



Instituto Universitário de Lisboa

Escola de Tecnologias e Arquitetura
Departamento de Ciências e Tecnologias da Informação

**Visualização de Informação
e
Visualização Analítica:**

**Mapa de visualização gráfica da informação agregada do país,
um sistema de apoio à decisão**

Nuno Miguel Antunes Ferreira

Dissertação submetida como requisito parcial para obtenção do grau de
Mestre em Sistemas Integrados de Apoio à Decisão

Orientadores:
Professor Doutor Duarte Trigueiros
Departamento de Ciências e Tecnologias da Informação,
ISCTE-IUL

Setembro, 2012

Agradecimentos

Quando me lancei no desafio de fazer o Mestrado em Sistemas de Apoio à Decisão, procurei concretizar o desejo de continuar a aprender ao longo da vida. O presente trabalho resulta de muitas horas dedicadas a um tema que me apaixonou, mas que não teria conseguido sem o apoio de muitas pessoas à minha volta. Não sendo possível particularizar todos, não gostaria no entanto de deixar de agradecer especialmente a algumas pessoas.

Gostava de dedicar este trabalho à pessoa que foi a minha musa inspiradora, a minha avó, Maria das Neves, por me ter tornado uma pessoa melhor. Gostaria de agradecer aos meus pais, Albertina e Idílio pela aposta que sempre fizeram na minha educação e por tudo o que de bom me transmitiram ao longo da vida. À minha namorada, Irene pelo apoio e motivação para levar por diante o desafio de concluir a dissertação e pelas horas de ausência. À minha irmã, Sandra, e ao meu cunhado, Rui, pelo apoio durante todo o mestrado e aos meus sobrinhos, Afonso e Dinis, pelos sorrisos sinceros que me foram fazendo esquecer todas as dificuldades. Aos amigos Nuno Ferro, Bruno Melo, José Silva, Ricardo Silva e Pedro Maia pelo apoio, troca de ideias e palavras de encorajamento. Ao João Martins e particularmente ao Rui Carneiro, pelo apoio e desafios lançados que me ajudaram a crescer. Ao Jorge Lourenço pelos conhecimentos técnicos e suporte, mas especialmente pela amizade. À Ana Rita pelo apoio, revisão, críticas construtivas e troca de ideias. Aos colegas da 3ª edição do MSIAD do ISCTE, e em particular aos colegas de grupo Hugo Afonso e Filipe Crispim, e àquele que se tornou em mais que um colega, num verdadeiro amigo, o Ricardo Sousa que tanto me incentivou para não desistir na reta final.

Ao meu orientador Professor Duarte Trigueiros agradeço pelo incentivo, orientação, estruturação e dedicação pessoal e profissional para a concretização desta dissertação. Ao Professor Alexandre Barão pela ajuda, orientação e críticas sempre construtivas que muito ajudaram no processo de revisão, e consolidaram o trabalho desenvolvido. Ao Mestre Pedro Perfeito pelas constantes trocas de ideias. Ao Professor Nelson António pelo apoio e troca de ideias sobre estratégia. Gostaria também de agradecer à Professora Maria José Trigueiros

pelas primeiras palavras de orientação no que seria o decurso de todo o mestrado. Finalmente ao Paulo Frade e novamente ao Rui Carneiro, por terem endossado a minha candidatura ao mestrado e terem acreditado nas minhas capacidades. A todos "Muito Obrigado"

Resumo

A sociedade atual é caracterizada pela elevada capacidade de produzir e armazenar dados. No entanto, esta crescente capacidade não se traduz num aumento da utilização dos mesmos para a produção de informação.

Por outro lado são várias as organizações não-governamentais que produzem grande volume de informação que apesar de muito relevante não chega ao domínio público. Seja pela fraca divulgação, pela falta de tratamento ou pelas deficientes interfaces concebidas existem enormes volumes de dados que não chegam cidadãos que ajudaram a financiar a criação desta mesma informação.

Esta dissertação aborda o tema e analisa o atual estado da disponibilização da informação de forma gráfica e como esta pode ajudar na consolidação do conhecimento.

Na vertente prática, esta dissertação procura apresentar algumas abordagens de representação analítica de informação relevante e disponibilizar, recorrendo a ferramentas SAP Business Objects, interfaces gráficas que ajudem a transformar estes dados em informação útil. Os Sistemas de Apoio à Decisão (SAD) desenvolvidos no âmbito desta tese procuram ajudar à tomada de decisões, seja por parte do cidadão, ou por outros agentes sociais e económicos ou para quem os indicadores estatísticos das economias ou países sejam relevantes.

Área de Aplicação: Sistemas de Apoio à Decisão.

Palavras-Chave: Sistemas de Apoio à Decisão, *Business Intelligence*, *Visualização de Informação*, *Visualização Analítica*, EuroStat.

Classificação ACM: H.2.4. Information Systems. Systems; H.2.7. Information Systems. Database Administration; H.4.2. Information Systems Applications. Types of Systems. Decision support.

Abstract

Today's Society is characterized for its high capability to produce and store data. However this growing capability is not translated in their increased use in order to produce information. On the other hand, there are several nongovernmental organizations that produce a large amount of relevant information but that never reach public domain. Whether by poor disclosure, lack of treatment or poorly designed interfaces, there's a huge amount of data that doesn't reach those who helped fund the creation of this same information.

We have on one side an excess of data, and on the other hand a lack of useful information.

This thesis seeks to approach the theme and analyze the current state of availability of the information in a visual graphic way and how it can help consolidate knowledge.

On a practical way, this dissertation seeks to present some approaches of analytical representation of relevant information that by using SAP tools, will produce graphical interfaces available to help transform this data into useful information.

The Decision Support System developed in the framework of this dissertation seek to help decision making, either by the citizen or by any other economic or social agent to whom the country statistic indicators might be relevant.

Title of Thesis: Development of Decision Support System to support visually the decision of citizens considering Social, Education and Development, Economic and financial indicators

Application Area: Decision Support Systems.

Keywords: Decision Support Systems, Business Intelligence, Information Visualization, Analytical Visualization, EuroStat.

ACM Classification: H.2.4. Information Systems. Systems; H.2.7. Information Systems.

Database Administration; H.4.2. Information Systems Applications. Types of Systems.

Decision support.

Índices

AGRADECIMENTOS.....	II
RESUMO	IV
ABSTRACT.....	V
ÍNDICES.....	VI
ÍNDICES DE FIGURAS	VIII
LISTA DE SIGLAS E ACRÓNIMOS.....	X
1. INTRODUÇÃO	1
2. ESTADO DA ARTE	3
2.1. ENQUADRAMENTO GERAL	3
2.2. VISUALIZAÇÃO DA INFORMAÇÃO NA ATUALIDADE	6
2.3. ENQUADRAMENTO DA VI NO CONTEXTO MUNDIAL	8
2.4. VISUALIZAÇÃO ANALÍTICA	9
2.5. TENDÊNCIAS TECNOLÓGICAS : MOBILIDADE	13
3. AS FONTES DE DADOS E A SUA ACESSIBILIDADE.....	15
3.1. O EUROSTAT COMO FONTE DE DADOS.....	17
4. ANÁLISE E DESENVOLVIMENTO DO SAD	19
4.1. IDEIA E PROPÓSITO	19
4.2. INFRA-ESTRUTURA TECNOLÓGICA.....	20
4.3. SISTEMA DESENVOLVIDO.....	21
4.3.1 <i>Modelo de dados</i>	22
4.3.2 <i>Fontes de dados</i>	24
4.3.3 <i>Processo ETL</i>	25
4.3.4 <i>Sistema Relacional</i>	27
4.3.5 <i>Criação da camada semântica</i>	28
4.3.6 <i>Criação de um report</i>	29
4.3.7 <i>Criação de um espaço de exploração</i>	30
4.3.8 <i>Interfaces móveis (APP)</i>	32
4.4. DASHBOARD	33
5. CONCLUSÕES	36

5.1. CONCLUSÃO DO ESTUDO DE CASO	36
5.2. TRABALHO FUTURO	37
BIBLIOGRAFIA.....	38
6. ANEXOS.....	40
6.1. EUROVIS2011.	40
6.2. CURRICULUM VITAE - FORMATO VISUAL	44
6.3. MANUAL TÉCNICO DATA SERVICES.....	45
6.3.1 <i>Object hierarchy</i>	45
6.3.2 <i>Designer window</i>	46
6.3.3 <i>Menu bar</i>	47
6.3.4 <i>Toolbar</i>	51
6.3.5 <i>Project Area</i>	52
6.3.6 <i>Tool palette</i>	52
6.3.7 <i>Workspace</i>	53
6.3.8 <i>Jobs</i>	56
6.4. MANUAL TÉCNICO INFORMATION DESIGN TOOL	57
6.4.1 <i>About the information design tool</i>	57
6.4.2 <i>Resources in the information design tool</i>	58
6.4.3 <i>Working with business layers</i>	60
7. LOMBADA DO TRABALHO	65

Índices de Figuras

FIGURA 1 – DADOS, INFORMAÇÃO CONHECIMENTO, SABEDORIA	4
FIGURA 2 - PARALELISMO ENTRE DATA MINING E VISUALIZAÇÃO DE INFORMAÇÃO	6
FIGURA 3 – TOR NØRRETRANDERS E A PERCEÇÃO DA INFORMAÇÃO PELOS SENTIDOS	7
FIGURA 4 – DIVISÃO DA VISUALIZAÇÃO E A RELAÇÃO COM A VC E VA.....	9
FIGURA 6 – HISTÓRIA DA VISUALIZAÇÃO ANALÍTICA	10
FIGURA 7 – ÂMBITO DA VISUALIZAÇÃO ANALÍTICA	11
FIGURA 8 -VISUALIZAÇÃO GAPMINDER.....	12
FIGURA 8 - VISUAL COMPLEXITY	13
FIGURA 12 – VISUALIZAÇÃO DAS TENDÊNCIAS DE EQUIPAMENTOS DE ACESSO À INTERNET.	14
FIGURA 10 - HANS ROSLING E O PROBLEMA DA ACESSIBILIDADES DAS FONTES DE INFORMAÇÃO.	15
FIGURA 11 - HANS ROSLING E O DESENVOLVIMENTO DE INTERFACES	16
FIGURA 12 – MODELO DA SOLUÇÃO IMPLEMENTADA.....	22
FIGURA 13 – ESTRUTURA DE DADOS DO EUROSTAT.	24
FIGURA 14 – IMPORTAÇÃO DOS FICHEIROS FONTE.....	26
FIGURA 15 – ESTRUTURA DO PROCESSO DE ETL EM SAP BUSINESS OBJECTS DATA SERVICES.	27
FIGURA 16 – MAPEAMENTO DAS FONTES DE DADOS E DAS TABELAS ASSOCIADAS.....	27
FIGURA 17 – MODELO FÍSICO DAS TABELAS.	28
FIGURA 18 – ESTRUTURA DA CAMADA SEMÂNTICA.....	29
FIGURA 19 – PAINEL DE QUERY EM WEB INTELLIGENCE	30
FIGURA 20 - CRIAÇÃO DO ESPAÇO EXPLORAÇÃO.....	30
FIGURA 21 - VISUALIZAÇÃO DO ESPAÇO DE EXPLORAÇÃO.	31
FIGURA 22 – VISUALIZAÇÃO DE UMA EXPLORATION VIEW NO BROWSER.	31
FIGURA 23 – DISPLAY DE UM ESPAÇO DE EXPLORAÇÃO NO IPAD	32
FIGURA 24 – EXPLORATION VIEW IPAD.	32
FIGURA 25 – ASPETO GERAL DO QUADRO DE INDICADORES.	34

FIGURA 26 – DETALHE PARA UM INDICADOR ESCOLHIDO..... 34

FIGURA 27 – COMPORTAMENTO DE LICITAÇÃO DE UM UTILIZADOR EM 2 PERÍODOS DE TEMPO. 41

FIGURA 28 – VISUALIZAÇÃO DA UTILIZAÇÃO DE UM SMARTPHONE DURANTE O DIA 42

Lista de Siglas e Acrónimos

BI	<i>Business Intelligence</i>
BI Launch Pad	<i>SAP BusinessObjects BI Launch Pad</i>
ETL	<i>Extract, Transform and Load</i>
DM	<i>Data Mining</i>
DSS	<i>Decision Support Systems</i>
DW	<i>Data Warehouse</i>
MSIAD	<i>Mestrado em Sistemas Integrados de Apoio à Decisão</i>
OLAP	<i>On-line Analytical Processing</i>
OLTP	<i>On-Line Transaction Processing</i>
SAD	<i>Sistemas de Apoio à Decisão</i>
SAP	<i>Systems Application and Products in Data Processing</i>
SAP BOBJ BI	<i>SAP BusinessObjects Business Intelligence Platform</i>
SAP BW	<i>SAP Netweaver Business Warehouse</i>
SI	<i>Sistema de Informação</i>
VA	<i>Visualização Analítica</i>
VC	<i>Visualização Científica</i>
VI	<i>Visualização de Informação</i>
Web	<i>World Wide Web</i>
Web Intelligence	<i>SAP BusinessObjects Web Intelligence</i>

1. Introdução

A presente dissertação aborda o problema da dificuldade de acesso à informação, não só pela ausência de divulgação, mas também pela forma como a informação é transmitida aos cidadãos.

A ideia para este projeto surgiu de uma necessidade muito real. A necessidade que a minha avó tinha em saber em quem votar. Esta necessidade acentuou-se com a dificuldade de ouvir e com a conseqüentemente incapacidade de acompanhar com regularidade os meios de comunicação social. Não obstante tratar-se de uma senhora de avançada idade, cujo acesso a todos os meios de comunicação atualmente disponíveis é limitado, a verdade é que o principal problema residia na dificuldade da total compreensão da informação que lhe era fornecida. Este problema não é exclusivo da minha avó. É transversal a toda a nossa sociedade. A informação para ser útil deve ser clara, precisa e estar facilmente disponível por forma a poder ajudar a tomada de decisão. No entanto a informação relevante para o cidadão muitas vezes não está acessível ou é apresentada de forma pouco clara. Esta problemática reflete a nossa realidade e é globalmente imune a diferenças de idade, género ou até de nível educacional e cultural.

Os dados disponíveis existem em cada vez maior número, e os meios de colocação da informação ao serviço dos cidadãos são cada vez mais abrangentes. No entanto, continuam a existir lacunas neste processo que impedem que a informação chegue de forma clara à maioria das pessoas. Torna-se desta forma fundamental a existência de interfaces adequadas que ajudem qualquer cidadão a aceder e a compreender de forma rápida e explícita a informação existente.

Neste sentido, o trabalho desenvolvido teve como principal objetivo a criação de interfaces gráficas, com a apresentação de indicadores que permitam avaliar questões que condicionam a vida dos cidadãos, contribuindo para a sua clarificação e melhoria do conhecimento.

O trabalho de que a presente dissertação trata foi desenvolvido com o uso de ferramentas de recolha e tratamento de dados e de Business Intelligence (BI), nomeadamente tecnologias da SAP Business Objects, tendo por base dados estatísticos disponibilizados pela Direção Geral de Estatística da Comissão Europeia (EuroStat).

A dissertação descreve o trabalho desenvolvido e está organizada em quatro partes. Após a presente introdução, no capítulo 2 é apresentada a análise e estudo do estado da arte da visualização da informação. No capítulo 3 é abordada a dificuldade da acessibilidade dos dados e o enquadramento do Eurostat como fornecedor de dados a nível europeu. O capítulo 4 apresenta a solução de interfaces gráficas e o trabalho efectuado na sua obtenção. Finalmente no capítulo 5 são apresentadas as conclusões e considerações sobre trabalho a desenvolver futuramente.

2. Estado da Arte

Este capítulo é composto pelo estudo dos conceitos, metodologias, disciplinas e soluções que orientam a temática da investigação. Assim na secção 2.1 é efectuado um enquadramento geral e a história da Visualização da Informação (VI). Na secção 2.1 é apresentada a perspectiva da VI na actualidade. A secção 2.3 aborda o enquadramento da VI no contexto mundial. A Visualização Analítica (VA) é abordada na secção 2.4.

2.1. Enquadramento geral

Vivemos atualmente numa sociedade que enfrenta um rápido crescimento da quantidade de informação que necessita ser tratada diariamente.

Praticamente todas as áreas da indústria, dos serviços, da comunicação e de muitos outros setores da vida social e política geram grandes quantidades diárias de informação, seja na forma de imagens, ficheiros, bases de dados, ou mesmo em suporte não digital. Esta realidade, juntamente com a crescente capacidade de recolher e armazenar dados, levou a que esse volume de informação tenha crescido a um ritmo superior ao da nossa capacidade de a processar, usar e tomar decisões.

Segundo um estudo efetuado no Global Information Industry Center da Universidade da Califórnia [1], estima-se que o mundo tenha produzido 14.7 exabytes¹ de novos dados em 2008, aproximadamente o triplo do volume de 2003.

Este crescente volume de dados disponíveis provoca problemas tanto para as organizações como para os indivíduos. A nível organizacional, o excesso de dados não tratados aumenta os níveis de ineficiência, contribuindo para que tempo e dinheiro sejam desperdiçados e consequentemente muitas oportunidades sejam perdidas. Ao nível do individuo, o impacto pode refletir-se tanto no aumento do nível de stresse causado pela permanente abundância de nova informação, como na limitação da sua capacidade de tomar decisões.

Num artigo de 2005 [2], E.M. Hallowell aborda o tema e identifica alguns destes problemas:

- O aumento do tempo necessário para a tomada de decisões;

¹ 1 Exabyte = 1.000.000.000 gigabytes

- Aumento dos erros de análise;
- Dificuldade na identificação de relações entre detalhes e a perspetiva global;
- Tornar-se altamente seletivo, ignorando grande quantidades de informação por incapacidade de análise de elevados volumes de dados.

O volume de dados está em constante crescimento. A forma como lidamos com eles no sentido de os transformar em informação útil é fundamental.

De forma a compreender a globalidade da questão, é importante fazer a distinção entre vários conceitos que por vezes são confundidos: Dados, Informação, Conhecimento e Sabedoria. Esta distinção é definida em inglês como a hierarquia DIKW - Data, Information, Knowledge e Wisdom.

No contexto de Sistemas de Informação, quando nos referimos a “Dados” considera-se que estes são o simples registo de factos, recolhidos num espaço temporal e desprovidos de contexto; por sua vez a “Informação” corresponde aos dados aos quais foi atribuído um contexto e uma organização; ao processo de interiorização da informação que poderá ser utilizada num processo de tomada de decisão denomina-se “Conhecimento”; já a “Sabedoria” poderá ser entendida como a capacidade de utilizar o conhecimento e, através de um processo de reflexão, complementar o conhecimento adquirido.

Como exemplificado na Figura 1 ², é possível ver a relação entre estes 4 conceitos. Assim conforme avançamos nos vários degraus da hierarquia DIKW e aumenta a contextualização, maior é o grau de compreensão sobre o tema analisado. Só quando a informação é devidamente contextualizada e percebida poderá produzir conhecimento e sabedoria.

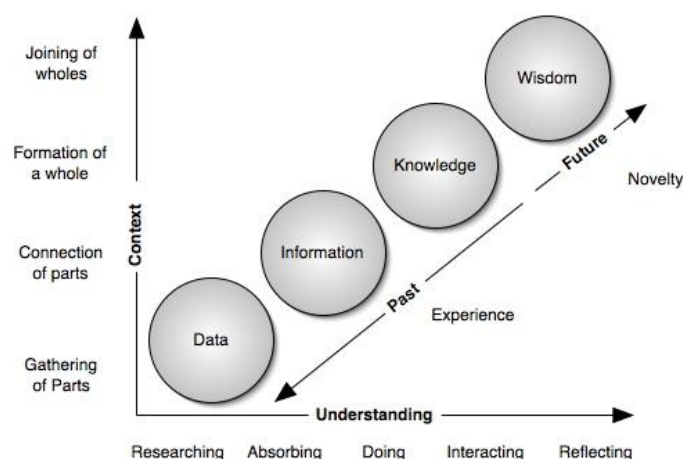


Figura 1 – Dados, Informação Conhecimento, Sabedoria

² Knowledge – Clark, Donald - <http://www.nwlink.com/~donclark/knowledge/data.html> - acedido em 09/06/2012

A contextualização pode ser vista em diferentes níveis, que vão desde a simples recolha até a ligação das várias partes. Por outro lado a compreensão pode ser vista também por níveis e que podem ir desde o processo de pesquisa até à reflexão.

Neste processo de passagem entre os vários níveis da hierarquia *DIKW*, os atuais Sistemas de Informação (SI) assumem um papel fundamental. Os dados contidos na maioria das aplicações, se não tratados, não apresentam valor *per si*. Para adquirirem valor, é necessário que sejam tratados para se poder extrair a informação neles contida. Os SI, sejam eles manuais ou automatizados, permitem o tratamento e o processamento dos dados de forma a ser obtida informação. Este é um passo essencial para a descoberta de conhecimento que apoia o processo de tomada de decisão.

A Visualização da Informação (VI) é um campo relativamente novo e interdisciplinar, que emergiu pelo desenvolvimento de várias áreas de investigação, nomeadamente das ciências computacionais, do estudo das relações Homem-Máquina, Desenho Gráfico, Design e Psicologia.

A VI tem as suas raízes no trabalho “*Semiologie Graphique*” [3] do cartógrafo francês Jacques Bertin que, em 1967, descreveu o enquadramento para a conceção dos elementos básicos de diagramas. Alguns anos mais tarde, Edward Tufte desenvolveu a teoria de gráficos de dados [4], maximizando a densidade de informações úteis, dentro dos limites de um diagrama. O campo de investigação científica de visualização da informação surgiu quando as teorias desenvolvidas por J. Bertin e E. Tufte foram usadas pela primeira vez em aplicações interativas de computador por Stuart K. Card, Ben Shneiderman e Jock D. Mackinlay [5]. Neste processo foram então identificados como objetivos a capacidade de interagir com grandes quantidades de informação dinâmica e a possibilidade do processo de pesquisa ser igualmente dinâmico. Na década seguinte, o número de visualizações aumentou, o que levou Ben Shneiderman a incentivar o desenvolvimento de taxonomias e quadros para descrever tecnologias de visualização em 1996. Um dos trabalhos referência de visualização da informação “*Readings in Information Visualization: Using Vision to Think*” [6], que descrevera a análise multidimensional, bem como a formalização do processo de visualização da informação. Em 2000, Colin Ware [7] apresentou o conceito de multidisciplinaridade do tema, onde enfatizou a importância da Visualização da Informação na percepção humana.

Novas abordagens surgiram com o potencial de novas tecnologias que permitiram armazenar cada vez mais dados. A área de Data Mining³ (DM), que se desenvolveu paralelamente à VI, teve uma abordagem diferente para encontrar a informação em grandes volumes de dados. O objetivo desta abordagem não é apoiar a percepção humana com visualizações interativas, mas sim permitir que os computadores efetuem o trabalho de um analista e contribuam para o processo de identificação de padrões, que através de processos manuais seria impossível. Como é possível ver na Figura 2, o DM procura a obtenção de um modelo computacional, que utilizado sobre um conjunto de dados procura levantar hipóteses, para assim obter conhecimento. A abordagem da VI, apesar de ter o mesmo ponto de partida e o mesmo objetivo, assenta na criação de modelos visuais para a identificação de padrões.

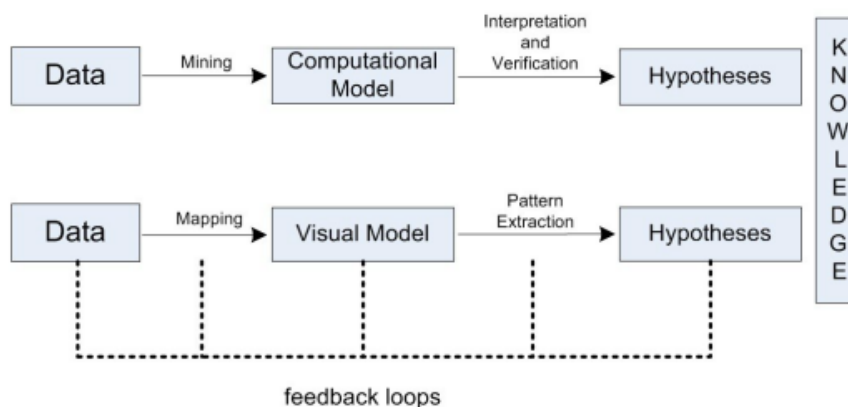


Figura 2 - Paralelismo entre Data Mining e Visualização de Informação [8]

2.2. Visualização da Informação na Atualidade

David McCandless⁴, defende que pode haver uma solução fácil para o problema do excesso de informação, e essa solução é utilizar mais os nossos olhos. Desta forma podemos visualizar a informação, para que possamos ver os padrões e conexões que interessam e depois projetar essa informação para que faça mais sentido, ou permitir-nos focar somente na informação que é importante. A importância da percepção visual é ilustrada num trabalho de Tor Nørretranders[9]. Neste estudo o autor estratifica e associa de forma original a quantidade de informação que pode ser absorvida por cada um dos nossos sentidos. Tor Nørretranders

³ Data Mining é uma área de investigação resultante da ciência computacional e estatística que se caracteriza como o processo de exploração de grandes quantidades de informação através de modelos matemáticos e estatísticos para identificação de padrões nos dados

⁴ "The beauty of data visualization"- TED Talk 2010 –

http://www.ted.com/talks/david_mccandless_the_beauty_of_data_visualization.html - acedido em 18/07/2012

defende que a visão é o nosso sentido mais rápido e tem aproximadamente a mesma largura de banda de uma rede de computadores. Depois o tato, que tem aproximadamente a velocidade de uma porta USB. Segue-se a audição e o olfato, que têm a mesma taxa de transferência que um disco rígido. O paladar é o sentido com menor capacidade e tem uma taxa de transferência de uma máquina calculadora de bolso. E de toda esta informação que nos chega pelos sentidos, apenas 0,7 por cento, é de facto a quantidade que estamos conscientes. Grande parte da nossa percepção – e a sua maioria é visual – é muitas vezes inconsciente, e é o olho humano o grande responsável por esta percepção pois é extremamente sensível a variações de cor, forma e padrão.

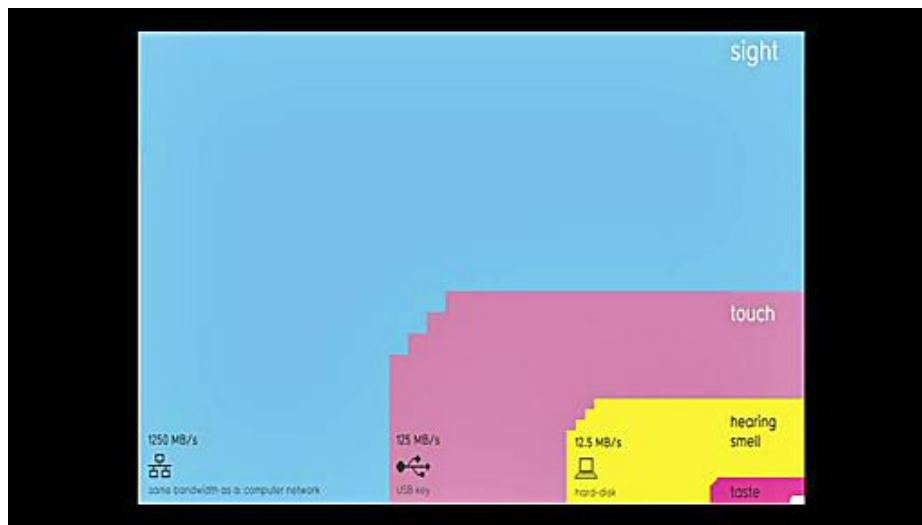


Figura 3 – Tor Nørretranders e a percepção da informação pelos sentidos

A capacidade da assimilação de informação pela visão é justificada pela sensibilidade dos olhos ao contraste, ao padrão, à orientação, entre outras. No âmbito dos trabalhos publicados na área de VI, as questões mais estudadas têm sido a criação de visualizações para determinados tipos de dados e o desenvolvimento de técnicas de interação com os dados através das visualizações e interfaces gráficas, correspondendo ao que é designado pela fase exploratória de uma nova disciplina. Antes de ser possível construir sistemas genéricos que possam ser usados para vários conjuntos de dados é necessário sistematizar quais as características de um sistema de Visualização de Informação. Nesta fase, devem ser identificados quais os tipos de dados a visualizar, quais os tipos de mapas e quais as formas de interação. É importante também ter em conta o tipo de utilizadores a que se destinam e as tarefas que deve executar. Quanto aos utilizadores a que se destina, considera-se que a VI serve um conjunto de utilizadores vasto e não necessariamente especializado, quer em técnicas de visualização, quer em relação ao domínio da informação a analisar.

Quanto ao tipo de tarefas a executar, de um modo geral existem diferentes situações a ter em conta, como por exemplo a Análise Exploratória dos Dados (AED)⁵, a apresentação final de resultados ou o acesso rotineiro para consulta, tendo cada uma das tarefas exigências diferentes. Ao analisar dados o objetivo é “descobrir” informação. A VI deverá permitir realçar diferentes aspetos, sejam eles globais ou parciais e de preferência, de uma forma rápida, não sendo normalmente relevante a qualidade final da imagem. O mesmo não se verifica quando o objetivo é obter imagens para uma apresentação sobre o estudo dos dados. Neste caso interessa ter imagens finais apuradas. Quanto ao acesso para consulta, interessa sobretudo ter uma forma pré-estabelecida e simples de visualizar os dados que estão disponíveis.

Não será, contudo, razoável ter um sistema que consiga responder a todas as situações. Isso seria naturalmente muito exigente, quer em termos de recursos computacionais, quer no esforço de aprendizagem requerido. Como o tipo de informação que pode ser considerado no âmbito da Visualização de Informação é vasto, a solução muitas vezes adotada passa pela identificação de várias subáreas e pelo desenvolvimento de técnicas próprias para cada uma delas.

2.3. Enquadramento da VI no contexto mundial

A VI é um tema bastante ativo e são várias as conferências sobre o tema. No entanto 3 conferências a nível mundial assumem grande destaque no estudo da matéria e com um crescente número de trabalhos apresentados. Estas 3 conferências (PacificVis para a Zona Ásia Pacífico ; EuroVis para a zona Europeia ;VisWeek no continente americano) são apoiadas pela IEEE⁶ (Institute of Electrical and Electronics Engineers).

As conferências decorrem anualmente e debruçam-se sobre os mais variados temas da Visualização da Informação e Visualização Analítica, sendo este o meio onde formam apresentados os trabalhos com maior relevância para o estudo da área.

No anexo 6.1 é apresentado um resumo da conferência EuroVis2011.

⁵ O conceito de AED foi defendida por John Tukey como sendo a abordagem de analisar conjuntos de dados sem recurso a um modelo estatístico e sem a formulação de hipóteses definidos.

⁶ A IEEE é uma associação profissional dedica-se ao estudo de avanços tecnológicos de inovação e excelência, tendo mais de 400.000 membros em mais de 160 países. <http://www.ieee.org> acedido em 08/02/2012

A temática da VC não será tema de desenvolvimento nesta dissertação, sendo a atenção focada na Visualização de Informação e na Visualização Analítica.

A VA foi definida por James J. Thomas and Kristin A. Cook [10] como a "ciência do raciocínio analítico através de interfaces visuais". O seu desenvolvimento surge na sequência de desenvolvimentos nas áreas de VI e VC, e procura focar-se no pensamento analítico, através do recurso a interfaces gráficas sendo resultado de várias áreas de investigação. Como se pode observar na Figura 6. Nesta imagem é possível analisar as várias áreas que contribuíram para o aparecimento e desenvolvimento da VA como área de estudo. Como já visto anteriormente, através do estudo da visualização e a utilização da computação gráfica, a VI surgiu como campo de estudo. A VA surge tendo por base estudos desta área, mas recebendo influências de muitas outras áreas, como a estatística, texto mining ⁹, o crescimento das bases dados e os processos de gestão de conhecimento, mas também do processo de diminuição de incerteza no processo de tomada de decisão e análise do comportamento cognitivo.

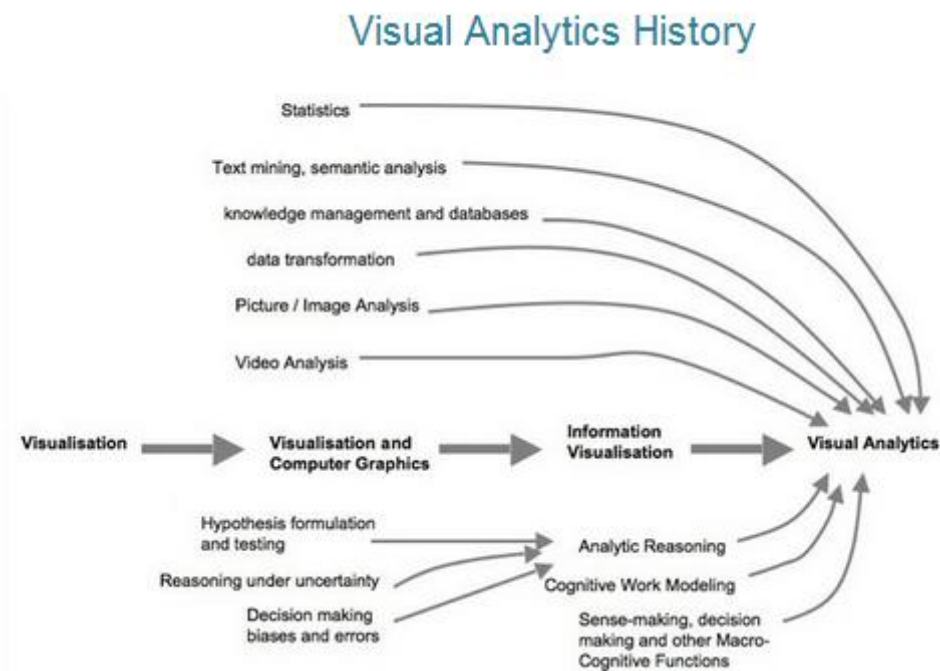


Figura 5 – História da Visualização Analítica 10

⁹ Text Mining - Processo analítico de obtenção padrões de dados através da análise de textos livres.

¹⁰ Imperial College London, Institute for Security Science and Tecnology

O âmbito da multidisciplinar da VA é analisada em por Keim, Daniel A. & Mansmann, Florian & Schneidewind, Jorn & Ziegler, Hartmut em “Challenges in Visual Data Analysis” [11]. Na Figura 7 é possível ver a o âmbito da VA, como campo de pesquisa interdisciplinar, que procura combinar as vantagens da capacidade cognitiva e percepção humana, o poder de processamento dos computadores, análise geoespacial, processos de gestão de informação e descoberta de conhecimento

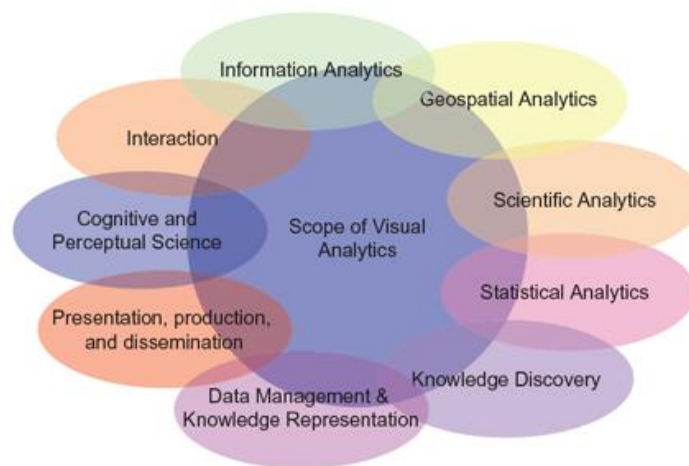


Figura 6 – Âmbito da Visualização Analítica ^[11]

Os problemas levantados são relacionados com o dimensionamento, complexidade, percepção e necessidade de análise e síntese. A área de Visualização Analítica (VA) procura obter avanços científicos e tecnológicos no raciocínio analítico, interação, transformações de dados e representações para computação e visualização de relatórios analíticos.

Alguns dos temas com impacto na VA são já tema aprofundado em outras áreas. Por exemplo, na área da Psicologia e ciências cognitivas estuda-se quais os elementos visuais de maior impacto na visualização da informação (Cor - Matriz; Saturação; Brilho; Textura; Orientação); Técnicas (Gráficos de linha, Scatter Plots, Grifos e ícones, Visualização volumétrica, Diagramas de Fluxo, etc.).

A forma de utilização destas técnicas tem sido o tema de vários livros de Stephen Few. A sua abordagem bastante prática, procura dar uma visão geral sobre todos os temas envolvidos na VA, e como estes podem ajudar a criar impacto. Não é apenas necessário apresentar os dados mas fazer com que os dados sejam percebidos. Segundo o autor este tem sido um dos

problema da área de *Business Intelligence* (BI)¹¹ : “O que a indústria de BI tem ignorado é o facto que a inteligência residir nos seres humanos, e que a informação apenas se torna valiosa quando é percebida e não apenas disponibilizada” Pag. 3. [12]

Este problema é apontado também em “Information Dashboard Design” [13], Stephen Few, onde são apresentados exemplos de vários fabricantes de software de BI – Pág. 10 a 34 – e como estes, que são considerados exemplos de portfólio, falham nos mais elementares conceitos de VA. O livro aborda a importância dos dashboards¹² como forma eficaz de transmitir informação de forma simplificada, mas que se a tecnologia for utilizada de forma pouco eficaz se traduz num desperdício de recursos financeiros e humanos.

A nível literário outros 2 autores merecem a referência pela sua influência:

- Hans Rosling – é um professor sueco, médico, estatístico, académico e orador público. Ele é professor de Saúde Internacional do Instituto Karolinska e Diretor da Fundação Gapminder, que desenvolveu o sistema de software Trendalyzer (Figura 8) . De 1967 a 1974, estudou dados estatísticos e medicina na Universidade de Uppsala e, em 1972, estudou a saúde pública no centro médico St. John's College, em Bangalore. Tem-se destacado recentemente por uma série de conferências TED¹³ no qual procura dar uma visão geral sobre o estado da saúde mundial através de mapas visuais.

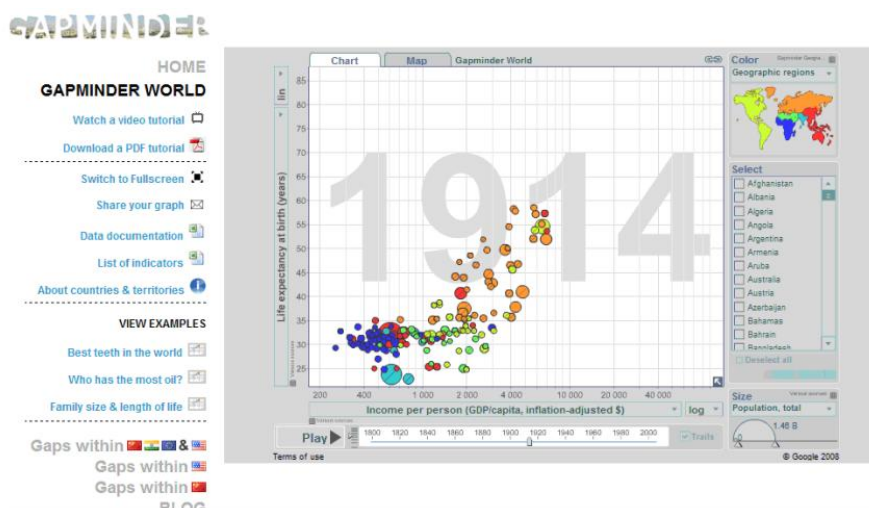


Figura 7 -Visualização GapMinder

¹¹ Refere-se ao processo de recolha , organização, análise e monitorização de informações relevantes para a gestão das empresas.

¹² Segundo Stephen Few um dashboard é “uma representação gráfica da informação mais importante necessária para atingir um ou mais objectivos e que cabe completamente num ecran de computador de forma a ser monitorizado num relace. ”- Pag 34. [13]

¹³ TED Talks (www.ted.com) - acedido em 10/06/2012

Manuel Lima – é um designer e investigador português. Foi nomeado, pela prestigiada revista americana *Creativity*, uma das 50 mentes mais criativas e influentes de 2009 pelo seu trabalho de investigação desenvolvido no site *VisualComplexity*¹⁴. A figura 8 representa alguns dos trabalhos apresentados no site. Manuel Lima tem dado particular enfoque na investigação de visualização de redes de informação. É também autor do livro *Visual Complexity: Mapping patterns of Information*

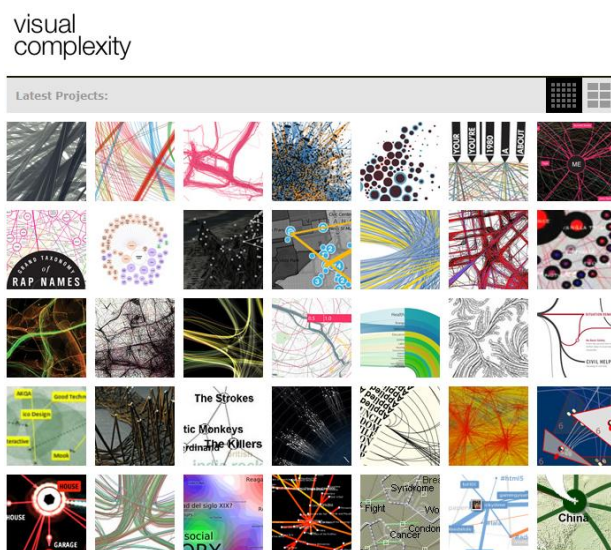


Figura 8 - Visual Complexity

2.5. Tendências Tecnológicas : Mobilidade

Actualmente a realidade dos sistemas dos sistemas de informação, não se resume apenas aos *desktops* ou aos computadores portáteis. Com a necessidade de tornar a informação acessível em qualquer lugar, a mobilidade dos dispositivos ganha cada vez maior relevância.

Segundo Alex Cocotas¹⁵, estamos atualmente num ponto de viragem (Figura 14). No ano passado, a venda de Smartphones ultrapassou a venda de computadores pessoais e a tendência é claramente ascendente.

¹⁴ www.visualcomplexity.com - acedido em 05/01/2012

¹⁵ Analista de investigação para Business Insider Intelligence.

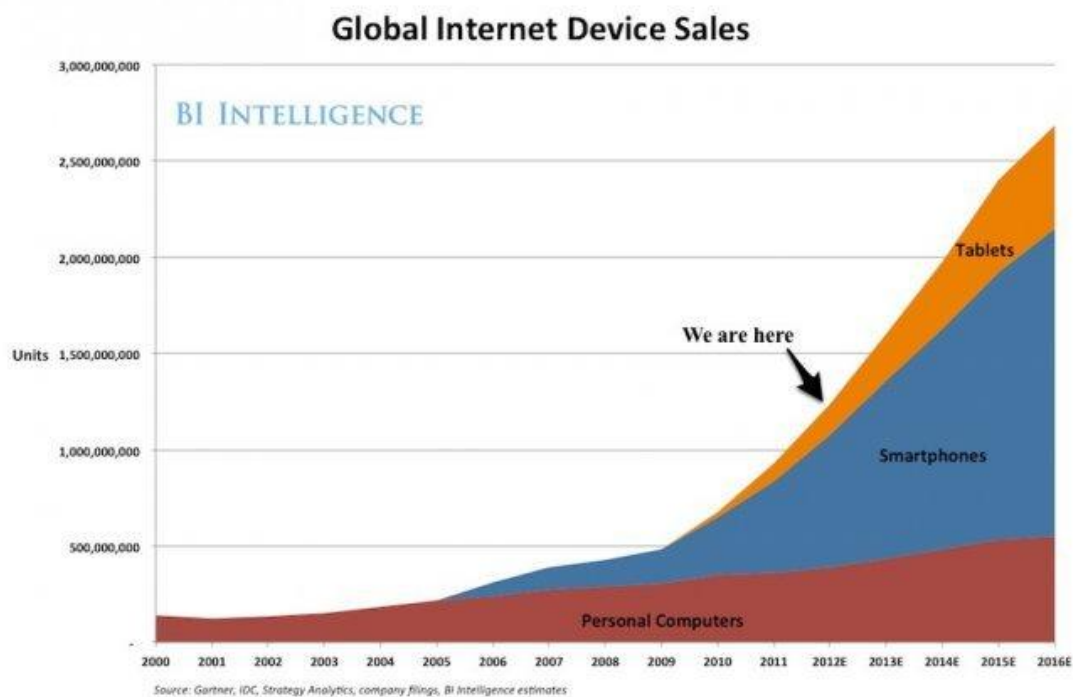


Figura 9 – Visualização das tendências de equipamentos de acesso à Internet.

O presente trabalho procura abordar esta realidade, e assim foi considerado que a inclusão da perspectiva da mobilidade no projeto seria não só uma mais-valia mas uma necessidade, e desta forma tornar mais fácil o acesso à informação, através de numa tecnologia que faz cada vez mais parte do nosso dia-a-dia.

3. As fontes de dados e a sua acessibilidade

No mundo atual a abundância de informação tornou os cidadãos não em info-excluídos mas sim em excluídos do conhecimento. A informação está disponível nos mais diversos formatos, no entanto o seu tratamento ou acesso a esta é por vezes limitado.

Na conferência TED "Stats that reshape your worldview" ¹⁶, o professor Hans Rosling apresenta algumas ideias interessantes e que refletem a diversidade de fontes de informação e a dificuldade da sua disponibilização de forma maciça:

“Porque é que isto não acontece? Porque razão não usamos os dados que temos? Temos dados nas Nações Unidas, nas agências nacionais de estatísticas, nas universidades e noutras organizações não-governamentais. Porque é que os dados estão escondidos nas bases de dados. E o público está ali, e a Internet está ali, mas ainda não usamos isto eficazmente. (...)”

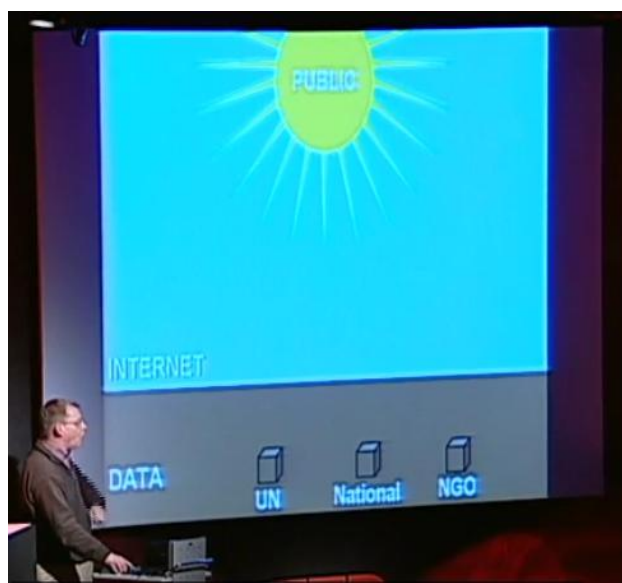


Figura 10 - Hans Rosling e o problema da acessibilidades das fontes de informação.

"Há algumas páginas Web como esta mas elas alimentam-se das bases de dados. Mas as pessoas colocam-lhe preços, palavras-chave estúpidas e estatísticas maçadoras.

E isso não funciona. Por isso, o que é preciso? Nós temos as bases de dados. Não é de novas bases de dados que precisamos (...)

¹⁶ http://www.ted.com/talks/hans_rosling_shows_the_best_stats_you_ve_ever_seen.html - 2006 - acedido em 06/09/2012

(...) Mas era isso que nós gostaríamos de ver, não era? Os dados publicamente financiados estão aqui em baixo. E gostaríamos que as flores crescessem na Internet. E um dos pontos cruciais é torná-los pesquisáveis, e que depois as pessoas podem usar as diferentes ferramentas de design para os animar ali (...) E a questão seria obter acesso às bases de dados gratuitamente, torná-las passíveis de pesquisa e transformá-las em formato gráfico, onde podem ser instantaneamente compreendidas. (...)”. “Visualizar a informação desta maneira é uma forma de compressão de conhecimentos. Uma maneira de apertar uma enorme quantidade informação e entendimento num pequeno espaço”.

A Figura 11 representa este desejo de Hans Rosling. As bases de dados nos seus vários formatos (Excel, Access, MySQL, XML, Webservices, ODBC) ligadas às ferramentas de design (PowerPoint, Freehand, Flash, Illustrator), para darem lugar às ‘flores’ na internet - interfaces gráficas.

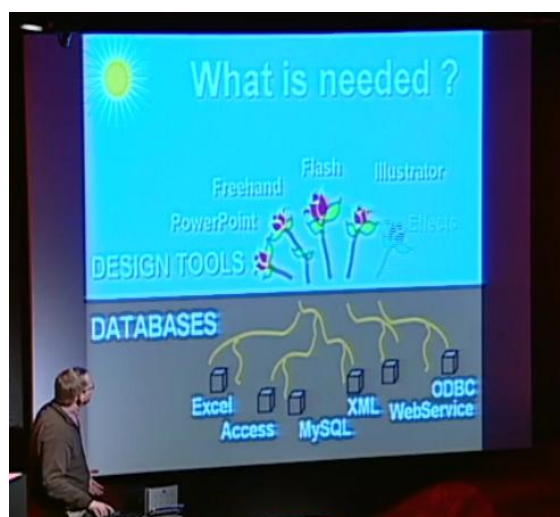


Figura 11 - Hans Rosling e o desenvolvimento de interfaces

Nesta conferência TED, e em outras que se seguiram, Hans Rosling expõe este problema que afeta atualmente todas as fontes de dados. São várias as instituições não governamentais mundiais que conseguiram agregar grandes quantidades de informação mas no entanto esta informação não está disponível ao público em geral ou o seu acesso é limitado. Só através do acesso livre e sem restrições poderá ser possível o acesso a dados que foram financiados pelos cidadãos e que a eles dizem respeito.

Neste contexto, faz sentido debruçar o estudo sobre aquele que é o principal fornecedor de dados estatísticos europeus - O EuroStat¹⁷.

¹⁷ <http://epp.eurostat.ec.europa.eu/portal/page/portal/eurostat/home/> - acedido em 10/09/2012

3.1. O EuroStat como fonte de dados

O EuroStat tem por missão, ser o "fornecedor de estatísticas de elevada qualidade da Europa". O EuroStat foi criado em 1953 como resposta aos requisitos da Comunidade do Carvão e do Aço. Ao longo dos anos a sua função foi sendo alargada e, quando a comunidade europeia foi fundada em 1958, o EuroStat transformou-se numa direção geral da comissão europeia. Com o desenvolvimento de políticas comunitárias o papel do EuroStat mudou. Hoje coleciona dados para os Estados Membros Europeus e desenvolve sistemas estatísticos para países candidatos. O papel principal do EuroStat é processar e publicar informações estatísticas comparáveis a nível europeu. No EuroStat procura-se a estatística na linguagem comum, abarcando conceitos, métodos, estruturas e técnicas *standards*. O EuroStat não recolhe informação. Esta função é cumprida a nível do estado membro pelas suas autoridades estatísticas. São estes os responsáveis por verificar analisar os dados nacionais e enviá-los para o EuroStat. Sendo o único fornecedor de estatísticas a nível europeu, o papel do EuroStat é o consolidar os dados e assegurar que são comparáveis usando metodologias harmonizadas. Desta forma fornece às direções gerais, à comissão e outras instituições europeias dados para que estas possam definir, implementar e analisar as políticas comunitárias. Resultado: o EuroStat fornece um rol muito alargado e interessante de dados que governos, empresas, sector educacional, jornalistas e público podem usar na sua vida diária. Estatísticas sólidas, objetivas e fiáveis são essenciais para o correto funcionamento de sociedades democráticas.

Ao nível dos decisores políticos, nos estados membros, nos governos locais e ao nível das empresas são necessárias estatísticas para apoiar a tomada de decisões nos contextos que se inserem. Por outro lado, o público e os meios de comunicação social precisam ter uma imagem da sociedade contemporânea e serem capazes de avaliar corretamente a performance dos políticos e de outros intervenientes da vida social.

As estatísticas existentes podem responder a muitas questões: Estará a sociedade a seguir na direção prometida pelos políticos? Estará o emprego a crescer ou a decrescer? Como estarão a evoluir as emissões de CO2 compradas com os últimos 10 anos? Qual o número de mulheres no mercado de trabalho? Como está a comportar-se a economia do seu país comparado com

os outros países? Estatísticas internacionais são uma forma de conhecer os países membros vizinhos e outros países fora da união europeia.

O EuroStat é neste contexto a instituição que pela sua credibilidade torna possível a manutenção de um sistema que permite a comparação do desenvolvimento entre os vários países, mas que, apesar de um trabalho notável na recolha e tratamento informativo, coloca ainda pouco enfoque na distribuição da informação de forma mais acessível a pessoas com pouca formação tecnológica ou fraco conhecimento de línguas.

4. Análise e desenvolvimento do SAD

Este capítulo apresenta os vários passos para a construção da solução tecnológica que suportado em dados extraídos do *site* do EuroStat, serviu de sustento às interfaces gráficas propostas.

Na seção 4.1 é apresentada a ideia e propósito por trás do projecto. Na seção 4.2 é analisada a infra-estrutura tecnológica. A seção 4.3 aborda o sistema desenvolvido, nomeadamente o modelo, as fontes, o processo de Extração, Transformação e Tratamento (*ETL*), o desenho do sistema relacional, a criação da camada semântica, o processo de criação de um *report*, a criação de um espaço de exploração e o desenvolvimento de interfaces móveis. Na secção é abordado o desenvolvimento do dashboard.

4.1. Ideia e propósito

No mundo em que vivemos onde o poder se sustenta pela construção e preservação de modelos hierárquicos puros, o decisor encontra-se geralmente no topo de uma pirâmide. Essa pirâmide, estratificada, tendo uma larga base sobre a qual assenta caracteriza-se por ter um pequeno número de decisores no topo. No entanto, e por contraponto, num modelo de democrático como o ocidental, em que o poder político é atribuído por via de eleições, os decisores são todos os eleitores, e como tal uma maioria.

Seja pelo excesso de informação, pela escassez desta, ou pela pouca clareza e objetividade da informação disponível o eleitor acaba, muitas vezes, por decidir com base na pouca informação a que tem acesso, e influenciado por questões destacadas pelos meios de comunicação.

A ideia para este projeto partiu do objetivo de tornar a informação que influencia a atividade de um país, acessível e objetiva.

Numa recente conferência TED, Aaron Koblin¹⁸ dizia: "Acho que os dados podem realmente tornar-nos mais humanos. Estamos a recolher e a criar todo o tipo de dados sobre como

¹⁸ Artfully visualizing our humanity - http://www.ted.com/talks/aaron_koblin.html - 2011 - - acedido em 09/08/2012

estamos a viver as nossas vidas e isso permite-nos contar algumas histórias fascinantes. Recentemente, um sábio teórico dos meios de comunicação 'twittou' - ["A cultura do século XIX foi definida pelo romance, a cultura do século XX foi definida pelo cinema, e a cultura do século XXI vai ser definida pela interface."] Acredito que isto vai ser verdade. As nossas vidas estão a ser guiadas por dados, e a apresentação desses dados é uma oportunidade”.

A oportunidade de existirem múltiplas fontes de dados, para as quais os interfaces estão pouco divulgados ou com acessibilidade limitada, criou a oportunidade de produzir uma interface gráfica que fosse intuitiva.

Na ideia inicial, era pretendido com este projeto criar apenas uma interface gráfica simples, que mostrasse a realidade dos dados de um país comparativamente, fosse em relação a outro país ou ao próprio país. No entanto o projeto foi-se tornando mais ambicioso. Com o aprofundar do tema foi-se tornando claro que uma única interface gráfica não responderia à generalidade das necessidades e assim foi concebido não apenas uma interface mas um conjunto de soluções e interfaces. À ideia inicial de conceber um único *dashboard*, juntaram-se outras necessidades. Necessidade de criar *reports* à medida e de criar um grau de abrangência maior de forma a responder ao utilizador mais exigente e que pretendesse fazer a sua análise, e não aceder a uma visão predefinida. Assim, o modelo proposto procura dar resposta a vários graus de exigência de utilizadores.

Os passos que se descrevem a seguir foram os passos necessários para a criação das visualizações, em especial do *dashboard*.

4.2. Infra-Estrutura Tecnológica

A solução implementada assentou em várias soluções informáticas presentes no mercado empresarial, nomeadamente nas soluções de BI da SAP BusinessObjects. A escolha das plataformas incidiu sobretudo no critério de domínio da tecnologia necessária.

A SAP (Systems, Applications, and Products in Data Processing) é uma empresa alemã, fundada em 1972, contando atualmente com mais de 55.000 trabalhadores, cerca de 197.000

clientes em 130 países, e uma faturação superior a 14 mil milhões de dólares. É atualmente a empresa líder mundial a nível de software aplicacional empresarial.

Em 2008, através da aquisição da *Business Objects*, a SAP apostou forte na consolidação da sua posição, alargando a gama de produtos para a área de soluções analíticas e em especial para a área de *Business Intelligence*.

As aplicações utilizadas nesta dissertação foram *BusinessObjects Data Services XI*, para o tratamento e integração de dados, *BusinessObjects Information Design Tool* para o processo de construção da camada semântica; *BusinessObjects BI Launch Pad* para a conceção de reports, *BusinessObjects Explorer* para a geração de um espaço de informação e o *BusinessObjects Xcelsius Dashboard Design* para conceção de *dashboard*.

A nível de *hardware*¹⁹ foi utilizado um servidor HP com processador *Xeon*, 16 Gigabytes (Gb) de memória RAM, um disco de 500 Gb e um router Netgear N600.

4.3. Sistema Desenvolvido

O sistema desenvolvido foi dividido em várias partes onde se faz o estudo de cada uma das componentes do desenvolvimento. Este assenta no tratamento de dados originários do *site* do EuroStat e importados para um sistema relacional de base de dados através de processos de Extração Transformação e carregamento (ETL). Posteriormente foi criada uma camada semântica, que faz a ponte de ligação entre os dados das tabelas e os conceitos a analisar. Baseado nesta camada semântica foram criados uma série de interfaces gráficas: um Information Space, uma Exploration View e um *dashboard*. Estas interfaces foram implementadas para a plataforma móvel e para a Web.

A Figura 12 representa o modelo implementado e que será explicado nos pontos seguintes.

¹⁹ Conjunto de componentes tecnológicos físicos de um sistema computacional.

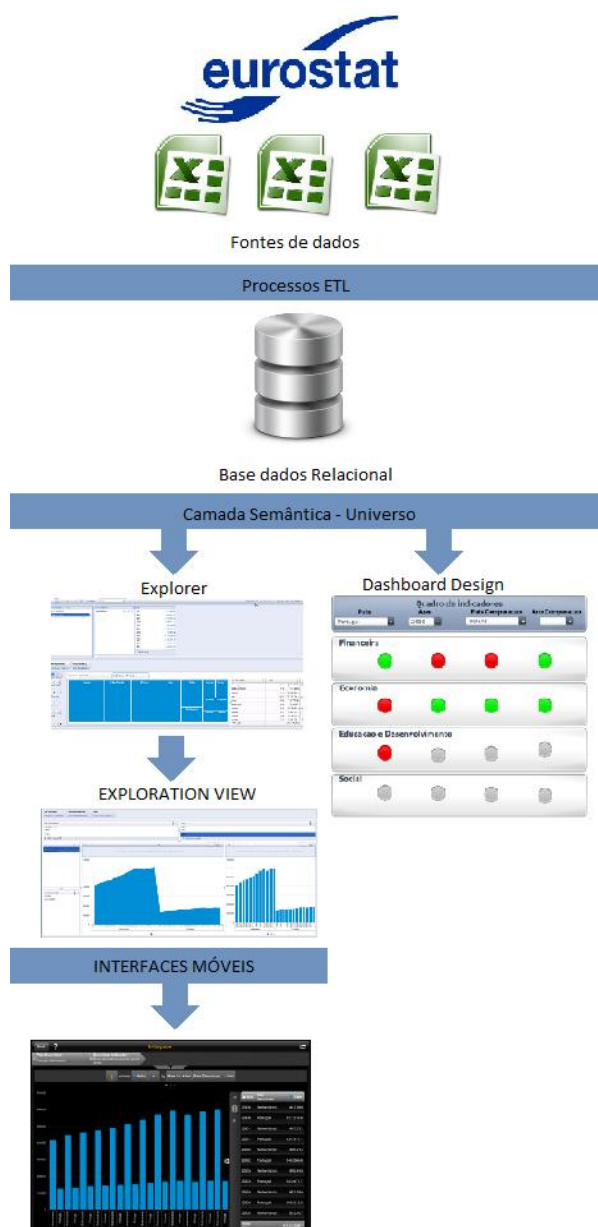


Figura 12 – Modelo da solução implementada

4.3.1 Modelo de dados

Um modelo dimensional é essencialmente composto 2 tipos de tabelas. Tabelas de Dimensão que são tabelas compostas por identificadores únicos aos quais correspondem descrições ; e tabelas factos que são tabelas compostas pelos identificadores únicos das tabelas de dimensão e as métricas a analisar. Face às necessidades de informação requeridas, foi concebido um sistema relacional simples composto por 3 tabelas. sendo uma tabela de factos e 2 tabelas de dimensão.

Tabela de Factos

Tabela MSIAD_MEDICOES

ID_INDICADOR; ID_PAIS; VALOR; ANO

Tabelas Dimensão

Tabela MSIAD_PAIS

ID_PAIS; PAIS

Tabela MSIAD_INDICADORES

ID_INDICADOR; DESCRICAO; POLARIDADE; FONTE; UNIDADE; AREA_ID

Este modelo foi concebido por forma a respeitar um modelo funcional de bases de dados, mas também por forma a facilitar o processo de concepção do interface, tendo em conta a ocupação de espaço físico no dispositivo de armazenamento.

4.3.2 Fontes de dados

No *site* do EuroStat a informação está organizada numa estrutura em árvore por áreas temáticas e áreas de interesse – Figura 13



Figura 13 – Estrutura de dados do EuroStat.

Os dados foram obtidos por exportação para o formato Excel diretamente do site do EuroStat, no dia 2012.09.30.

A sua escolha foi feita tendo por base aqueles indicadores que, depois de feita investigação, pareceram ser indicadores importantes para a caracterização de um país.

Para tirar uma fotografia do país foram escolhidas 4 vertentes: Social, Educacional e Desenvolvimento, Económica e Financeira. A estas está associada um horizonte temporal de

impacto na sociedade. Sendo a vertente financeira o momento atual do país, a vertente económica terá um impacto no momento atual mas também num horizonte temporal de 2 ou 3 anos. Já a vertente educacional terá um impacto a médio prazo (10 anos), ao passo que a vertente social tem um impacto a longo prazo (20 anos).

Assim para cada uma das áreas, os indicadores escolhidos foram:

Social - População total; População em risco de exclusão social; Desigualdade social; Violência;

Educacional e Desenvolvimento / Capital Humano - Abandono escolar precoce; Investimento em Inovação e Desenvolvimento; População com qualificações médias ou superiores; Desenvolvimento de inovação e tecnologia;

Económica - Indústria; Agricultura; Serviços e Desemprego;

Financeira – PIB; Inflação; Exportações; Consumo final.

Alguns indicadores foram particularmente difíceis de trabalhar. Definir qual seria a sua vertente primordial e qual a sua área de maior impacto foram um desafio. O desemprego é um desses casos. Deverá ser considerado como indicador social? Económico? Ou mesmo financeiro?

Naturalmente que o sistema está sujeito a interpretações várias, conforme os indicadores escolhidos. No entanto esta análise representa uma abordagem. Todos os temas sociais, económicos, financeiros ou educacionais carecem de uma análise de especialistas. Especialistas esses que poderão, com maior grau de profundidade, estabelecer quais os indicadores mais relevantes para a área em estudo.

4.3.3 Processo *ETL*

Para o processo de integração de dados foi utilizada a ferramenta de ETL SAP BusinessObjects Data Services (ver informação técnica no Anexo 6.3 – Manual Técnico Data Services)

De forma a permitir a sua inter-relação, os dados foram posteriormente importados para um sistema de bases de dados. O processo de carregamento de ficheiros é um processo linear, pois trata-se de um processo que já tem nas suas fontes de dados a definição de *Primary Keys* (PK), bem como a normalização de nomes ou integridade de relações.

Os ficheiros foram mapeados utilizando a definição de importação de ficheiros Excel (Figura 14). Na identificação da fonte de dados é indicada qual a localização do ficheiro a importar, quais as *Sheets* e quais os campos e a sua caracterização a nível do formato.

Os ficheiros importados estão no formato Excel 2007 e contêm os cabeçalhos das colunas. A sua tipificação quanto ao formato e dimensão é identificado na localização do ficheiro.

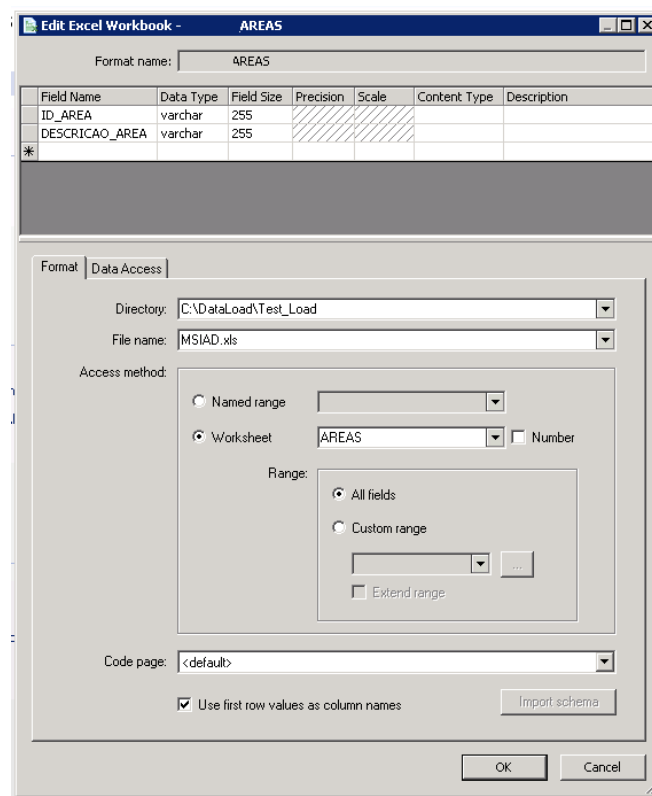


Figura 14 – Importação dos ficheiros fonte.

O processo consiste de um Batch Job (Figura 15) , que integra em si 2 Data Flows. Um para o carregamento dos dados, tal como vêm dos ficheiros, e outro para fazer o tratamento complementar de dados, como é o caso de indicadores que se baseiam em outros indicadores. Um destes casos é o PIB per Capita, que está dependente da população total e do PIB.

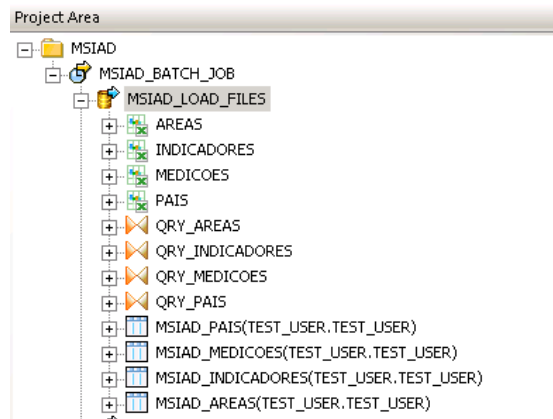


Figura 15 – Estrutura do Processo de ETL em SAP Business Objects Data Services.

O mapeamento (Figura 16) foi efectuado com definição da fonte de dados, seguido de uma query transformation, e finalmente qual a tabela target.

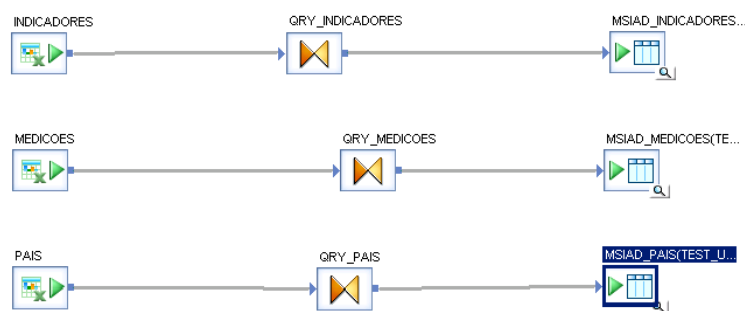


Figura 16 – Mapeamento das fontes de dados e das tabelas associadas.

A figuras na esquerda da imagem representam os ficheiros fonte, com identificação de cada um dos campos. Nos objectos de *Query Transformation*, no centro da imagem, é definido todo o tipo de transformações para os dados a importar. Nos objectos do lado direito, *Targets*, é identificado qual a tabela destino dos dados a importar.

4.3.4 Sistema Relacional

Os dados foram posteriormente passados do formato de Excel para uma base de dados (Oracle 11g neste caso). Para tal, foi utilizada a ferramenta de ETL (Extract Transform and Load) da SAP Business Objects, Data Services XI, Version 12.2.

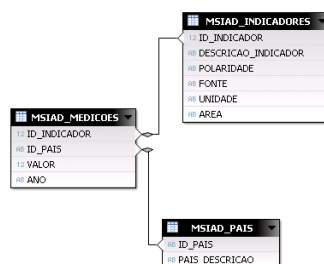


Figura 17 – Modelo físico das tabelas.

4.3.5 Criação da camada semântica

Para o processo de criação de camada semântica, que fizesse a ligação entre os conceitos técnicos de bases de dados e a área de análise, foi utilizada a ferramenta SAP BusinessObjects Information Design Tool (ver informação técnica no Anexo 6.4. – Manual técnico Information Design Tool)

Na arquitectura de BI (Business Intelligence) da SAP Business Objects, os objectos de dados (Tabelas ou Views), relacionam-se numa camada semântica, chamada Universo.

Um universo é um ficheiro que contém:

- Parâmetros de conexão a um ou vários middlewares²⁰ de base de dados;
- Estruturas SQL denominadas objectos que representam as estruturas SQL reais da base de dados, tais como tabelas e funções de bases de dados. Os objectos estão agrupados em classes e tanto os objectos como as classes são visíveis para os utilizadores dos relatórios gerados a partir de um universo;
- Um modelo de dados de tabelas e combinações utilizadas numa base de dados. Os objectos definem-se a partir das estruturas da base de dados.

Esta camada tem por objectivo criar um ambiente mais *user friendly*, permitindo que cada uma das colunas que se pretende analisar tenham um aspecto mais familiar na perspectiva do analista de dados.

O desenvolvimento do universo é efectuado utilizando a ferramenta Information Design Tool.

²⁰ O Middleware é o conjunto de software que faz o interface entre o sistemas operativo e as restantes aplicações.

Esta ferramenta oferece uma interface gráfica para seleccionar e ver as tabelas de uma base de dados. Num modelo de dados as tabelas da base de dados estão representadas por símbolos de tabela. Pode-se utilizar esta interface para manipular as tabelas, criar combinações entre tabelas, criar tabelas com alias, criar contextos e resolver os loops do modelo.

O Universo desenvolvido inclui classes (Indicadores, Medições, e País). Dentro destas classes estão mapeados os objectos que podem ser Dimensões (quadrados azuis), no caso de ser um objecto que caracteriza um valor, ou Métricas, (retângulos cor de laranja) no caso de ser o valor que se pretende analisar.

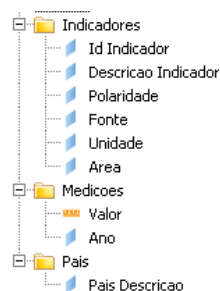


Figura 18 – Estrutura da camada semântica.

4.3.6 Criação de um *report*

Definida esta camada semântica, poderá ser utilizada a vertente de *reporting*. Neste caso foi utilizada a ferramenta Web Intelligence (Figura 22).

Web Intelligence permite que os utilizadores criem relatórios, realizem consultas ad-hoc, analisem dados e apliquem formatação numa interface *drag&drop* (Figura 21). Web Intelligence oculta a complexidade das origens de dados subjacentes. Nesta ferramenta utilizam-se os objectos do Universo para definir quais os objectos que se pretende analisar e/ou utilizar como filtro.

Objetos de resultado

Filtros de consulta

Para filtrar a consulta, arraste os filtros predefinidos para esta área ou arraste os objetos para esta área, em seguida, use o Editor de Filtros para definir filtros personalizados.

Visualização de dados

Pais Descricao	Valor	Ano	Descricao In...	Unidade	Descricao Ar...
Bulgaria	7310.5	2013	Total populati...	1000 persons	SOCIAL
France	65894.6	2013	Total populati...	1000 persons	SOCIAL
Poland	38152.6	2013	Total populati...	1000 persons	SOCIAL
Hrvatska	322.7	2013	Total populati...	1000 persons	SOCIAL
Turkey		2013	Total populati...	1000 persons	SOCIAL
Austria	208473.6	2000	Gross dome...	Millions of eu...	FINANCEIRO
Sweden	268252.6	2000	Gross dome...	Millions of eu...	FINANCEIRO
European Un...	958899.4	2001	Gross dome...	Millions of eu...	FINANCEIRO
Germany	2101900	2001	Gross dome...	Millions of eu...	FINANCEIRO
Spain	680397	2001	Gross dome...	Millions of eu...	FINANCEIRO
Latvia	9216.2	2001	Gross dome...	Millions of eu...	FINANCEIRO
Slovenia	22828.2	2001	Gross dome...	Millions of eu...	FINANCEIRO
Sweden	253743.2	2001	Gross dome...	Millions of eu...	FINANCEIRO
European Un...	9951999.9	2002	Gross dome...	Millions of eu...	FINANCEIRO
Euro area (1...	7328611	2002	Gross dome...	Millions of eu...	FINANCEIRO
Denmark	184743.6	2002	Gross dome...	Millions of eu...	FINANCEIRO
Spain	729258	2002	Gross dome...	Millions of eu...	FINANCEIRO

Figura 19 – Painel de query em Web Intelligence

4.3.7 Criação de um espaço de exploração

Uma outra forma da análise de dados desenvolvida foi a utilização de espaços de exploração. O espaço de exploração (Information Space) é uma vertente na qual podemos analisar os dados, utilizando um cubo de informação pré-calculado e que permite ao utilizador descobrir um pouco mais sobre a informação contida nesse espaço, sem necessidade de reprocessar a query cada vez que seja feita uma pequena alteração. A informação é pré processada e posteriormente colocada em disco, e efectuada a sua indexação. Desta forma o tempo de acesso aos dados é acelerado. Assim, é permitido ao utilizador explorar os dados, sem que tenha uma ideia da informação, e ir descobrindo, conforme vai utilizando o espaço. No processo de criação do espaço de informação são definidas as métricas e as dimensões pela qual se quer analisar estas métricas.

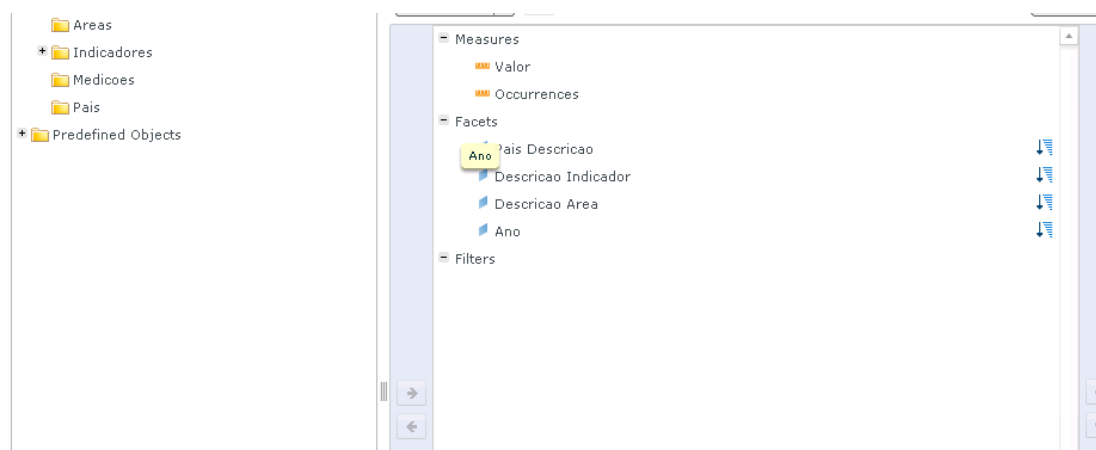


Figura 20 - Criação do espaço exploração.

Neste caso de estudo foi definido o valor do indicador como métrica a analisar, sendo que o número de ocorrências é uma métrica predefinida e sempre incluída em qualquer espaço de informação criado. Como dimensões foram definidas o Ano, o País, a Área e o Nome do indicador. Desta forma será possível analisar as tendências dos principais indicadores pelas várias dimensões de análise. A Figura 21 representa uma possível visualização de dados através de um espaço de exploração no *browser* de um computador.

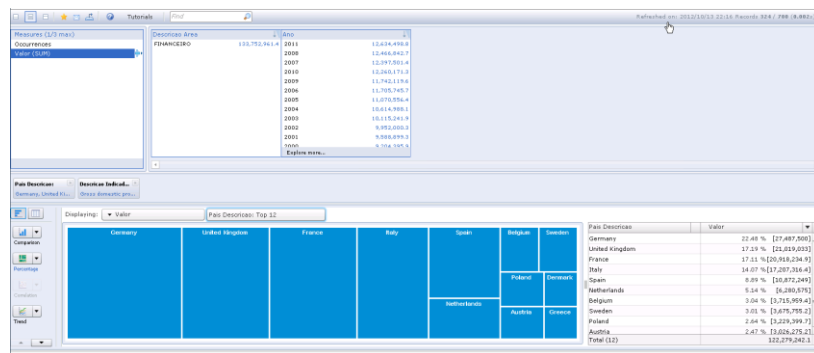


Figura 21 - Visualização do espaço de exploração.

Com o recurso ao espaço de exploração criado, foi possível construir uma Visualização de Exploração - *Exploration View*- (Figura 22). Esta visualização de exploração consiste na criação de uma visualização pré-definida utilizando as várias representações gráficas e definindo quais as dimensões pela qual poderá ser feita a filtragem.

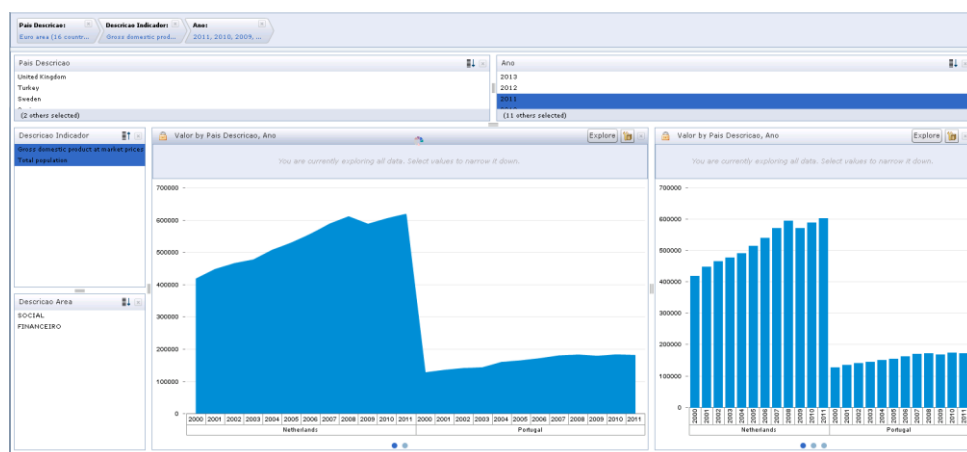


Figura 22 – Visualização de uma Exploration view no Browser.

4.3.8 Interfaces móveis (APP)

Foi um dos objetivos definidos para este trabalho criar interfaces visuais para dispositivos móveis (Figura 23). Os interfaces que foram concebidos para o ambiente de browser, foram posteriormente integrados para poderem ser utilizados nos mais variados dispositivos móveis, fossem *Smartphones* ou *Tablets*. Com a instalação e configuração de uma aplicação disponível na Apple Store ou na Google Play é possível a aceder tanto ao Information Space como a Exploration View.

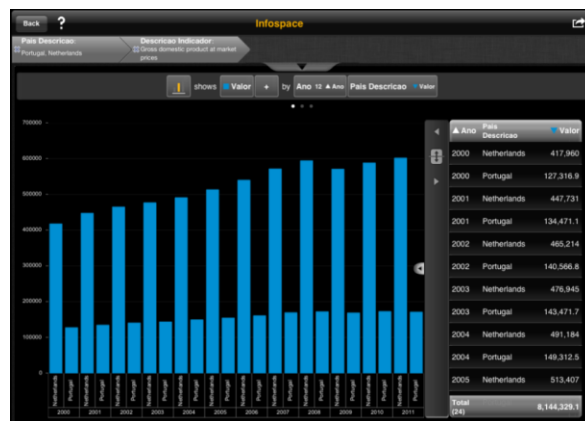


Figura 23 – Display de um espaço de exploração no iPad

A visualização conseguida é em tudo semelhante à exploração conseguida no browser, mas a utilização de um equipamento com ecrã táctil torna a experiência mais natural, sem a necessidade de recurso ao *mouse* do computador- Figura 24

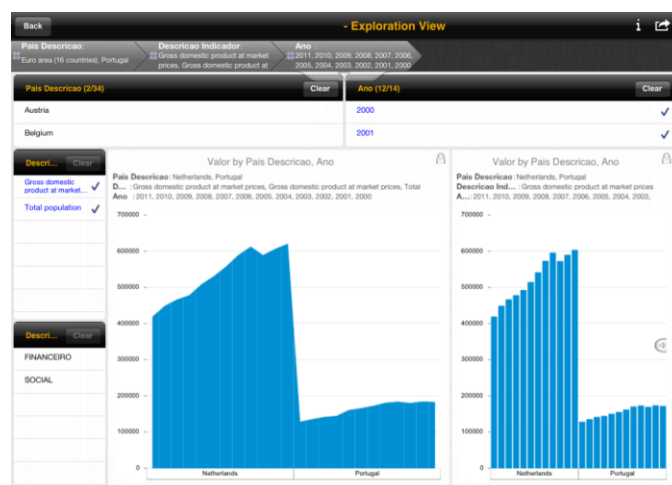


Figura 24 – Exploration View iPad.

4.4. *Dashboard*

As formas de representação de informação analítica anteriormente apresentadas procuravam colocar o utilizador no papel do explorador da informação e descobrir por si mesmo a informação. No entanto, esta forma de exploração poderá levar o utilizador a perder-se na informação e continuar a não ter uma visão geral.

O *dashboard* aqui apresentado (Figura 25) constitui a ideia inicial desta dissertação. O seu propósito surgiu de uma necessidade identificada: Como ajudar a dar uma ideia geral do país, nas suas várias perspectivas, de uma forma simples, intuitiva, e objectiva?

Segundo Stephen Few, um Dashboard é uma "representação visual da informação mais importante necessária para atingir um ou mais objectivos, que cabe completamente num ecrã de forma a poder ser monitorizada num relance" [13].

Com o *dashboard* desenvolvido e recorrendo aos dados anteriormente integrados, procura-se dar uma visualização simples, que consiga reflectir as várias vertentes do país em análise.

O modelo de *dashboard* proposto consiste num quadro, dividido em 5 áreas.

Na zona superior será possível, através de *DropBrox* escolher qual o país que pretendemos analisar e qual o ano de referência, bem como qual o país de comparação e em que ano.

As outras 4 zonas correspondem a indicadores de cada uma das áreas escolhidas para a análise: Social, Educação e Desenvolvimento, Financeira e Económica. Estas áreas contêm círculos coloridos que apresentam cada um dos indicadores escolhidos. Ao fazer passar o rato sobre cada de um dos indicadores será apresentada uma caixa com o nome do mesmo.

Numa perspectiva de melhoria de interface, e segundo estudos Jakob Nielsen²¹, a interface gráfica deveria incluir grafismos específicos que ajudassem a evitar a navegação desnecessária para consumo de conteúdos informativos.

²¹ Investigador na área de desenho de interfaces e interfaces homem-máquina



Figura 25 – Aspeto geral do quadro de indicadores.

As bolas adquirem a cor de acordo com a tendência de comportamento positivo, negativo, neutro ou sem dados apresentados pelo indicador associado.

Verde para um desempenho positivo em relação ao país e período escolhido, vermelho para um desempenho negativo, amarelo para um desempenho neutro e cinzento para situações de ausência de dados para o período e/ou país escolhido.

Clicando num indicador, é possível obter uma apresentação gráfica do mesmo ao longo dos anos e tendo como termo de comparação o valor do indicador para o outro país e ano seleccionados- Figura 26.



Figura 26 – Detalhe para um indicador escolhido.

Desta forma, é possível relacionar valores entre os vários países, e assim tirar conclusões sobre os valores absolutos para um país em comparação com outro, analisar variações, ter uma noção da proporcionalidade das variações, criando um contexto adicional de relação e evitar que os valores sejam analisados de forma isolada.

O *dashboard* desenvolvido procura ser uma ferramenta interativa, e que através da sua utilização e com o recurso de um sistema simples de cores, seja possível chamar a atenção do utilizador e ajudar a perceber quais as áreas mais desenvolvidas ou sub desenvolvidas. O objectivo desta interface é possibilitar que um utilizador, mesmo sem um elevado grau de conhecimento informático, possa obter uma visão geral da sociedade. Visão essa muito necessária na perspectiva da tomada de decisão.

5. Conclusões

5.1. Conclusão do estudo de caso

O atual contexto social, político e económico tornou-se cada vez mais complexo, e a nossa capacidade de assimilar a informação que nos chega pelos meios de comunicação é limitada. As fontes de informação são múltiplas e algumas mais fidedignas que outras.

O objetivo deste trabalho foi procurar contribuir de alguma forma para ajudar a resolver este problema.

Pela curiosidade suscitada por alguns temas, fui ficando cada vez mais interessado pela área da Visualização da Informação e da Visualização Analítica e pela forma como estão atualmente a mudar o modo como se aborda a análise de grandes quantidades de informação, ajudando a uma mais rápida assimilação dos conteúdos.

Com o estudo destes temas e o seu crescente impacto na perceção da informação, apercebi-me que esta poderá ser uma das formas de lidarmos com a torrente de informação. Criar formas de transmitir a informação de forma objetiva, intuitiva e perceptível é um desafio. Combinando os conhecimentos adquiridos durante o mestrado e os conhecimentos profissionais, propus-me a fazer um sistema simples de forma a disponibilizar alguma informação.

Recorrendo à informação do EuroStat, como instituição credível e único fornecedor de estatísticas a nível europeu, procurei transmitir a objetividade e credibilidade necessária e essencial. Estes dados serviram como ponto de partida para a criação de um Sistema de Apoio à Decisão.

A ideia foi criar um *dashboard* que ajudasse a aclarar a situação de um país nas vertentes Social, Educacional e Desenvolvimento, Económica e Financeira, e o fizesse de forma clara. A esta ideia inicial foram-se juntando outras ideias e identificadas outras necessidades. Permitir ao utilizador "descobrir" a informação por si mesmo ou poder ter acesso à informação de forma móvel, foram desafios que me pareceram alcançáveis.

Utilizando o *dashboard* como ponto de partida, pretende-se dar uma ideia geral da situação social, educacional, económica e financeira de um país. Pretende-se que suscite curiosidade. Querer investigar e saber mais. E que esta informação, depois de observada e interiorizada, gere o conhecimento para poder tomar decisões mais conscientes e informadas.

Ao colocar-se as tecnologias de informação ao serviço dos cidadãos, ao ajudar a clarificar alguma dessa informação, estaremos a contribuir para que os processos de tomada de decisões sejam mais sustentados.

5.2. Trabalho Futuro

Como trabalho futuro, considero relevante a criação de processos automatizados para a importação e tratamento de dados. Este processo, apesar de não ser de um elevado grau de dificuldade face ao trabalho já desenvolvido, implicaria um interface automatizado para importação de dados. Esta importação de dados está limitada pelo acesso disponibilizado através do próprio *site* do EuroStat, não sendo possível neste momento uma importação automatizada.

Será também relevante o desenvolvimento de *icons* que ajudem a perceber qual o indicador associado em cada um dos círculos no *dashboard*.

Outro passo importante passa pela divulgação do trabalho desenvolvido. Esta divulgação poderá passar naturalmente pela publicação científica sobre o tema e eventualmente pelos meios de comunicação tradicionais ou pelo recurso a redes sociais.

Bibliografia

- [1] **Short, James E. & Bohn, Roger E. & Baru, Chaitanya** – “How Much Information? 2010 - Report on Enterprise Server Information” – 2011
http://hmi.ucsd.edu/pdf/HMI_2010_EnterpriseReport_Jan_2011.pdf
- [2] **Hallowell, Edward M.** - "Overloaded Circuits: Why Smart People Underperform" –2005- Harvard Business Review – <http://hbr.org/2005/01/overloaded-circuits-why-smart-people-underperform/ar/1> - acessado em 15/07/2012
- [3] **Bertin, Jacques** – “Semiologie Graphique” - 1967 – Editions de l'Ecole des Hautes Etudes en Sciences Sociales - ISBN:978-2713220272
- [4] **Edward R. Tufte** – “The Visual Display of Quantitative Information” – 1983 - Graphics Press USA - ISBN:978-0961392147
- [5] **Stuart K. Card, Ben Shneiderman e Jock D. Mackinlay** – “The cognitive coprocessor architecture for interactive user interfaces” – 1989 - ACM Press – ISBN10:0-89791-335-3
- [6] **Stuart K. Card, Ben Shneiderman e Jock D. Mackinlay** - “Readings in Information Visualization: Using Vision to Think” – 1999 – Morgan Kaufmann Publishers - ISBN:978-1558605336
- [7] **Colin Ware** – “Information Visualization: Perception for Design” – 2000 - Morgan Kaufmann Publishers - ISBN:978-1558605114
- [8] **Daniel Keim , Jörn Kohlhammer , Geoffrey Ellis and Florian Mansmann** - "Mastering the Information Age - Solving problems with visual analytics " – 2010 - <http://www.vismaster.eu/wp-content/uploads/2010/11/VisMaster-book-lowres.pdf> - acessado em 08/02/2012
-

[9] **Tor Norretranders** - “The User Illusion: Cutting Consciousness Down to Size” – 1999 - Penguin Books - ISBN: 978-0140230123

[10] **Thomas, James J. & Cook, Kristin A.** - "Illuminating the Path: The R&D Agenda for Visual Analytics" – 2005 - National Visualization and Analytics Centre

[11] **Keim, Daniel A. & Mansmann, Florian & Schneidewind, Jorn & Ziegler, Hartmut** - “Challenges in Visual Data Analysis” – 2006 - <http://www.inf.uni-konstanz.de/gk/pubsys/publishedFiles/KeMaSc06.pdf> - acedido em 06/09/2012

[12] **Stephen Few** – “Now You See it” - 2009 - Analytics Press – ISBN : 978-0970601988

[13] **Stephen Few** - “Information Dashboard Design” – 2006 - O'Reilly Media - ISBN: 978-0596100162

6. Anexos

6.1. *EuroVis2011.*

Em Maio de 2011 realizou-se a conferência EuroVis 2011 em Bergen, Noruega. Durante 4 dias (31 de Maio a 3 Junho) foram apresentados vários temas ligados à visualização da informação e Visualização Analítica

No âmbito da Visualização Analítica os temas abordados foram:

- Visual Analytics Frameworks and Tools ;
- Visual Analytics Theory;
- Visual Analytics Methodologies;
- Visual Analytics Applications;
- Visual Analytics for Graph Visualization.

Na parte dedicada à Visualização de Informação foram abordados os temas:

- Graphs and Diagrams;
- Illustrative Visualization;
- Geospatial Visualization;
- Topology;
- Scalar and Vector Data ;
- Groups and Graphs;
- Large Graphs;
- Time-Series Data Uncertainty ;
- Medical Visualization ;
- Visualization and Analysis for Life Sciences ;
- Time Series Data;
- Flow Visualization;
- Text and Documents;
- Visual Analysis in Science and Engineering;
- Interaction Design and Usability;
- Visualization for Decision Support ;
- Volume visualization;
- Multi-Dimensional and Multi-Variate Data.

A VA encontra-se ainda num estado mais embrionário sendo que atualmente ainda se discutem sobretudo metodologias e conceitos teóricos. É no entanto um processo natural face a uma área que ainda está no início da sua autonomização.

Sendo a maioria da audiência constituída por elementos do mundo académico, a conferência desenrolou-se através da apresentação de vários *papers* de investigação, sendo alguns deles apoiados por empresas para áreas específicas. Na audiência estavam também presentes elementos ligados a algumas grandes empresas da área tecnológica, nomeadamente IBM e Microsoft.

Foram vários os trabalhos apresentados relativos à VA. Pela sua clareza e componente prática destaque 2.

Um dos trabalhos apresentados que captou maior atenção, foi a análise das ofertas registadas num *site* que tem um sistema de pagamento por licitação, "*Visual Analysis of User Behaviour in Pay-Per-Bid Auctions*"²². Através de uma análise visual, foi possível retirar conclusões quanto ao comportamento dos indivíduos e perceber algumas das estratégias para vencer um leilão e o seu comportamento de licitação. Na Figura 27 é possível analisar o comportamento de um utilizador X em Março (a) e Maio (b) de 2010 e a forma como uma mudança de estratégia justificava um aumento de vitórias nos leilões desde Maio. Até Março a sua estratégia passava pela licitação manual, e depois por licitação automática. A partir de Maio a sua estratégia mudou para um padrão de licitações automáticas, entrando apenas no leilão numa fase posterior. Esta mudança de estratégia levou a um aumento significativo dos seus ganhos.

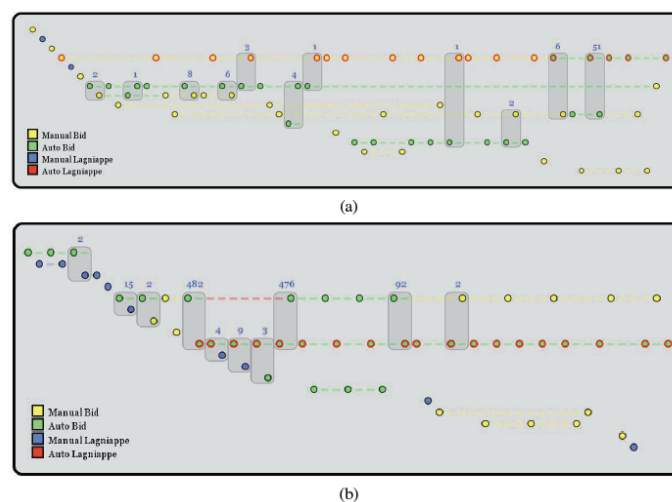


Figura 27 – Comportamento de licitação de um utilizador em 2 períodos de tempo.

²² Walker, Rick & Serban, Llyr ap Cenydd Pop & Roberts, Jonathan C. -School of Computer Science, Bangor University, UK

Outro trabalho muito interessante prendia-se com a utilização de telemóveis “Smartphones” ao longo do dia.

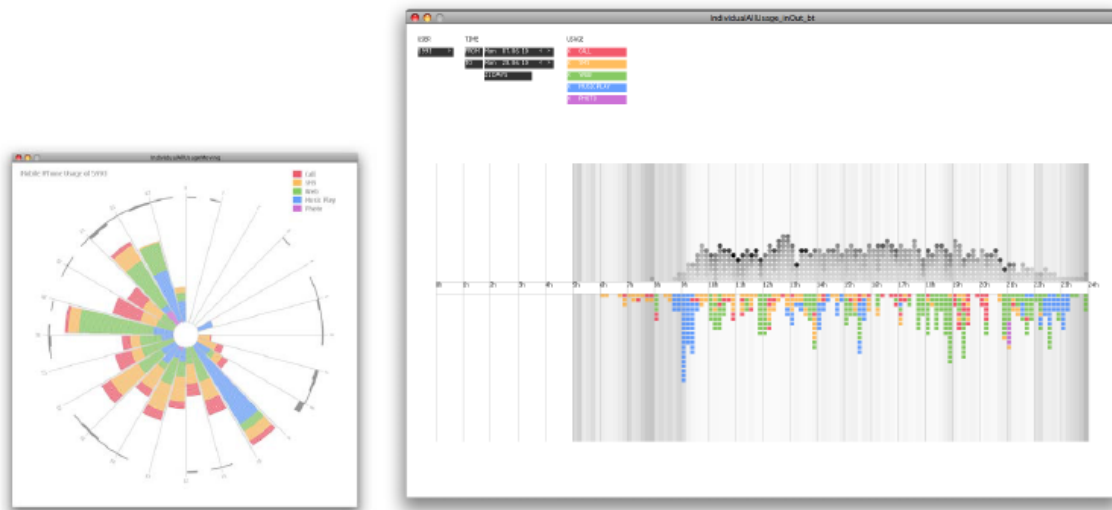


Figura 28 – Visualização da utilização de um smartphone durante o dia

Este *paper* intitulado "Exploring Complex Mobile Life through Lightweight Visualizations"²³, utilizou 2 tipos de visualização. A representação radial (Figura 11 - lado esquerdo) representa os intervalos das 24 horas do dia, um pouco à imagem de um relógio mas com 24 horas, e mostra os vários padrões de utilização do *SmartPhone*, sendo o vermelho chamadas, laranja SMS, verde acesso web, azul para utilização do leitor de música e roxo para a utilização da máquina fotográfica. Esta representação permitiu analisar padrões de utilização e levantou questões. Quais seriam os motivos por trás dos picos de utilização.

Apesar de uma visualização rica, a informação analisada era ainda insuficiente, e não permitia tirar algumas conclusões. Foi assim criada outra visualização (Figura 28 - lado direito) com maior nível de informação. Neste caso através da representação em formato *Timeline*, foi possível adicionar mais informação à representação gráfica. Através da utilização do GPS, foi possível definir o fundo da *Timeline* (igualmente representando as 24 horas do dia) de forma a ficar mais cinzento de acordo com a mudança de localização e como tal representando períodos em que o utilizador estaria em movimento. Foi adicionada ainda informação através da análise da comunicação do *Bluetooth*, detetando qual o número de pessoas que estariam à volta igualmente com *Bluetooth* ligado. Apesar de uma margem de erro, foi considerado este

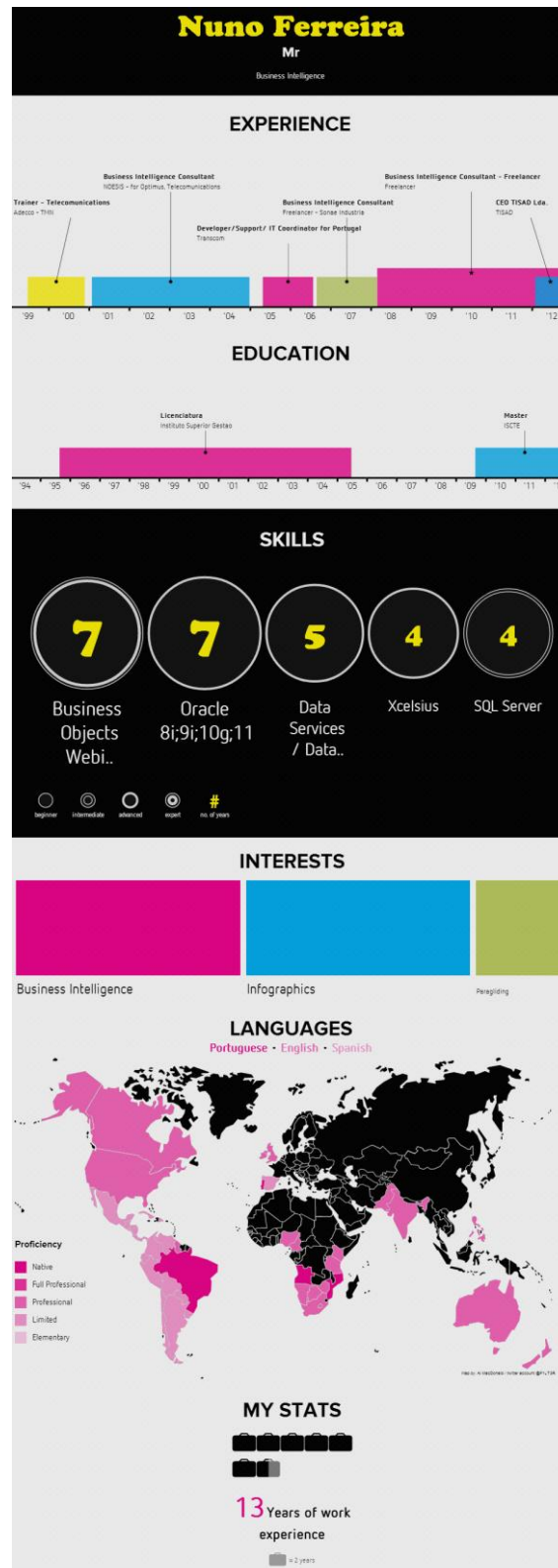
²³ Kim, Tanyoung & Stasko, John - Georgia Institute of Technology, Atlanta, GA, USA - & Blom, Jan Nokia Research Center, Lausanne, Switzerland

padrão para mostrar o número de pessoas à volta do utilizador e qual o seu grau de proximidade, indicado pela intensidade das bolas cinzentas, na parte superior do gráfico. Na parte inferior, utilizando o mesmo padrão de cores, foram indicadas as chamadas, SMS, acesso Web, utilização do leitor de música e utilização da máquina fotográfica.

A análise destes dados permitiu concluir que se tratava de um trabalhador a full-time, com um período matinal de deslocação para um centro urbano e um período vespertino de deslocação para os arredores, que possivelmente estaria inserido num ambiente laboral tecnológico (indicado pelo elevado número de equipamentos com *Bluetooth* ao longo do dia), e que possivelmente almoçava em locais com elevada densidade populacional (através do análise do pico de utilizadores de *Bluetooth*).

Muitos outros trabalhos foram apresentados durante os 4 dias de apresentações intensivas e consecutivas. Durante estes dias foi possível ver qual o grau de abrangência da Visualização da Informação e da Visualização Analítica bem como a sua crescente importância.

6.2. Curriculum Vitae - formato visual



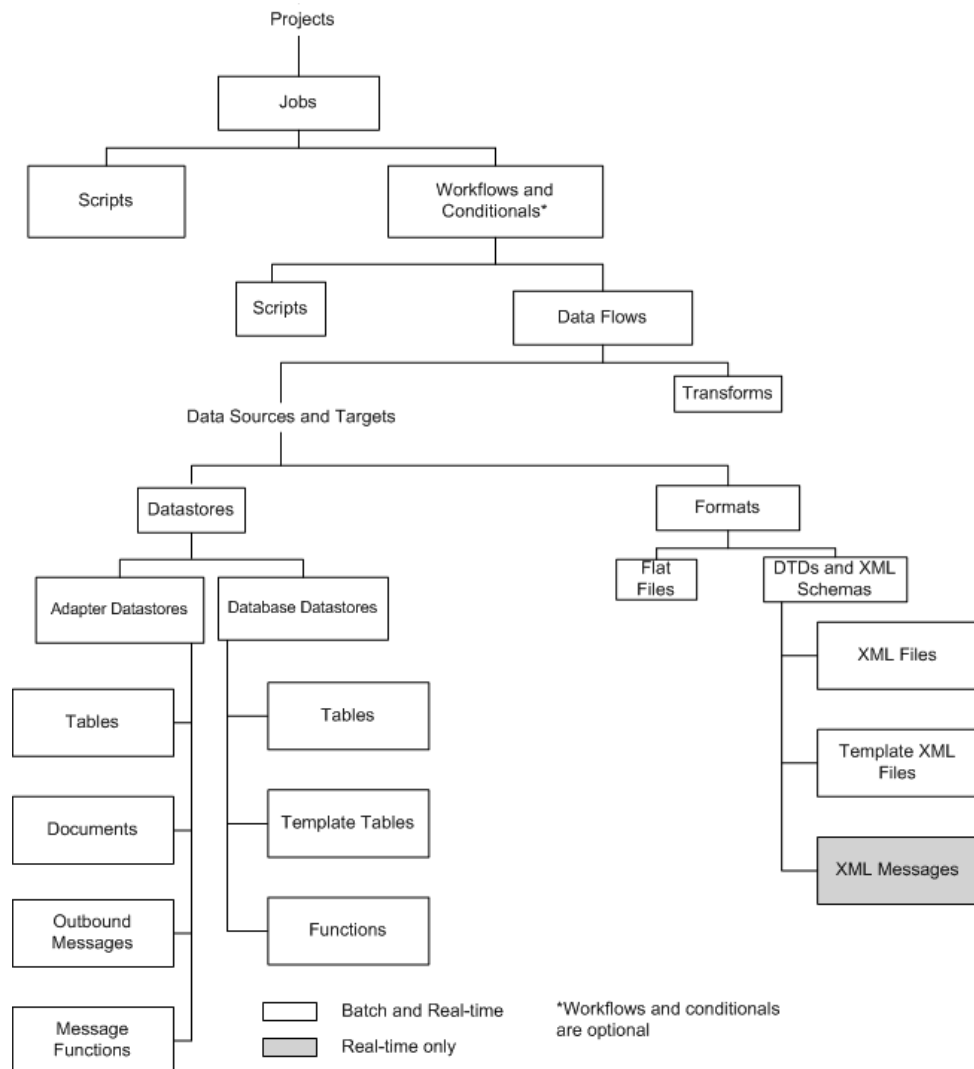
6.3. Manual Técnico Data Services

Este anexo é composto pelo manual técnico de SAP BusinessObjects Data Services²⁴.

Designer User Interface

This section provides basic information about the Designer's graphical user interface.

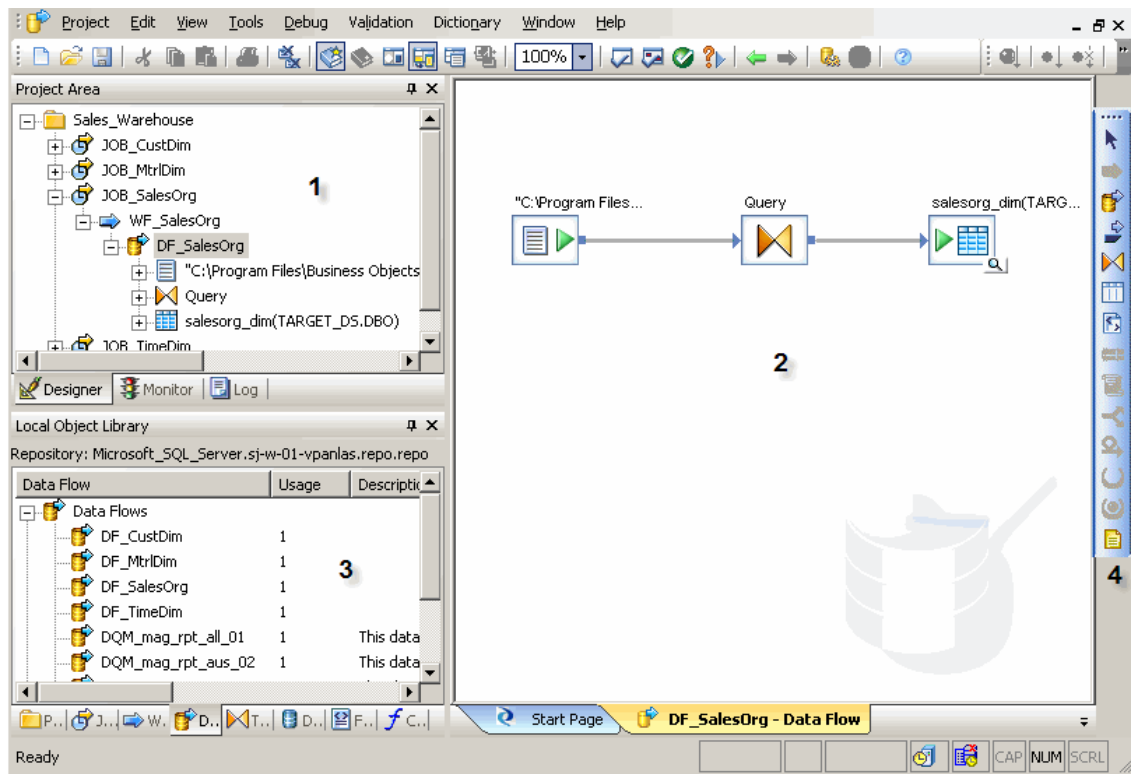
6.3.1 Object hierarchy



²⁴ Manual completo disponível em http://help.sap.com/boall_en/ - acessado em 15/09/2012

6.3.2 Designer window

The Designer user interface consists of a single application window and several embedded supporting windows.



Area	Description
1. Project area	Contains the current project (and the job(s) and other objects within it) available to you at a given time. In the software, all entities you create, modify, or work with are objects.
2. Workspace	The area of the application window in which you define, display, and modify objects.
3. Local object library	Provides access to local repository objects including built-in system objects, such as transforms, and the objects you build and save, such as jobs and data flows.
4. Tool palette	Buttons on the tool palette enable you to add new objects to the workspace.

6.3.3 Menu bar

This section contains a brief description of the Designer's menus.

Project menu

The project menu contains standard Windows as well as software-specific options.

Option	Description
New	Define a new project, batch job, real-time job, work flow, data flow, transform, datastore, file format, DTD, XML Schema, or custom function.
Open	Open an existing project.
Close	Close the currently open project.
Delete	Delete the selected object.
Save	Save the object open in the workspace.
Save All	Save all changes to objects in the current Designer session.
Print	Print the active workspace.
Print Setup	Set up default printer information.
Exit	Exit Designer.

Edit menu

The Edit menu provides standard Windows commands with a few restrictions.

Option	Description
Undo	Undo the last operation.
Cut	Cut the selected objects or text and place it on the clipboard.
Copy	Copy the selected objects or text to the clipboard.
Paste	Paste the contents of the clipboard into the active workspace or text box.

Delete	Delete the selected objects.
Recover Last Deleted	Recover deleted objects to the workspace from which they were deleted. Only the most recently deleted objects are recovered.
Select All	Select all objects in the active workspace.
Clear All	Clear all objects in the active workspace (no undo).

View menu

A check mark indicates that the tool is active.

Option	Description
Toolbar	Display or remove the toolbar in the Designer window.
Status Bar	Display or remove the status bar in the Designer window.
Palette	Display or remove the floating tool palette.
Enabled Descriptions	View descriptions for objects with enabled descriptions.
Refresh	Redraw the display. Use this command to ensure the content of the workspace represents the most up-to-date information from the repository.

Tools menu

An icon with a different color background indicates that the tool is active.

Option	Description
Object Library	Open or close the object library window.
Project Area	Display or remove the project area from the Designer window.
Variables	Open or close the Variables and Parameters window.
Output	Open or close the Output window. The Output window shows errors that occur such as during job validation or object export.
Profiler Monitor	Display the status of Profiler tasks.
Run Match Wizard	Display the Match Wizard to create a match data flow. Select a transform in a data flow to activate this menu item. The transform(s) that the Match Wizard generates will be placed downstream from the transform you selected.

Match Editor	Display the Match Editor to edit Match transform options.
Associate Editor	Display the Associate Editor to edit Associate transform options.
User-Defined Editor	Display the User-Defined Editor to edit User-Defined transform options.
Custom Functions	Display the Custom Functions window.
System Configurations	Display the System Configurations editor.
Substitution Parameter Configurations	Display the Substitution Parameter Editor to create and edit substitution parameters and configurations.
Profiler Server Login	Connect to the Profiler Server.
Export	Export individual repository objects to another repository or file. This command opens the Export editor in the workspace. You can drag objects from the object library into the editor for export. To export your whole repository, in the object library right-click and select Repository > Export to file .
Import From File	Import objects into the current repository from a file. The default file types are ATL, XML, DMT, and FMT. For more information on DMT and FMT files, see the Upgrade Guide.
Metadata Exchange	Import and export metadata to third-party systems via a file.
BusinessObjects Universes	Export (create or update) metadata in BusinessObjects Universes.
Central Repositories	Create or edit connections to a central repository for managing object versions among multiple users.
Options	Display the Options window.
Data ServicesManagement Console	Display the Management Console

Debug menu

The only options available on this menu at all times are **Show Filters/Breakpoints** and **Filters/Breakpoints**. The **Execute** and **Start Debug** options are only active when a job is selected.

All other options are available as appropriate when a job is running in the Debug mode.

Option	Description
Execute	Opens the Execution Properties window which lets you execute the selected job.
Start Debug	Opens the Debug Properties window which lets you run a job in the debug Mode.
View Design-Time Data	Opens data panes in the transform editor, which let you view and analyze the input and output for a data set in real time as you design a transform.
View Automatically	Let you view input and output data automatically after you modify a transform.
Filter Input Dataset	To filter the number of data rows displayed in the Design-Time Data Viewer panes.
Options	Opens a window in which you can configure the number of data rows displayed and the time allowed for updates before a time out.
Show Filters/Break-points	Shows and hides filters and breakpoints in workspace diagrams.
Filters/Breakpoints	Opens a window you can use to manage filters and breakpoints.














Validation menu



The Designer displays options on this menu as appropriate when an object is open in the workspace.

Option	Description
Validate	Validate the objects in the current workspace view or all objects in the job before executing the application.
Show ATL	View a read-only version of the language associated with the job.
Display Optimized SQL	Display the SQL that Data Services generated for a selected data flow.

6.3.4 Toolbar




In addition to many of the standard Windows tools, the software provides application-specific tools, including:

Icon	Tool	Description
	Close all windows	Closes all open windows in the workspace.
	Local Object Library	Opens and closes the local object library window.
	Central Object Library	Opens and closes the central object library window.
	Variables	Opens and closes the variables and parameters creation window.
	Project Area	Opens and closes the project area.
	Output	Opens and closes the output window.
	View Enabled Descriptions	Enables the system level setting for viewing object descriptions in the workspace.
	Validate Current View	Validates the object definition open in the workspace. Other objects included in the definition are also validated.
	Validate All Objects in View	Validates the object definition open in the workspace. Objects included in the definition are also validated.
	Audit Objects in Data Flow	Opens the Audit window to define audit labels and rules for the data flow.
	View Where Used	Opens the Output window, which lists parent objects (such as jobs) of the object currently open in the workspace (such as a data flow). Use this command to find other jobs that use the same data flow, before you decide to make design changes. To see if an object in a data flow is reused elsewhere, rightclick one and select View Where Used .
	Go Back	Move back in the list of active workspace windows.
	Go Forward	Move forward in the list of active workspace

		windows.
	Management Console	Opens and closes the Management Console window.
	Contents	Opens the Technical Manuals PDF for information about using the software.

6.3.5 Project Area

The project area provides a hierarchical view of the objects used in each project. Tabs on the bottom of the project area support different tasks. Tabs include:

 Designer	Create, view and manage projects. Provides a hierarchical view of all objects used in each project.
 Monitor	View the status of currently executing jobs. Selecting a specific job execution displays its status, including which steps are complete and which steps are executing. These tasks can also be done using the Administrator.
 Log	View the history of complete jobs. Logs can also be viewed with the Administrator.

6.3.6 Tool palette

The tool palette is a separate window that appears by default on the right edge of the Designer workspace. You can move the tool palette anywhere on your screen or dock it on any edge of the Designer window.














The icons in the tool palette allow you to create new objects in the workspace. The icons are disabled when they are not allowed to be added to the diagram open in the workspace.

To show the name of each icon, hold the cursor over the icon until the tool tip for the icon appears, as shown.

When you create an object from the tool palette, you are creating a new definition of an object. If a new object is reusable, it will be automatically available in the object library after you create it.

For example, if you select the data flow icon from the tool palette and define a new data flow, later you can drag that existing data flow from the object library, adding a call to the existing definition.

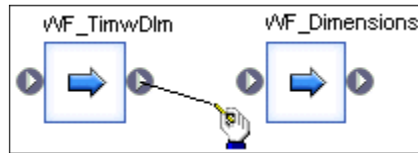
The tool palette contains the following icons:

Icon	Tool	Description (class)	Available
	Pointer	Returns the tool pointer to a selection pointer for selecting and moving objects in a diagram.	Everywhere
	Work flow	Creates a new work flow. (reusable)	Jobs and work flows
	Data flow	Creates a new data flow. (reusable)	Jobs and work flows
	ABAP data flow	Used only with the SAP application.	
	Query transform	Creates a template for a query. Use it to define column mappings and row selections. (single-use)	Data flows
	Template table	Creates a table for a target. (single- use)	Data flows
	Template XML	Creates an XML template. (single-use)	Data flows
	Data transport	Used only with the SAP application.	
	Script	Creates a new script object. (single-use)	Jobs and work flows
	Conditional	Creates a new conditional object. (single-use)	Jobs and work flows
	Try	Creates a new try object. (single-use)	Jobs and work flows
	Catch	Creates a new catch object. (single- use)	Jobs and work flows
	Annotation	Creates an annotation. (single-use)	Jobs, work flows, and data flows

6.3.7 Workspace

When you open or select a job or any flow within a job hierarchy, the workspace becomes "active" with your selection. The workspace provides a place to manipulate system objects and graphically assemble data movement processes.

These processes are represented by icons that you drag and drop into a workspace to create a workspace diagram. This diagram is a visual representation of an entire data movement application or some part of a data movement application.



Moving objects in the workspace area

Use standard mouse commands to move objects in the workspace.

To move an object to a different place in the workspace area:


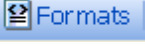

1. Click to select the object.
2. Drag the object to where you want to place it in the workspace.

To open the object library

- Choose **Tools > Object Library**, or click the object library icon in the icon bar.

The object library gives you access to the object types listed in the following table. The table shows the tab on which the object type appears in the object library and describes the context in which you can use each type of object.

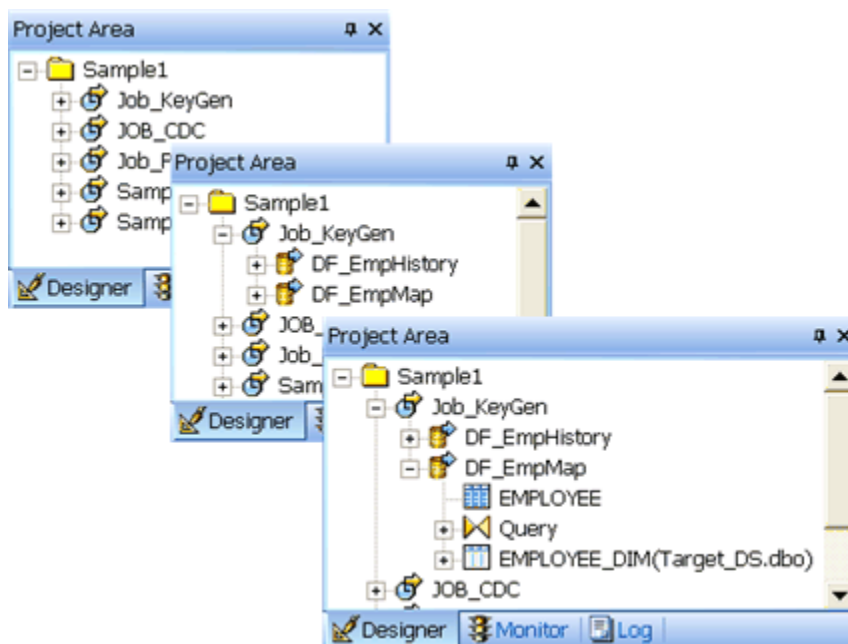
Tab	Description
Projects	Projects are sets of jobs available at a given time.
Jobs	Jobs are executable work flows. There are two job types: batch jobs and real-time jobs.
Work Flows	Work flows order data flows and the operations that support data flows, defining the interdependencies between them.
Data Flows	Data flows describe how to process a task.
Transforms	Transforms operate on data, producing output data sets from the sources you

	specify. The object library lists both built-in and custom transforms.
 Datastores	Datastores represent connections to databases and applications used in your project. Under each datastore is a list of the tables, documents, and functions imported into the software.
 Formats	Formats describe the structure of a flat file, XML file, or XML message.
 Custom Functions	Custom Functions are functions written in the software's Scripting Language. You can use them in your jobs.

Objects that make up a project

The objects in a project appear hierarchically in the project area. If a plus sign (+) appears next to an object, expand it to view the lower-level objects contained in the object. The software shows you the contents as both names in the project area hierarchy and icons in the workspace.

In the following example, the Job_KeyGen job contains two data flows, and the DF_EmpMap data flow contains multiple objects.



Each item selected in the project area also displays in the workspace:



6.3.8 Jobs

A job is the only object you can execute. You can manually execute and test jobs in development. In production, you can schedule batch jobs and set up real-time jobs as services that execute a process when the software receives a message request.

A job is made up of steps you want executed together. Each step is represented by an object icon that you place in the workspace to create a job diagram. A job diagram is made up of two or more objects connected together. You can include any of the following objects in a job definition:

- Data flows
 - Sources
 - Targets
 - Transforms
- Work flows
 - Scripts
 - Conditionals
 - While Loops
 - Try/catch blocks

If a job becomes complex, organize its content into individual work flows, then create a single job that calls those work flows.

Real-time jobs use the same components as batch jobs. You can add work flows and data flows to both batch and real-time jobs. When you drag a work flow or data flow icon into a job, you are telling the software to validate these objects according the requirements of the job type (either batch or real-time).

There are some restrictions regarding the use of some software features with real-time jobs.

6.4. *Manual Técnico Information Design Tool*

Este anexo é composto pelo manual técnico de SAP BusinessObjects Information Design Tool ²⁵.

6.4.1 About the information design tool

The information design tool is an SAP BusinessObjects metadata design environment that enables a designer to extract, define, and manipulate metadata from relational and OLAP sources to create and deploy SAP BusinessObjects universes.

A universe is an organized collection of metadata objects that enable business users to analyze and report on corporate data in a non-technical language. These objects include dimensions, measures, hierarchies, attributes, pre-defined calculations, functions, and queries. The metadata object layer, called the business layer, is built on a relational database schema or an OLAP cube, so the objects map directly to the database structures via SQL or MDX expressions. A universe includes connections identifying the data sources so queries can be run on the data.

The role of the universe is to provide the business user with semantically understandable business objects. The user is free to analyze data and create reports using relevant business language regardless of the underlying data sources and structures.

Universes created using the information design tool can be used by the following SAP BusinessObjects data analysis and reporting applications starting with version BI 4:

- SAP BusinessObjects Web Intelligence
- SAP Crystal Reports for Enterprise
- SAP BusinessObjects Explorer
- SAP BusinessObjects Dashboard Design
- SAP Visual Intelligence (for relational universes starting with SAP BusinessObjects BI platform version 4.0 SP4 and SAP Visual Intelligence version 1.0.3)


To enable the designer to create universes, the information design tool provides the resources necessary to do the following:






²⁵ Manual completo disponível em http://help.sap.com/boall_en/







- Create connections to data sources.
- Extract a complete OLAP cube schema.
- Extract tables and joins to build a relational schema called a data foundation.
- Create metadata objects from the cube or the data foundation. These objects are contained and organized in a business layer. The SQL and MDX expressions within objects can be validated and queries run against the target databases to test the business layer.
- Share resources to allow multiple designers to work on the same resources concurrently.
- Publish a universe, which compiles the business layer, the data foundation, and the connections into a single universe file (.unx):
- Publish a universe to a repository to be implemented in deployments of SAP BusinessObjects data analysis and reporting applications.
- Publish a universe locally, to be implemented by client applications in standalone mode (for example Web Intelligence Rich Client).
- Create security profiles to define user access to universe data and metadata.

6.4.2 Resources in the information design tool

The information design tool provides the following design resources to extract metadata and build universes.

Resource	Description
 Project	<p>A project is a named local workspace that contains the resources used to build one or more universes. A project can be shared so that multiple designers can work on the same resources.</p> <p>A project can contain any number of independent resources, for example data foundations, business layers, and connections. All resources contained within a project can be used interchangeably, for example a connection can be used by several data foundations within the same project.</p> <p>Projects and their resources are displayed in the Local Projects View. To open a resource in the editor, double-click the resource in the Local Projects View.</p>

<p>Connection</p> <p> OLAP</p> <p> Relational</p>	<p>A connection is a named set of parameters that define how a universe can access a relational or OLAP data source. A universe is always associated with at least one connection. A connection is an independent resource and can be used by several universes. You can build a multisource-enabled universe that references one or more relational connections.</p> <p>Connections can be local (stored in a local file) or secured (an object in a shared repository that is referenced by a connection shortcut).</p> <p>Local connections are stored in the local project as .cnx files.</p>
<p> Connection shortcut</p>	<p>A connection shortcut is an object in the local project that references a secured connection in a repository. You use a connection shortcut to refer to secure connections when creating data foundations and business layers based on secure connections.</p> <p>Connection shortcuts are stored in the local project as .cns files.</p>
<p> Data foundation</p>	<p>A data foundation is a schema that defines the relevant tables and joins from one or more relational databases. You enhance the data foundation by adding derived tables, alias tables, calculated columns, additional joins, contexts, prompts, lists of values, and other SQL definitions. The data foundation becomes the basis of one or more business layers.</p> <p>Data foundations are stored in the local project as .dfx files.</p>
<p> Business layer</p>	<p>A business layer is a collection of metadata objects that provides an abstraction of relational database entities or OLAP cubes, understandable by a business user. Objects map via SQL expressions to an underlying data foundation, or via MDX expressions to an underlying OLAP cube. These objects include dimensions, hierarchies, measures, attributes, and predefined conditions.</p> <p>You can add dimensions, hierarchies, measures, attributes, and other objects as the universe design requires. You can validate the SQL or the MDX at any time. You can create queries, lists of values, parameters (also called prompts), and navigation path objects.</p> <p>The business layer is the universe under construction, and when the business layer is complete, it is compiled with the connections or connection shortcuts and data foundation, published, and deployed as a universe.</p>

	Business layers are stored in the local project as .blx files.
 Query	A query is a set of objects that define a request to the database for data. A query can be defined and saved in the business layer as a metadata object to be used to test objects in the business layer.
 Parameter	A parameter is a variable in the universe that requires a value at query time. Parameters are often defined to prompt the user to supply a value, and in this case are referred to as prompts.
 List of values	A list of values is a collection of data values that can be associated with an object in the universe, allowing the user to choose values for a prompt. Parameters and lists of values can be defined in the data foundation. They are inherited by all business layers based on that data foundation. Parameters and lists of values can also be defined in the business layer.
 Universe	A universe is a compiled file that includes all resources used in the definition of the metadata objects built in the design of the business layer. The universe is used by SAP BusinessObjects data analysis and reporting applications, where the business layer objects are visible for analysis and reporting. Universes are stored either locally or in a repository as .unx files.
Security Profiles  Data  Business	A security profile is a group of security settings that controls the data and metadata that are displayed to users and modifies the parameters defined in the data foundation and/or business layer. Security profiles are defined on published universes and stored in the repository.

6.4.3 Working with business layers

About business layers

A business layer is a collection of metadata objects that map to SQL or MDX definitions in a database, for example, columns, views, database functions, or pre-aggregated calculations. The metadata objects include dimensions, hierarchies, measures, attributes, and predefined conditions. Each object corresponds to a unit of business information that can be manipulated

in a query to return data. Business layers can be created directly on an OLAP cube, or on a data foundation that is built on a relational database.

When a business layer is complete, it is published to a repository or a local folder as a universe. A universe is a published .unx file that includes a business layer and its connection to an OLAP cube, or a business layer and its corresponding data foundation. The universe is available in the repository to SAP BusinessObjects data analysis and report creation applications.

The principle role of the business layer is to define and organize metadata before it is published as a universe. An alternative way to understand the business layer is to think of it as a metadata workbench that a designer uses to assemble and modify a metadata set before publishing as a universe for data analysis and report creation applications.

About business layer objects

The **Business Layer** objects pane contains the metadata objects that make up the business layer.






Every object in the business layer has a name that can be edited. When naming objects, use business vocabulary familiar to the users who will be using the universe for querying, analysis, and reporting.

Objects can have three states:

- **Active:** Object is visible in the Query Panel. This is the default state.
- **Hidden:** Object is valid but not available in the Query Panel (used by other objects as a hidden object).
- **Deprecated:** Object is hidden and not valid. One possible use for this state is when the target database field no longer exists, but you want to keep the object for possible future use.

Each object in the business layer has properties that are applied in the published universe. You can define properties when you insert an object, and modify object properties at any time. See the related topics about inserting and editing objects.

Depending on the type of data source, you can create and edit the following types of objects in the business layer:

Object	Description
 Folder	<p>A folder is a container that holds a group of related objects. You create folders to group objects that have a common purpose in the business layer. The folder has no role in a query. It is only used to organize objects.</p>
 Dimension	<p>A dimension is an object that maps to one or more table columns or a function in a database and represents an axis of analysis in a query. For example, Product, Geography, Time, and Employee are common dimensions. Each dimension classifies an aspect of an activity in a business environment.</p> <p>In a business layer, dimensions represent contextual information (the axes of analysis).</p>
 Measure	<p>Measures are objects that represent calculations and aggregate functions that map to statistical and analytic data in the database.</p> <p>In a business layer, measures represent the factual information (data). Numeric data is usually, but not always, the source of a measure.</p> <p>Aggregating the information must make sense for the object to be a measure. For example, summing up sales revenue makes sense, so Sales Revenue is a measure. Summing product list prices is not necessarily useful, so List Price is a dimension, or perhaps an attribute of the Product Dimension.</p> <p>You can create measures from non-numeric objects by counting things. This can result in measures like Number of Orders.</p>
 Attribute	<p>An attribute is an object attached to a parent object that provides additional descriptive information about the parent. Attributes can be defined for dimensions, hierarchies, and levels.</p>
 Filter	<p>A filter is a condition object that limits the data returned in a query. Filters can be inserted into the Query Filters pane in the Query Panel to be applied to the query.</p> <p>Native filters are defined by an SQL WHERE clause on data foundation tables. Native filters apply to business layers based on data foundations.</p> <p>Business filters are defined by creating and combining conditions on dimensions and measures in the business layer.</p>

How to build a relational business layer

Before you begin:

- You need a project in the Local Projects View.
- You need a data foundation saved in the same local project.

For links to more detailed information on each step, see the Related Topics.

1. Right-click the project folder in the Local Projects View and select **New > Business Layer**.
2. Follow the steps in the New Business Layer wizard. Select **Relational data foundation** as the data source.

The business layer is created in a .blx file in the local project folder. It opens automatically in the

Business Layer Editor.

3. Build the business layer:

If you created the business layer with the **Automatically create folders and objects** option selected (default), all objects are created in the business layer as dimensions. You need to specify the measures explicitly using the **Turn into Measure with Aggregation Function** command.

Otherwise, insert the business layer objects:

- a. Insert folders and subfolders to organize the business layer.

Note:

When you drag a table into the business layer from the data foundation, a folder is automatically inserted.

- b. Drag and drop tables and columns into the desired folders and rename the objects if needed.
- c. Specify the measures using the **Turn into Measure with Aggregation Function** command.

4. You can enhance the function of the business layer in several ways, for example:

- Insert attributes to provide descriptive information for dimensions
- Insert additional measures
- Insert pre-defined filters (mandatory or optional) that can limit data returned in queries
- Insert parameters with optional prompts
- Insert custom lists of values to be associated with a prompt

- Insert navigation paths to define drill paths
- Create business layer views to restrict the objects seen in the Query Panel
- Set SQL options and SQL generation parameters in the business layer properties
- Set up aggregate awareness to improve query performance

5. Run an integrity check to validate the dependencies, object expressions, and parameters and lists of values. Right-click the business layer name in the **Business Layer** pane, and select **Check Integrity**.

6. Save the business layer by clicking the **Save** icon in the main tool bar.

Some commands to help you maintain the business layer are listed below.

- If you change object definitions, use **Show Local Dependencies** to find the other business layer objects and data foundation objects that might be impacted by the changes.
- If the related data foundation is modified, you must refresh the business layer manually. The related topic describes commands to help you do this.
- Use **Change Data Foundation** to change the source data foundation for the business layer.
- For multisource-enabled business layers, use **Compute Statistics** to improve query performance.

The business layer is based on a data foundation. You can select any data Relational foundation in the current project folder.

7. Lombada do trabalho

