheiro KML e GML;
- Imagens no formato GeoTIFF.

De forma a possibilitar a utilização de objetos e funções geográficas, para transformação ou cálculo, o GeoTools foi construído sobre o OpenGIS[3], um projeto *open-source* que disponibiliza primitivas geográficas. Por exemplo, a definição de geometrias e cálculos de distância.

4.2.1 Fonte de dados

No GeoTools, uma fonte de dados é um contentor de *features*, sendo possível interrogar o contentor para obter, adicionar, editar e remover *features*.

Cada fonte de dados funciona como um *plugin* adicionado ao GeoTools, ficando também disponível no GeoServer para representação da informação na *Web* através da camada de serviços (ver secção 3.5 do Capítulo 3).

O GeoTools organiza as fontes de dados de acordo com a Figura 4.5. Cada fonte de dados contém *features* sobre vários temas, apresentando cada tema as suas características. Por exemplo, uma fonte de dados para representação de um mapa, necessita ter informação sobre as vias, regiões e edifícios, sendo cada um destes tipos de dados temas de informação com características diferentes.

A Figura 4.6, apresenta os objetos que definem uma fonte de dados no GeoTools. De forma a estabelecer uma analogia com a organização das fontes de dados apresentada na Figura 4.5, a interface DataStore

[3] http://www.opengis.com/

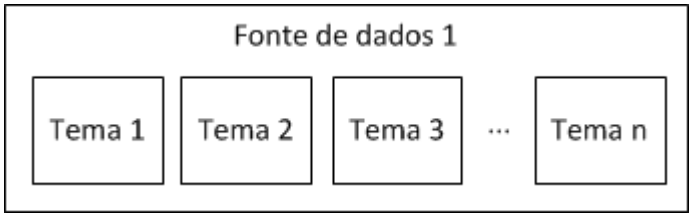

Figura 4.5. Organização das fontes de dados.

corresponde à fonte de dados e a interface FeatureSource corresponde ao tema.

A interface FeatureReader representa um iterador permitindo navegar sobre as *features* do conjunto. O iterador é obtido através de uma interrogação à implementação da interface FeatureSource.

Para a implementação de uma fonte de dados é necessário implementar as interfaces apresentadas. De forma a facilitar a implementação de uma fonte de dados, existem duas concretizações em classes abstratas das interfaces DataStore e FeatureSource: ContentDataStore e ContentFeatureSource, respetivamente.

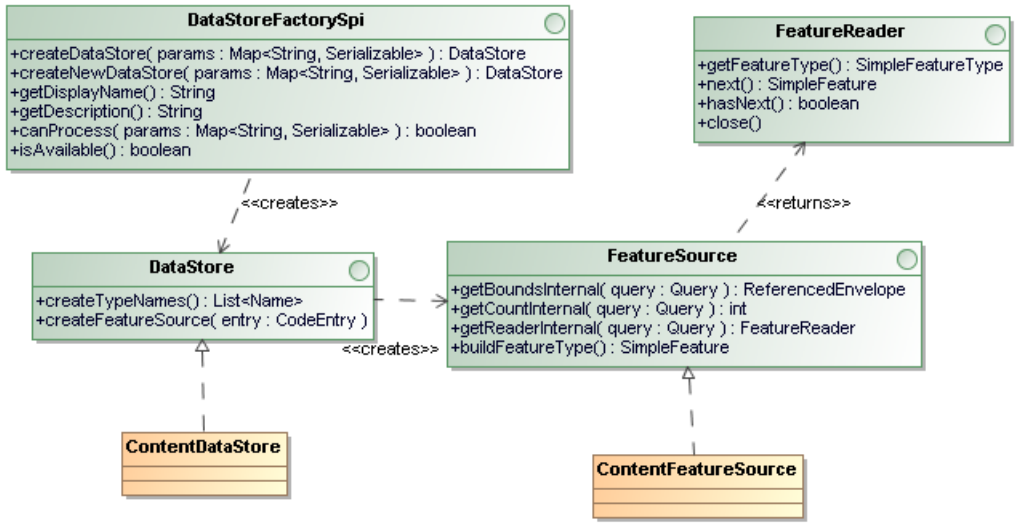

Figura 4.6. Representação dos objetos que definem uma fonte de dados.

A interface DataStoreFactorySpi, permite a instalação da fonte de dados como um *plugin*. Sendo necessária a criação do ficheiro no seguinte diretório com a informação da Listagem 4.3:

- META-INF/services/org.geotools.data.DataStoreFactorySpi.

```
{package}.{className}
```

Listagem 4.3. Exemplo do conteúdo do ficheiro DataStoreFactorySpi.

O ficheiro deverá conter o nome qualificado da *factory* que instancia a fonte de dados, sendo descoberta e instanciada em tempo de execução de forma dinâmica.

Para a instalação da fonte de dados no GeoServer, é necessário gerar um ficheiro JAR com a fonte de dados e as suas dependências. De seguida, este deverá ser copiado para a diretoria /lib do GeoServer, ficando assim instalado.

4.3 OpenLayers

O OpenLayers é uma *framework Javascript* que permite a visualização de mapas num *browser*. O OpenLayers foi desenvolvido possibilitando o consumo de informação através dos standards WMS e WFS definidos pelo OGC (ver secção 3.5 e 3.7 do Capítulo 3). [31]

Para a obtenção das imagens e informação geográfica, é necessária a utilização de um servidor de mapas, sendo utilizado o GeoServer (ver secção 4.1). A informação disponibilizada pelo GeoServer será posteriormente obtida pelo OpenLayers através de pedidos HTTP, para a representação da informação através de imagens ou de forma vetorial. [31]

Para a representação de um tema sobre o mapa, a Listagem 4.4 apresenta um exemplo da utilização do OpenLayers para a representação da informação.

```
var options = {
  controls: [],
  maxExtent: bounds,
  maxResolution: 0.0979640270761948,
  projection: "EPSG:4326",
  units: 'degrees'
};
```

```
var map = new OpenLayers.Map('div_element', options);
var tiled = new OpenLayers.Layer.WMS(
  "concelhos",
  "http://localhost:8080/geoserver/statistics/wms",
  {
    LAYERS: 'concelhos',
    STYLES: '',
    format: 'image/png',
    tiled: true,
    tilesOrigin : map.maxExtent.left + ',' + map.maxExtent.bottom
  },
  {
    buffer: 0,
    displayOutsideMaxExtent: true,
    isBaseLayer: true,
    yx : {'EPSG:4326' : true}
  }
);
map.addLayers([layer]);
```

Listagem 4.4. Exemplo da utilização do OpenLayers para a representação de um tema obtido através dos serviços do GeoServer.

No exemplo da Listagem 4.4, é adicionado um mapa à página, devendo este ser desenhado dentro de um contentor com o id: 'div_element'. Para a representação do mapa, é indicado o sistema de projeção, a unidade de medida e a resolução máxima. Esta informação será enviada no pedido ao GeoServer, sendo este capaz de reprojetar a informação se necessário.

De seguida, é criado um objeto para a representação da informação disponibilizada no serviço WMS. Este objeto irá efetuar pedidos ao servidor, para o tema concelhos, dispondo posteriormente as imagens sobre o mapa.

A Figura 4.7, apresenta o resultado da execução do código presente na Listagem 4.4, sendo representado um mapa com os concelhos de Portugal.

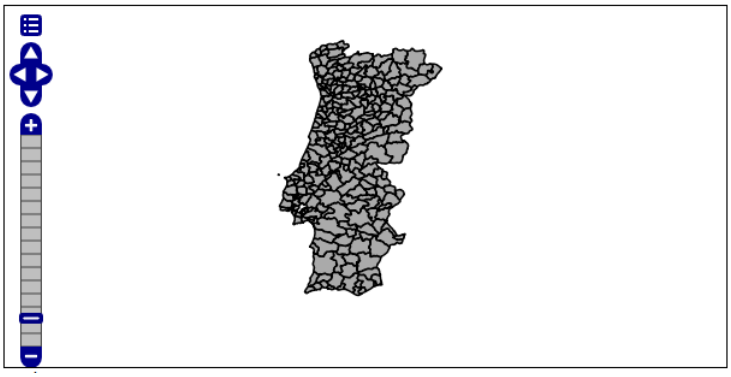

Figura 4.7. Representação de um mapa com os concelhos de Portugal.

5

Implementação

Neste Capítulo é apresentada e implementada uma proposta para a resolução do problema apresentado no Capítulo 1. A arquitetura geral da solução e todas as peças constituintes serão aqui apresentadas.

Ao longo do capítulo são também debatidos alguns aspetos da implementação, assim como as decisões tomadas na abordagem ao problema.

5.1 Desafios

Um dos grandes desafios para a implementação da solução é a integração de dados estatísticos de várias fontes de dados, permitindo a sua análise e cruzamento a nível aplicacional. Deverá ser possível a integração posterior de novas fontes de dados sem ser necessário realizar alterações ao sistema em funcionamento. Com um exemplo concreto, pretende-se cruzar indicadores do INE com indicadores do Banco de Portugal, sendo posteriormente adicionados ao sistema indicadores sobre o desemprego, provenientes de uma nova fonte de dados.

Para além do problema de integração, existe ainda para algumas fontes de dados a questão da propriedade dos dados. Ou seja, o INE ou o Banco de Portugal, por exemplo, disponibilizam os dados gratuitamente para consulta, mas estes não podem ser copiados para outro sistema, sendo requisito que permaneçam nos sistemas de origem. Desta forma, não é possível consolidar os dados numa única fonte de dados e posteriormente apresentá-los. No entanto, pretende-se analisar estes dados e compará-los entre si.

Para além das questões relacionadas com a propriedade dos dados, existem também questões relacio-

nadas com a privacidade dos dados. Estas questões, embora relevantes, não se encontram no âmbito do projeto e ficarão por abordar.

5.2 Arquitetura geral

De forma a corresponder aos objetivos apresentados na secção 1.3 do Capítulo 1, e tendo em conta os desafios apresentados na secção 5.1, a Figura 5.1 apresenta a arquitetura global da solução. Como representado na legenda da Figura 5.1, com uma tonalidade mais escura encontram-se os componentes externos ao sistema e fora do âmbito do projeto. Os componentes externos ao sistema, são utilizados como parte integrante da solução tendo sido desenvolvidos por terceiros.

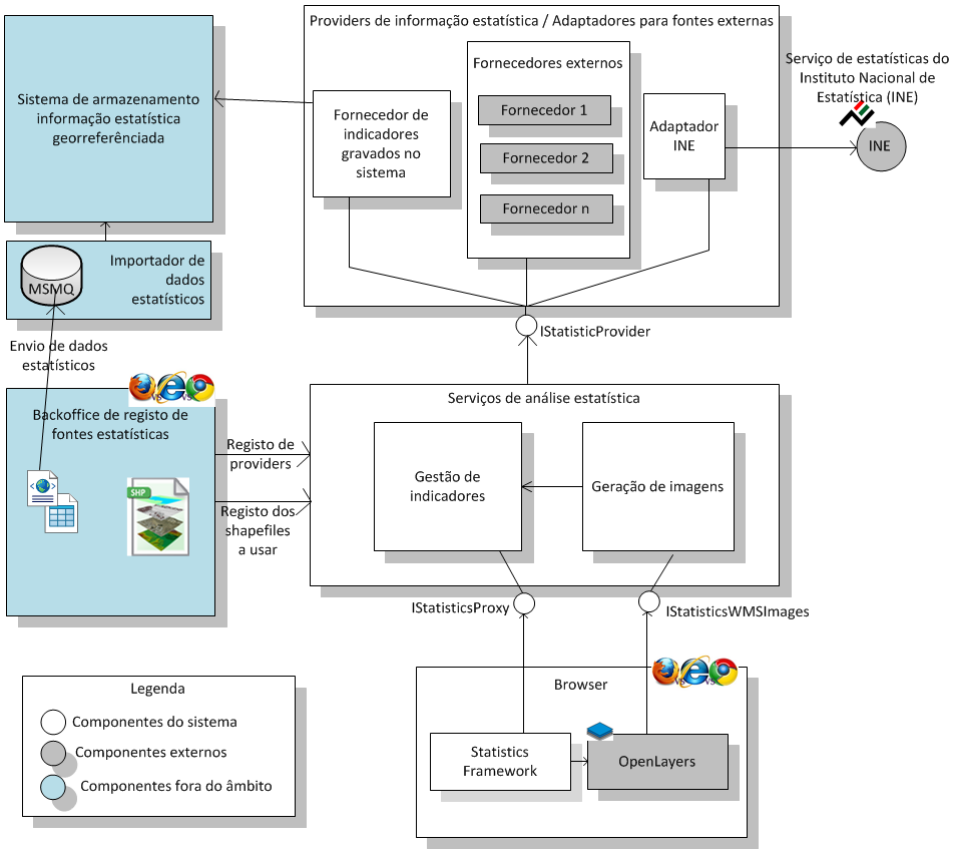

Figura 5.1. Arquitetura geral do sistema de análise estatística de informação georreferenciada.

Analisando apenas os blocos à direita na Figura 5.1, é possível verificar que a arquitetura está organizada em três camadas:

- Camada de apresentação (*Browser*), para a qual foi desenvolvida uma *framework Javascript* de forma a facilitar a integração do sistema em aplicações *Web*;
- Camada de negócio (Serviços de análise estatística), para integração de indicadores. A esta camada foram atribuídas as seguintes responsabilidades:
  - Gestão das fontes de dados disponíveis;
  - Gestão dos indicadores de cada uma das fontes de dados;
  - Intermediação da comunicação com fontes de dados. A comunicação com as fontes de dados é efetuada através desta camada;
  - Geração de imagens temáticas a partir dos dados dos indicadores.
- Camada de dados (*Providers* de informação estatística). Esta é a camada que permite o acesso aos dados dos indicadores estatísticos, unificando a interface de comunicação. Os subsistemas presentes nesta camada poderão ser adaptadores para sistemas já existentes, convertendo a interface dos sistemas de origem para a interface comum do sistema. Esta abordagem permite incorporar os indicadores do INE no sistema.

De forma a permitir que os dados permaneçam nos sistemas de origem, a comunicação entre as várias camadas é feita através de serviços. A comunicação entre serviços é efetuada com SOAP [32] sobre HTTP, permitindo interoperar entre sistemas e tecnologias [33].

Para a integração de dados estatísticos no sistema, um fornecedor deverá implementar o contrato definido na secção 5.4. De seguida, o serviço deverá ser publicado na *Web* e posteriormente registado no sistema.

A abordagem orientada a serviços, para além da possibilidade de integração de novas fontes de dados, possibilita ainda o desenvolvimento de outras aplicações para exploração dos dados estatísticos. Por exemplo, poderá ser desenvolvida uma aplicação móvel que possibilite a visualização destes dados.

Uma desvantagem, tendo por base a orientação a serviços, encontra-se no facto de por cada interação ser realizado um pedido HTTP com toda a verbosidade inerente à mensagem SOAP [34]. Devido à importância dos objetivos a alcançar, na secção 5.5.2 será discutida uma forma de minimizar o impacto deste problema. No entanto, este é um compromisso que será assumido tendo em vista os seguintes pontos:

- Os dados poderão permanecer nos sistemas de origem;
- Interoperabilidade entre sistemas, possibilitando o desenvolvimento de uma aplicação móvel para exploração dos dados ou definir uma fonte de dados numa tecnologia diferente.

Do lado esquerdo na Figura 5.1, foram identificados os seguintes subsistemas que se consideram essenciais para a disponibilização da solução como um produto:

- Subsistema para armazenamento de indicadores (Sistema de armazenamento de informação estatística);
- Subsistema de importação de informação estatística (Importador de dados estatísticos);
- *Backoffice* de administração de fontes de dados e indicadores (*Backoffice* de registo de fontes estatísticas).

Estes subsistemas não se encontram implementados uma vez que se encontram fora do âmbito do projeto.

5.3 Suporte a *data warehouses* existentes

Os *data warehouses* existentes contêm indicadores com informação geográfica sob a forma alfanumérica, contendo o nome ou um código identificativo da zona geográfica a que se refere [18]. Assim, de forma a permitir a integração destes indicadores, são separados os conceitos de dados estatísticos e geográficos. Por um lado existe informação estatística, devolvida pelas fontes de dados, por outro, informação geográfica onde estão definidas as regiões e as suas delimitações.

De forma a desacoplar as fontes de dados da geografia associada à informação, é necessário georreferenciar a informação no processo de análise geográfica. A georreferenciação da informação associa os dados à definição de uma área geográfica.

5.3.1 Georreferenciação

O processo de georreferenciação dos dados estatísticos, consiste na associação do dado estatístico à definição de uma área geográfica. Esta associação é efetuada através do código associado à área geográfica com a informação alfanumérica de localização definida pela fonte de dados.

No entanto, as áreas geográficas, estão relacionadas com a forma como os dados foram recolhidos. Ou

seja, se foram efetuadas agregações para um indicador considerando delimitações das circunscrições administrativas do País, diferentes das delimitações definidas na Carta Administrativa Oficial de Portugal 2010 (CAOP) [1], então não será possível representar os dados geograficamente uma vez que não se conhecem as áreas.

Para que o utilizador possa efetuar agregações dos dados utilizando códigos ou regiões não oficiais, é necessário que para cada indicador sejam indicadas as áreas a utilizar.

Os formatos para a definição das áreas, suportando a georreferenciação, são apresentados na secção 3.6 do Capítulo 3, sendo que neste momento apenas são suportados *Shapefiles* no sistema (ver secção 3.6.1 do Capítulo 3).

5.4 Camada de dados

A camada de dados tem como objetivo fornecer o acesso aos dados dos indicadores estatísticos. O acesso aos dados é efetuado através de um serviço que deverá implementar o contrato definido na Figura 5.2.

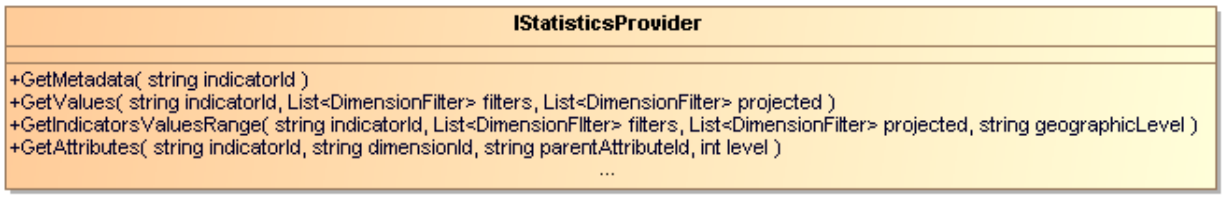

Figura 5.2. Contrato com a definição de um serviço da camada de dados.

O contrato do serviço da Figura 5.2, apresenta dois tipos de operações:

- Operações de configuração do indicador (GetMetadata e GetAttrubutes);
- Operações para interrogação e obtenção dos factos do indicador (GetValues e GetIndicatorValuesRange).

[1] http://www.igeo.pt/produtos/cadastro/caop/inicial.htm

5.4.1 Características do indicador

As operações de configuração devolvem as características do indicador, apresentando as dimensões e os atributos que o constituem, de forma a possibilitar as interrogações (ver secção 2.2 do Capítulo 2).

A operação GetMetadata tem como objetivo devolver todas as configurações associadas ao indicador, indicando as dimensões, os atributos de cada dimensão e informação sobre o indicador.

As dimensões podem ser hierárquicas podendo conter vários atributos. A dimensão geográfica é um exemplo de uma dimensão hierárquica. Esta pode conter vários níveis de agregação, formando estes níveis uma hierarquia (Distrito, Concelho e Freguesia), contendo assim apenas para Portugal mais de 5000 atributos. A operação GetAttributes, pretende diferir o carregamento dos atributos das dimensões para cada um dos níveis da hierarquia, por exemplo, estando a visualizar o nível de distrito, não é necessário ter todos os concelhos previamente carregados, sendo estes carregados à medida que se expande cada um dos distritos, para a visualização do nível inferior da hierarquia.

O contrato de dados para suporte às configurações do indicador é apresentado na Figura 5.3, sendo definido pela seguinte informação:

- Identificador do indicador;
- Nome do indicador;
- Se possível, um URL para visualização do indicador no site da fonte de dados;
- Dimensões, respetivos códigos e nomes;
- Hierarquia de atributos das dimensões;
- Indicação se o carregamento dos atributos da dimensão hierárquica deverá ser diferido.

Na versão atual do sistema, para a exploração de dados, apenas são suportados atributos categóricos (ver secção 2.1 do Capítulo 2), sendo assumido pelo sistema que todos os atributos devolvidos são do tipo categórico.

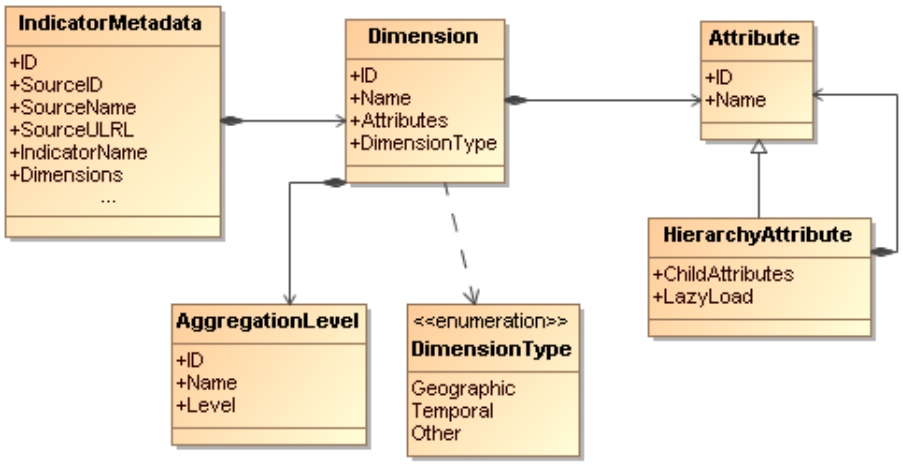

Figura 5.3. Definição dos metadados de um indicador.

5.4.2 Factos do indicador

As operações que devolvem os factos dos indicadores separam, à semelhança da linguagem MDX [35], as dimensões para filtrar e limitar o conjunto de análise, e as dimensões a projetar, sobre as quais deverão ser efetuadas as agregações. Por exemplo, imagine-se as seguintes características do indicador A:

- Dimensão Ano com os atributos: 2011, 2012;
- Dimensão Género com os atributos: Masculino e Feminino.

Considerando todos os atributos do indicador A, existem quatro factos diferentes:

- Ano 2012 Género Masculino -> Facto A;
- Ano 2012 Género Feminino -> Facto B;
- Ano 2011 Género Masculino -> Facto C;
- Ano 2011 Género Feminino -> Facto D.

O exemplo anterior, apresenta a projeção dos atributos da dimensão Ano e Género. No entanto, caso se pretenda projetar apenas os atributos da dimensão Género, é necessário efetuar uma agregação de modo a que os atributos da dimensão projetada não se repitam na resposta, sendo devolvida a seguinte informação:

- Género Masculino, filtrado para Ano 2011 e 2012 -> Facto E (Agregado);

- Género Feminino, filtrado para Ano 2011 e 2012 -> Facto F (Agregado).

Desta forma, a agregação dos factos de análise é da responsabilidade das fontes de dados, com a aplicação da função de agregação. A Figura 5.4, apresenta o contrato de dados para a devolução dos factos de um indicador.

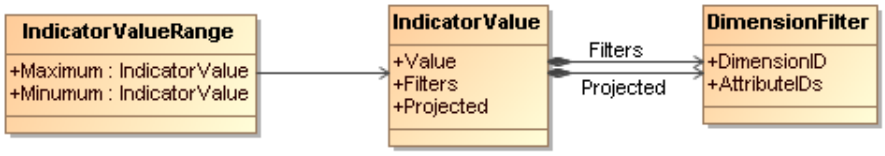

Figura 5.4. Definição da agregação de valores de um indicador.

Na versão atual do sistema, apenas são suportados factos aditivos, sendo assumido que se poderão efetuar todas as combinações dos atributos de forma a agregar a informação (ver secção 2.3 do Capítulo 2).

O serviço de dados deverá suportar comunicação através de mensagens SOAP sobre HTTP, possibilitando a sua integração no sistema e registo na camada de negócio.

5.5 Gestão de indicadores

O componente para a gestão de indicadores da Figura 5.1, é um módulo da camada de negócio sendo criado com o propósito de integrar os diferentes *providers* de informação estatística na aplicação. Esta camada foi implementada segundo o padrão de software Facade [36], constituindo o ponto único de acesso à informação estatística das várias fontes de dados e assegurando diversas funções no sistema:

- Gestão e configuração de fontes de dados e indicadores;
- Intermediação da comunicação com as fontes de dados;
- Cache das respostas das fontes de dados;
- Autorização e autenticação do utilizador no acesso aos dados.

A Figura 5.5, apresenta a arquitetura do módulo responsável pela gestão dos indicadores e comunicação com as fontes de dados, para a obtenção dos dados dos indicadores registados.

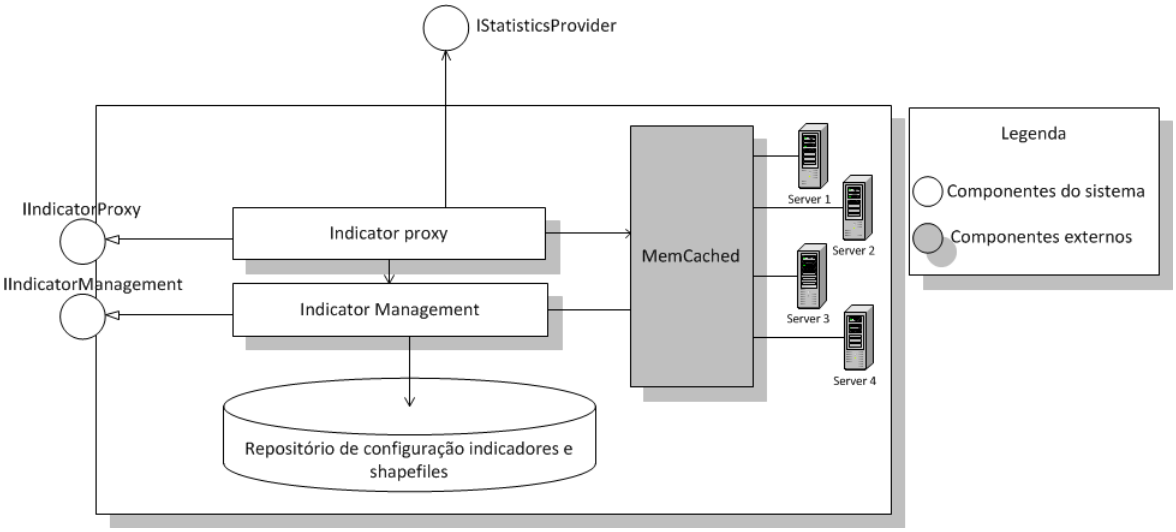

Figura 5.5. Arquitetura da gestão de fontes de dados e indicadores.

Este módulo disponibiliza uma interface de comunicação com as fontes de dados e uma interface para gestão das fontes de dados e indicadores estatísticos, possibilitando a adição de novas fontes de dados e indicadores.

Para suporte ao registo de fontes de dados e configuração da representação dos seus indicadores, foi implementado o Repositório de configuração de indicadores e *Shapefiles*, responsável pelo armazenamento das configurações associadas a cada indicador e às áreas geográficas para a georreferenciação dos dados no processo de análise geográfica.

O módulo de *cache*, representado na Figura 5.5 pelo nome de "MemCached"[2], está representado com uma tonalidade mais escura, uma vez que é uma peça que foi desenvolvida no contexto de um projeto *open-source* sendo utilizado no sistema. Este módulo é discutido na secção 5.5.2.

5.5.1 Configuração de indicadores e fontes de dados

De forma a integrar os indicadores de uma fonte de dados no sistema, é necessário indicar um conjunto de configurações. A configuração contém a seguinte informação sobre os indicadores e fontes de dados:

[2] http://memcached.org/

- URL onde o serviço da fonte de dados está disponível (O serviço da fonte de dados, deverá implementar o contrato definido na secção 5.4);
- Indicação se as respostas do serviço poderão ser colocadas em *cache*, e qual o tempo de *cache* associado;
- Indicação dos níveis geográficos de agregação da informação e associação de cada um dos níveis a um *Shapefile*, de forma a possibilitar a geração de imagens temáticas.

Para cada um dos indicadores da fonte de dados, de forma a permitir desacoplar as fontes de dados da definição das áreas geográficas, é necessário que sejam identificados os níveis de agregação de cada indicador (ver secção 5.3). A cada um dos níveis de agregação disponíveis, deverá ser associado um *Shapefile* possibilitando a georreferenciação dos dados. Cada *Shapefile*, deverá conter a seguinte informação:

- Identificador da área geográfica, coincidente com o identificador devolvido nos dados;
- Definição da área geográfica;
- Nome da área geográfica.

As áreas definidas no *Shapefile* deverão estar definidas no sistema de coordenadas WGS84 (ver secção 3.3 do Capítulo 3).

5.5.2 *Cache*

Existem indicadores cuja informação está em constante alteração, sendo por isso impeditivo a utilização de *caches* uma vez que o utilizador pretende visualizar sempre a informação mais atual. No entanto, existem indicadores com informação com períodos de atualização mais longos. Desta forma, a informação destes indicadores pode ser colocada em *cache* tendo em conta a sua fraca volatilidade.

De forma a aumentar o desempenho global do sistema, evitando assim a consulta à fonte de dados, sempre que a configuração o permitir, os dados de um indicador deverão ser colocados em *cache* para reutilização futura. Pedidos subsequentes sobre os mesmos dados, não terão de consultar a fonte de dados.

A *cache* de dados, ao nível do subsistema de gestão de indicadores, embora esteja prevista na arquitetura, não se encontra implementada, tendo sido prevista a utilização de MemCached para gestão dos dados em *cache* com um tempo de vida associado, funcionando de forma distribuída. [37]

5.5.3 Autenticação e autorização

Embora seja considerado um tópico fora do âmbito do projeto, a autenticação e autorização são um tema bastante relevante na visualização de informação estatística. Dependendo dos indicadores, podemos estar a lidar com informação sensível que não se pretende expor ao público. Por exemplo, indicadores sobre o património, se for possível visualizar a informação a um nível geográfico muito detalhado, por exemplo o lote, em alguns casos seria possível a identificação dos indivíduos.

A informação estatística a apresentar necessita de ser filtrada de acordo com o utilizador, sendo a camada de negócio, neste sistema, o ponto para a implementação da autenticação e autorização no acesso aos dados. Este mecanismo de autorização poderá ser implementado de acordo com o mecanismo de papeis e utilizadores definidos pelo modelo RBAC [38].

Nesta solução, a fonte de dados permite ao sistema o acesso à informação, sendo esta devidamente filtrada na camada de negócio quando requisitada pelos utilizadores.

5.6 Geração de imagens temáticas

Para a geração das imagens temáticas existem duas opções de representação:

- Renderização da informação no cliente, através das funcionalidades suportadas pelo *browser*;
- Geração de uma imagem com a informação estatística no servidor.

Para a renderização da informação no cliente é necessário a obtenção dos dados estatísticos e das áreas correspondentes a cada um dos dados, procedendo posteriormente à representação da informação. No entanto, se a área de análise for vasta, por exemplo a Europa, será transferida uma grande quantidade de informação que será mantida em memória pelo *browser* para a representação das áreas.

De forma a diminuir a quantidade de informação devolvida em cada pedido, a representação poderia ser segmentada, ou seja, a área de análise é dividida numa grelha, tal como ilustrado na Figura 5.6, sendo efetuado um pedido por cada quadricula. Esta solução apenas diminui a quantidade de informação devolvida em cada pedido, tendo o *browser* de manter a informação em memória de forma a suportar o desenho das áreas.

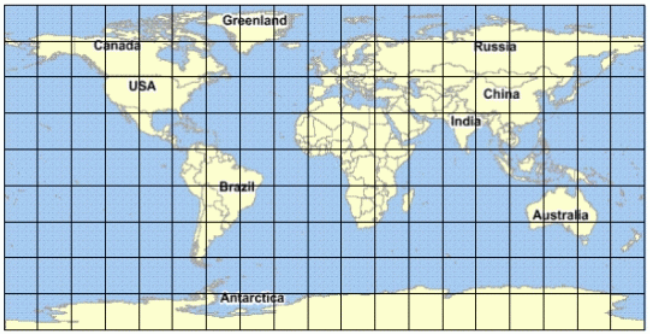

Figura 5.6. Segmentação da área de representação.

Uma alternativa ao desenho da informação no cliente, é a utilização da estratégia de segmentação da informação, ilustrado na Figura 5.6, com a geração da imagens temáticas no servidor. Com esta abordagem, é reduzida a quantidade de informação devolvida em cada pedido e a quantidade de informação que o *browser* mantém em memória na representação das áreas. Assim, para a representação da informação no cliente, é usada a técnica da geração de imagens temáticas no servidor.

A Figura 5.7 apresenta a arquitetura para a geração das imagens temáticas para os diferentes níveis geográficos.

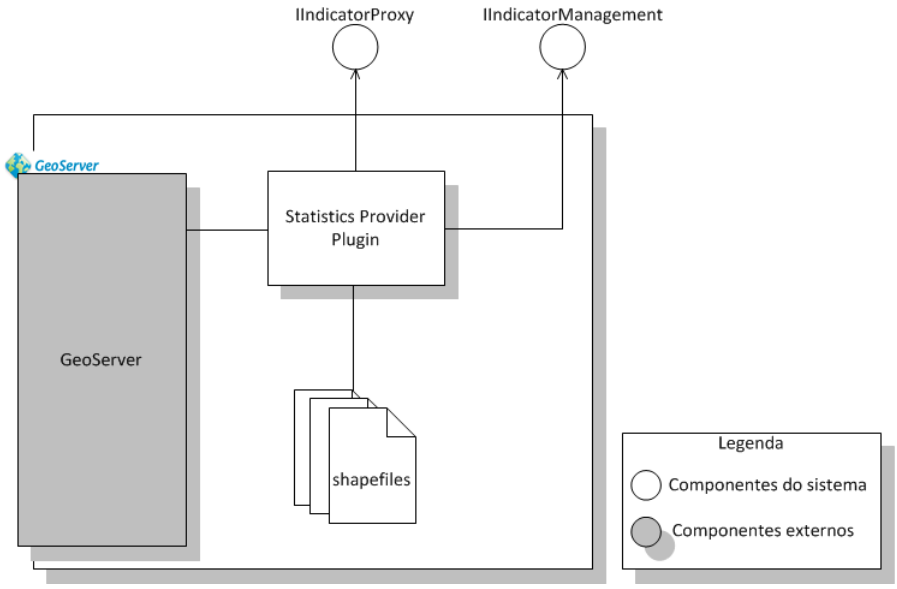

Figura 5.7. Arquitetura da geração de imagens temáticas.

Com um tom mais escuro, está representado o GeoServer, responsável pelo desenho das imagens temáticas e disponibilização da informação como um serviço (ver secção 4.1 do Capítulo 4).

Para a geração das imagens temáticas, o GeoServer necessita da seguinte informação:

- As áreas;
- Os dados estatísticos;
- A associação entre as áreas e os dados estatísticos.

A junção desta informação é assegurada pelo *plugin* "Statistics Provider Plugin"que deverá ser instalado no GeoServer (ver secção 4.2.1 do Capítulo 4), para a geração das imagens temáticas.

5.6.1 Statistics Provider Plugin

O Statistics Provider Plugin é a implementação de uma fonte de dados do GeoTools, com o objetivo de juntar a informação estatística com a informação geográfica. A associação entre os dados estatísticos e os dados geográficos é posteriormente devolvida ao GeoServer para a geração das imagens (ver secção 4.2.1 do Capítulo 4).

Os dados devolvidos ao GeoServer pela fonte de dados, são *features* com o seguinte esquema:

- Polígono com a definição da área geográfica;
- Nome da área geográfica;
- Identificador da área geográfica;
- Dimensões usadas para filtrar os dados;
- Dimensões usadas para identificar a agregação/projeção de dados;
- Identificador do indicador;
- Identificador da fonte de dados;
- Nível de agregação da informação;
- Valor absoluto do dado estatístico;
- Valor relativo do dado estatístico em percentagem.

Os filtros aplicados à informação e o estilo aplicado no GeoServer, usam os campos definidos no esquema.

Para a aplicação de estilos, são usados os campos do valor absoluto e do valor relativo, em percentagem, que através da linguagem de definição de estilos SLD, definidos no GeoServer, influenciam a representação da área geográfica (ver secção 3.5.3 do Capítulo 3).

Para filtrar a informação, são usados os campos para identificar o indicador, a fonte de dados, nível de agregação, dimensões para filtrar e agregar a informação. Os filtros são indicados ao GeoServer através da linguagem de interrogação CQL (ver secção 4.1.1 da Capítulo 3), sendo posteriormente aplicados à fonte de dados.

A Figura 5.8 apresenta a arquitetura do *plugin*, para a devolução dos dados, possibilitando a geração das imagens temáticas. Com um tom mais escuro, é representado o GeoTools, componente a que será adicionado o *plugin* desenvolvido (ver secção 4.2 do Capítulo 4).

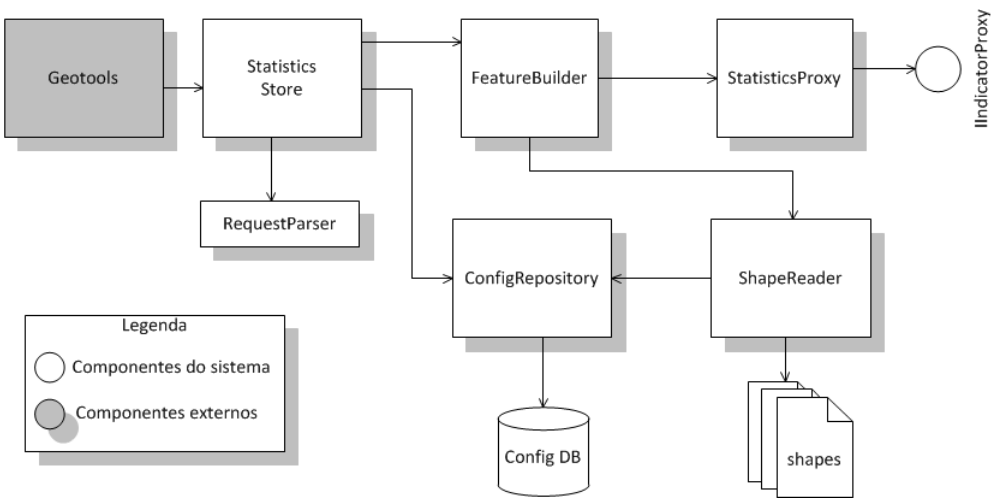

Figura 5.8. Arquitetura da geração de imagens temáticas.

A arquitetura do *plugin* apresenta a separação lógica dos objetos em *packages*, sendo a divisão feita de acordo com a responsabilidade de cada um dos objetos.

O *package* "Statistics Store", contém a definição do plugin, estendendo assim a funcionalidade do GeoTools e permitindo a instalação do *plugin* no GeoServer. Por cada pedido, é feita a separação dos filtros, de acordo com o esquema das *features* no componente RequestParser.

A configuração das áreas geográficas, contendo os níveis de agregação e localização dos dados, para cada indicador, é obtida através dos objetos do *package* ConfigRepository, podendo as áreas geográficas ser acedidas através dos objetos do *package* ShapeReader. No sistema, apenas são suportados *Shapefiles* para a definição das áreas geográficas. No entanto, estes componentes foram desenvolvidos de forma a possibilitar a sua extensão, podendo suportar novos formatos de dados (ver secção 3.6 do Capítulo 3).

Para a construção das *features*, são usados os objetos do *package* FeatureBuilder, que contém a definição do esquema das *features* e associa as áreas geográficas aos dados estatísticos.

Para representação da informação, o *plugin* é adicionado ao GeoServer, sendo apenas gerados mapas coropletos com o método de intervalos iguais (ver secção 2.5 do Capítulo 2) e usadas sete classes para representação da informação. No entanto, tal como o suporte às áreas geográficas, é possível estender a funcionalidade para suportar novos tipos de mapas temáticos e métodos. Para a alteração do número de classes, de acordo com a distribuição dos dados, é necessária a geração dinâmica do estilo.

5.7 Camada de apresentação

A Camada de apresentação é responsável pela representação da informação estatística e exploração da componente geográfica dos indicadores.

Para a visualização da informação estatística, foi desenvolvida uma *framework* em *Javascript*, permitindo a reutilização da tecnologia, possibilitando a sua integração numa aplicação *Web*. A *framework* oferece a visualização dos dados através de mapas temáticos (ver secção 2.4), gráficos [3] e tabelas *pivot*.

A arquitetura da *framework*, é apresentada na Figura 5.9. A arquitetura da camada de apresentação está separada em vários blocos, separando responsabilidades e fornecendo isolamento a cada um dos blocos. A interligação dos blocos é feita através de interfaces ou em alguns casos através do BUS de eventos global.

A implementação da *framework* de visualização de dados estatísticos, Figura 5.9, tem por base o padrão de software MVC [39]. Os modelos são responsáveis por transportar a informação do servidor, e são

[3] Gráficos circulares e gráficos de barras

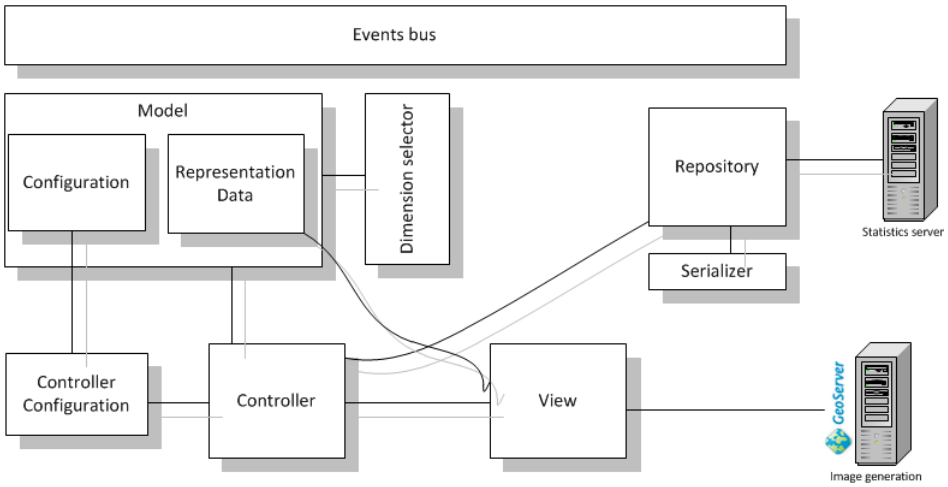

Figura 5.9. Arquitetura da camada de apresentação.

partilhados entre vistas e controladores. Os controladores são responsáveis pela coordenação global do sistema, obtendo novos dados e atualizando as vistas. As vistas têm como responsabilidade a representação visual dos modelos.

O BUS de eventos é global à *framework* e é utilizado para o envio e receção de notificações, interligando os vários componentes diminuindo o acoplamento entre as partes. Sempre que ocorra um evento que um componente pretenda publicar, é enviada uma mensagem para o BUS de eventos, podendo assim outros componentes serem notificados. Por exemplo, se ocorre uma ação numa vista que deva implicar o carregamento de mais informação, é enviado um evento para o BUS que se encontra subscrito por um controlador. O controlador ao receber o evento, consulta a informação e atualiza a vista.

A comunicação com o servidor é efetuada através da camada de repositório, sendo configurado o endereço e o seriador da informação a enviar. O papel do seriador é preparar a informação a enviar ao serviço num formato que este consiga entender, convertendo a informação dos modelos para a definição de dados do serviço.

Os mapas temáticos são um caso especial relativamente à comunicação com o servidor, uma vez que se tratam de imagens, apenas será configurado o URL da imagem sendo o *browser* a efetuar o pedido da informação.

5.7.1 Pedidos a domínios diferentes do domínio da aplicação

O desenvolvimento de uma API *Javascript* permite ao utilizador integrar a ferramenta de análise estatística na sua aplicação, possibilitando a exploração de indicadores. A *framework* encarrega-se de obter os dados provenientes dos serviços e apresentá-los ao utilizador de forma transparente para a aplicação. De notar que a aplicação que pretenda integrar a ferramenta de análise estatística, será um site alojado na *internet* em qualquer domínio.

Uma vez que se tratam de aplicações *Web* com interface para *browser*, é necessário ter em consideração as limitações impostas pelas várias versões dos *browsers*. Nas versões dos *browsers* (e.g: Internet Explorer 6, 7 e 8, Firefox 2 e 3 e Google Chrome), anteriores às que suportam HTML5, existem restrições de segurança que inibem os pedidos AJAX a um domínio diferente do domínio do site [40][41]. Estas restrições de segurança impossibilitam a reutilização da *framework* em vários sites, uma vez que não seria possível obter a informação. No entanto, uma vez que existem mais *frameworks* com necessidades semelhantes, foram encontradas, pela comunidade, várias soluções de forma a contornar o problema.

Uma solução possível é a configuração de um proxy no servidor do site, sendo os pedidos encaminhados para o *endpoint* do serviço. Assim, os pedidos AJAX são relativos ao domínio do site e a restrição de segurança do *browser* não impede o pedido. Esta solução tem a desvantagem de obrigar à existência de um servidor HTTP, não podendo o utilizador utilizar páginas estáticas.

A comunicação através de uma iframe, é outra alternativa. No entanto, devido à mesma restrição de segurança [40], uma vez que se trata de um domínio diferente, não seria possível aceder ao conteúdo da iframe, sendo esta apenas uma solução para o envio de informação.

Por fim, a solução encontrada baseia-se no tipo de conteúdo bloqueado pelos *browsers* para garantir que são cumpridos os requisitos de segurança. Os *browsers*, de acordo com a restrição [40] bloqueiam apenas os pedidos AJAX através da utilização dos objetos XMLHttpRequest e ActiveXObject, em pedidos a domínios diferentes do domínio do site. No entanto, não colocam qualquer restrição quanto à obtenção de outros recursos a partir de domínios diferentes, por exemplo, imagens e ficheiros *script* (utilizando as *tags* img e *script* respetivamente). Uma vez que estes conteúdos não são bloqueados, esta estratégia consiste em fazer pedidos a ficheiros *Javascript* (gerados dinamicamente), e interpretar o seu conteúdo. Com um

exemplo concreto, pretende-se ilustrar a solução:

Considere-se a existência de um serviço com a seguinte especificação:

- O serviço encontra-se instalado no endereço: http://www.myservice.com;
- Aceita um parâmetro com o nome *callback* que define o nome da função a evocar quando o *script* é carregado;
- É devolvido um ficheiro de *script* por cada invocação.

Se o código da listagem 5.1, for incluído numa página:

```
<script type="text/javascript" src="http://www.myservice.com?callback=foo"></
    script>
```

Listagem 5.1. *Tag* de *script* a incluir na chamada ao serviço

O serviço devolve um ficheiro com o conteúdo definido na listagem 5.2

```
1 foo("Hello, world!");
```

Listagem 5.2. Resposta do serviço ao pedido pelo recurso

Para que a comunicação possa ser efetuada, a função *foo* (indicada no parâmetro) deverá estar previamente definida na página, estabelecendo-se assim a comunicação com o servidor.

Esta abordagem, foi a abordagem escolhida para a implementação da comunicação com os serviços, possibilitando a reutilização da *framework* em diferentes domínios. No entanto, esta abordagem tem também alguns problemas, nomeadamente:

- Apenas é possível realizar pedidos com o verbo HTTP GET;
- Na sua implementação alguns *browsers* limitam o tamanho da *query string*, variando este de *browser* para *browser* [4].

Os *browsers* mais recentes que implementam a norma de HTML5, já não têm esta limitação podendo efetuar pedidos a outros domínios desde que devidamente autorizados [41]. De acordo com a arquitetura da camada de apresentação, o utilizador pode definir uma nova implementação para o repositório de comunicação, uma vez que na implementação as dependências estão ao nível das interfaces. Assim, é

[4] http://www.boutell.com/newfaq/misc/urllength.html

possível a implementação de um repositório que tire partido das funcionalidades disponibilizadas em *browsers* mais recentes com suporte a HTML5.

5.7.2 Exploração geográfica

A exploração geográfica dos dados é efetuada através de mapas temáticos. Para a visualização dos dados sobre o mapa, é usado o OpenLayers (ver secção 4.3 do Capítulo 4) sendo adicionada uma camada temática com os dados a explorar.

Na Figura 5.9 que define a arquitetura, existe um controlador para o mapa, que faz a gestão dos filtros selecionados e do nível de agregação da informação e uma vista responsável pela apresentação da informação. A vista é composta por um componente OpenLayers que apresenta a informação e uma legenda. O controlador tem a responsabilidade de coordenar o mapa com a respetiva legenda, sendo os pedidos pelas imagens temáticas geridos pelo OpenLayers.

Como camada base na informação do mapa, definindo estradas, ruas e edifícios, é usada a informação do openstreetmap [5]. Para colocar uma camada temática sobre a informação do openstreetmaps, é necessário que a nova camada de informação tenha o mesmo sistema de coordenadas e a mesma projeção, de forma que as imagens possam ficar alinhadas sobre o mapa (ver secção 3.7 do Capítulo 3). Assim, para a representação das imagens temáticas, é usado o sistema de coordenadas WGS84 com a projeção Mercator (ver secção 3.3 e 3.4.1 do Capítulo 3).

5.8 Armazenamento de informação estatística

O subsistema de armazenamento de informação estatística, embora não se encontre no âmbito do projeto, está previsto na arquitetura global da Figura 5.1. Este subsistema é responsável pelo armazenamento e consolidação dos dados de várias fontes, nos casos onde não se coloque a restrição da propriedade dos dados.

A existência de um sistema comum para armazenamento de informação estatística, poderá viabilizar o uso desta plataforma por pequenos negócios, uma vez que não terão de suportar os custos de um sistema de armazenamento de informação, bem como os custos de desenvolvimento de um serviço de consulta

[5] http://www.openstreetmap.org/

da informação. No entanto, se for pretendido é possível que os dados permaneçam num outro sistema de informação, podendo ser integrados através de serviços.

Para a integração deste subsistema, poderá ser usada a estratégia das fontes de dados, onde seria desenvolvido um serviço com a capacidade de consulta dos dados presentes neste subsistema.

Para o desenvolvimento deste subsistema, é necessário ter em consideração alguns aspetos relacionados com os dados, nomeadamente:

- Características dos dados, definidos na secção 2.1 do Capítulo 2;
- Definição das dimensões de análise e respetivas hierarquias (ver secção 2.2 do Capítulo 2);
- Definição do tipo de facto (ver secção 2.3.1 do Capítulo 2).

6

Conclusão

De acordo com o crescimento exponencial da informação, a análise estatística geográfica fornece uma análise EDA sobre os dados, possibilitando a identificação de padrões e comparação de dados.

Como prova de conceito da solução proposta e implementada, foram utilizados os dados disponibilizados pelo INE, possibilitando a análise geográfica sobre os dados. Para a introdução dos dados do INE no sistema, foi desenvolvida uma fonte de dados que unifica a interface de comunicação, fazendo a intermediação na comunicação. Para cada um dos indicadores, foram identificados os níveis de agregação da informação, e associado um *Shapefile* a cada nível, contendo a definição das áreas geográficas.

Comparativamente com o SAPO Mapas e ArcGIS, este sistema possibilita a integração de indicadores estatísticos de diferentes fontes de dados, para análise geográfica.

Em comparação com o Pordata, devido ao suporte na análise geográfica dos dados, este é o sistema mais semelhante à solução implementada, suportando também análise geográfica sobre mapas coropletos. No entanto, em contraste a esta solução apresentada neste projeto, o Pordata renderiza a informação no browser, ao invés da utilização de imagens. Esta abordagem, enfraquece o desempenho do sistema, devido à grande quantidade de informação transferida e suportada em memória no cliente.

Comparativamente ao Excel, são suportadas menos representações de mapas temáticos, a solução implementada apenas suporta mapas coropletos. No entanto, foi seguida uma estratégia semelhante para a análise geográfica da informação. A informação estatística foi separada da informação geográfica, existindo posteriormente uma fase de georreferenciação dos dados, permitindo a associação da informação a uma localização geográfica.

6.1 Trabalho futuro

Como trabalho futuro, foram identificados os seguintes tópicos para a evolução do sistema:

- Sistema para armazenamento de indicadores;
- Suporte à análise geográfica com os vários tipos de mapas temáticos e métodos de cálculo de classes;
- Criação de indicadores com base na relação de indicadores existentes.

Identificado na secção 5.8 do Capítulo 5, um subsistema para armazenamento de indicadores estatísticos, seria um possível componente a adicionar ao sistema no futuro. Este sistema para além de garantir a independência das fontes de dados, poderia tal como o Excel, fornecer análise sobre um conjunto de dados fornecidos pelo utilizador.

A utilização dos vários tipos de mapas temáticos através da combinação do método de cálculo das classes, possibilitam a construção de mapas diferentes sobre o mesmo conjunto de dados (ver secção 2.4 do Capítulo 2). A representação da informação em diferentes tipos de mapas, oferece ao utilizador perspetivas diferentes sobre o mesmo conjunto de dados, permitindo ao utilizador validar e conhecer a distribuição dos dados. Num caso prático, na representação de dados sobre eleições, a apresentação do partido com o maior número de votos em cada região facilitaria a leitura. No entanto, esta representação implica representar para além do facto, a dimensão que identifica o partido, formando um mapa temático com a projeção de dois dados, o partido dominante na região e o facto.

Através de indicadores existentes e do cruzamento dos mesmos, é possível a criação de novos indicadores possibilitando fazer previsões da distribuição dos dados no futuro. Por exemplo, considerando os indicadores:

- Densidade populacional;
- Condições meteorológicas;
- Receitas de turismo.

Considerando que as receitas de turismo estão relacionadas com as condições meteorológicas e com a densidade populacional, de acordo com os factos passados em cada um destes indicadores, poderia ser criado um novo indicador com as previsões das receitas de turismo atendendo às condições meteorológicas previstas para os próximos dias.

Referências

1. F. B. Clyde W. Holsapple, *Handbook on Decision Support Systems 1: Basic Themes (International Handbooks on Information Systems)*. Springer, 2008.
2. S. Davis, *GIS for Web Developers: Adding 'Where' to Your Web Applications*. Pragmatic Bookshelf, 2007.
3. Ogc, "OpenGIS Styled Layer Descriptor Profile of the Web Map Service Implementation Specification." `http://www.opengeospatial.org/standards/sld`. Consultado em: 20/08/2012.
4. M. Pumphrey, "Geoserver user 's guide." Consultado em: 20/08/2012.
5. J. de la Beaujardiere, ed., *Web Map Service Implementation Specification, Version 1.1.1*. Open Geospatial Consortium Inc., 2002.
6. A. Bayer, H. Bittencourt, J. Rocha, and S. Echeveste, "A estatística e sua história," *XII Simpósio Sul-Brasileiro de Ensino de Ciências*, 2004.
7. P. Burrough, "Gis and geostatistics: Essential partners for spatial analysis," *Environmental and Ecological Statistics*, vol. 8, no. 4, pp. 361–377, 2001.
8. P. Costigan-Eaves and M. Macdonald-Ross, "William playfair (1759-1823)," *Statistical Science*, vol. 5, no. 3, pp. 318–326, 1990.
9. M. Friendly, "A brief history of data visualization," *Handbook of data visualization*, pp. 7–13, 2008.
10. W. Cleveland and R. McGill, "Graphical perception: Theory, experimentation, and application to the development of graphical methods," *Journal of the American Statistical Association*, pp. 531–554, 1984.
11. W. Cleveland and R. McGill, "Graphical perception: The visual decoding of quantitative information on graphical displays of data," *Journal of the Royal Statistical Society. Series A (General)*, pp. 192–229, 1987.
12. K. Johnston, J. Ver Hoef, K. Krivoruchko, and N. Lucas, *Using ArcGIS geostatistical analyst*, vol. 300. Esri Redlands, CA, 2001.
13. Y. Dodge, D. Cox, and D. Commenges, *The Oxford dictionary of statistical terms*. Oxford University Press, USA, 2006.
14. A. Gelman, "Exploratory data analysis for complex models," *Journal of Computational and Graphical Statistics*, vol. 13, no. 4, pp. 755–779, 2004.

15. L. Anselin and S. Rey, "Perspectives on spatial data analysis," *Perspectives on Spatial Data Analysis*, pp. 1–20, 2010.
16. D. Pyle, *Data Preparation for Data Mining (The Morgan Kaufmann Series in Data Management Systems)*. Morgan Kaufmann, 1999.
17. A. A. da Silva, *Gráficos e Mapas: representação de informação estatística*. Lidel, 2006.
18. R. Kimball and M. Ross, *The Data Warehouse Toolkit: The Complete Guide to Dimensional Modeling (Second Edition)*. Wiley, 2002.
19. T. Pedersen and C. Jensen, "Multidimensional database technology," *Computer*, vol. 34, no. 12, pp. 40–46, 2001.
20. E. Tufte, *The visual display of quantitative information*, vol. 7. Graphics press Cheshire, CT, 1983.
21. A. Briney, "Thematic maps: Thematic maps display data on a map." `http://geography.about.com/od/understandmaps/a/thematicmaps.htm`. Consultado em: 20/08/2012.
22. M. M. Fischer and J. Wang, *Spatial Data Analysis: Models, Methods and Techniques (SpringerBriefs in Regional Science)*. Springer, 1st edition. ed., Sept. 2011.
23. M. Neteler and H. Mitasova, *Open Source GIS: A GRASS GIS Approach*. Springer, 2007.
24. D. Maling, "Coordinate systems and map projections for gis," *Maguire, DJ*, 1991.
25. P. A. Vretanos, "Web Feature Service Implementation Specification," tech. rep., OGC, May 2005.
26. I. Esr, *ESRI Shapefile Technical Description*. Environmental Systems Research Institute, Inc., July 1998.
27. Open Geospatial Consortium, "OGC KML Standard - Version 2.2." http://www.opengeospatial.org/standards/kml. Consultado em: 20/08/2012.
28. C. Portele, "OpenGIS Geography Markup Language (GML) Encoding Standard (OGC 07-036)." OpenGIS Standard, Aug. 2007. Consultado em: 20/08/2012.
29. "OpenGIS catalogue services specification, version 2.0.0 with corregendum OpenGIS® implementation specification OGC 04-021r3.," 2005.
30. "Geotools user guide." Consultado em: 20/08/2012.
31. E. Hazzard, *OpenLayers 2.10*. Packt Publishing, Mar, 2011.
32. N. Mitra and Y. Lafon, "SOAP version 1.2 part 0: Primer (second edition)," tech. rep., W3C, Apr. 2007. http://www.w3.org/TR/2007/REC-soap12-part0-20070427/.
33. K. Ballinger, D. Ehnebuske, C. Ferris, M. Gudgin, C. K. Liu, M. Nottingham, and P. Yendluri, "Ws-i basic profile version 1.1," 2006. Consultado em: 20/08/2012.
34. D. Davis and M. Parashar, "Latency performance of soap implementations," in *Cluster Computing and the Grid, 2002. 2nd IEEE/ACM International Symposium on*, pp. 407–407, IEEE, 2002.
35. B. C. Smith, C. R. Clay, and H. Consulting, *Microsoft SQL Server 2008 MDX Step by Step*. Microsoft Press, 2009.
36. E. Gamma, R. Helm, R. Johnson, and J. Vlissides, *Design Patterns: Elements of Reusable Object-Oriented Software*. Addison-Wesley Professional, 1994.
37. B. Fitzpatrick, "Distributed caching with memcached," *Linux journal*, vol. 2004, no. 124, p. 5, 2004.

38. R. Sandhu, E. Coyne, H. Feinstein, and C. Youman, “Role-based access control models,” *Computer*, vol. 29, no. 2, pp. 38–47, 1996.
39. M. Fowler, *Patterns of enterprise application architecture*. Addison-Wesley Professional, 2003.
40. D. Broyle and H. Saiedian, “Security vulnerabilities in the same origin policy: Implications and alternatives,” *Computer*, no. 99, pp. 29–31,34, 2011.
41. A. van Kesteren, “Cross-origin resource sharing,” W3C working draft, W3C, Mar. 2009. http://www.w3.org/TR/2009/WD-cors-20090317/.