



Universidade de Évora

Mestrado em Engenharia Informática

**SISTEMA INFORMÁTICO
DE
GESTÃO E ANÁLISE DE GENEALOGIAS**

João Pedro Pimenta Leitão

(Licenciado)

Orientador:

Professor Doutor Carlos Alberto Zany Pampulim Martins Caldeira



177961

Évora - Janeiro de 2011

*Aos meus pais, **Pedro e Maria João;**
Aos meus irmãos, **Nuno e Pedro Miguel;**
À minha futura esposa, **Ana Sofia;**
À demais família e amigos.*

*Os que desprezam os manuscritos genealógicos
atiram fora o melhor oiro da história civil,
política e religiosa de nossa terra.
(**Camilo Castelo Branco**)*

*"Porque é frágil a memória dos homens e para que,
com o tempo, não caiam no esquecimento os feitos
dos mortais, nasceu o remédio da escrita
para que, por meio dele, os factos passados se
conservem como presentes para o futuro."*

(Arquivo do Museu de Grão Vasco – Viseu)

*"Mandas-me, ó Rei, que conte declarando
De minha gente a grã genealogia;
Não me mandas contar estranha história:
Mas mandas-me louvar dos meus a glória."*

*(**Luís de Camões** - Os Lusíadas)*

AGRADECIMENTOS

Não posso concluir este trabalho sem agradecer àqueles que directa ou indirectamente contribuíram para este estudo se tornasse uma realidade. Assim gostaria de expressar a minha gratidão:

- Ao Professor Doutor Carlos Pampulim Caldeira, por todos os esforços, na orientação desta dissertação e pela atenção que sempre dedicou nos assuntos relacionados com a mesma.
- À Professora Doutora, Fernanda Olival, que sempre se prontificou a esclarecer todos os pontos relacionados com o projecto, com a História e as demais áreas para uma correcta compreensão do problema.
- Aos meus pais, irmãos e futura esposa, que viveram comigo as tormentas deste trabalho e a quem devo tudo aquilo que sou, pelo apoio, pelo amor, carinho e paciência essencialmente nos maus momentos, e que sempre me levaram a confiar nas minhas capacidades permitindo assim que conseguisse concluir este estudo no prazo previsto.
- A todos o meu bem-haja!

RESUMO

A selecção do tema e conseqüente trabalho de que emerge o título desta dissertação decorreu do facto de se ter tomado conhecimento da necessidade que os membros do projecto FCOMP-01-0124-FEDER-007360 – Inquirir da Honra: Comissários do Santo Ofício e das Ordens Militares em Portugal (1570 – 1773) tiveram para satisfazer alguns objectivos em particular relacionados com a Genealogia da rede de Comissários. O sistema de trabalho manual que até aqui era utilizado, continha uma quantidade considerável de informação complexa, descrevendo ao pormenor as características não só dos indivíduos, mas também do que estava associado ao mesmo, incluindo quem e como se relacionava com as demais figuras.

O principal objectivo consistiu assim em responder à pergunta: “Como será possível efectuar uma gestão de toda a informação genealógica recolhida no papel e permitir a sua análise no computador, recorrendo a tecnologias que, por um lado sejam eficientes, e por outro, fáceis de aprender pelos utilizadores?”.

Para conseguir responder à questão, foi necessário conhecer em primeira mão, o universo da Genealogia e a forma como opera, para que posteriormente, se desenhasse e moldasse toda uma aplicação às necessidades do utilizador. No entanto, a aplicação não se centra apenas em permitir ao utilizador efectuar uma gestão, recorrendo a um sistema de gestão de bases de dados *MySQL* e permitir a análise genealógica “tradicional” em programas como o *Personal Ancestral File*. Pretende-se sobretudo, que o utilizador faça uso e responda às perguntas “do presente” esperando que a própria aplicação sirva de motivação para novas perguntas, com a integração da tecnologia XML e do Sistema de Informação Geográfico, Google Earth, permitindo assim a análise de informação genealógica no mapa-mundo.

Palavras-Chave:

Base de dados prosopográfica; Genealogia; Integração de Sistemas; Interoperabilidade; KML; MySQL; PAF; Sistema de Informação.

ABSTRACT

The choice of this essay's work subject is set on the need to accomplish determinate goals related with the Genealogy of the network Inquisition Commissioners on behalf of the project FCOMP-01-0124-FEDER-007360 members - Inquirir da Honra: Comissários do Santo Oficio e das Ordens Militares em Portugal (1570 – 1773)- To Inquire Honor: Inquisition Commissioners and the Military Orders in Portugal. The manual work system used till now presented a considerable amount of complex information, describing in detail characteristics not only of individuals but also of what is associated to it, including who and how.

The main goal aimed at thus responding to: «How could it be possible to select and examine all the genealogical data registered on paper and allow it to be analyzed on computer, by means of technology that on one hand are efficient and on other hand easy to learn by its users? »

In order to get to the answer to that matter, it was necessary to acknowledge firstly the Genealogy's universe so afterwards it could be possible to outline and shape an entire application to user needs. Nevertheless, the application does not only focus on allowing the user to carry out the system's management, using MySQL database management system and allowing the "traditional" genealogical management in programs such as the Personal Ancestral File. Above all the user should get involved with it and answer the key questions of 'the present' hoping that the application serves by itself as motivation to arouse new questions with the integration of XML technology and Geographic Information System, Google Earth, thus allowing the analysis of genealogical information worldwide.

Keywords:

Genealogy; Information System; Interoperability; KML; MySQL; PAF; Prosopographic database; Systems Integration.

ÍNDICE DE FIGURAS

FIGURA 1 - EXEMPLO DA LOCALIZAÇÃO DO DEPARTAMENTO DE INFORMÁTICA DA UNIVERSIDADE DE ÉVORA NO GOOGLE EARTH.....	26
FIGURA 2 - EXEMPLO DA APLICAÇÃO DOROTREE 2.1	31
FIGURA 3 - EXEMPLO DA APLICAÇÃO FAMTREE 4.3.....	32
FIGURA 4 - EXEMPLO DA APLICAÇÃO WINFAMILY2009	33
FIGURA 5 - EXEMPLO DA APLICAÇÃO REUNION 9.....	34
FIGURA 6 - EXEMPLO DA APLICAÇÃO THE MASTER GENEALOGIST 7	35
FIGURA 7 - EXEMPLO DA APLICAÇÃO FAMILY HISTORIAN 4.....	36
FIGURA 8 - EXEMPLO DA APLICAÇÃO LEGACY FAMILY TREE 7	37
FIGURA 9 - EXEMPLO DA APLICAÇÃO ROOTS MAGIC 4.....	38
FIGURA 10 - COMPARAÇÃO DE GRAFISMO ENTRE AS APLICAÇÕES ANCESTRAL QUEST E PERSONAL ANCESTRAL FILE	39
FIGURA 11 - COMPARAÇÃO DE GRAFISMO ENTRE AS APLICAÇÕES FAMILY TREE MAKER E FAMILY TREE BUILDER .	41
FIGURA 12 - PORTAL DOS MEMBROS DO PROJECTO SPARES – INQUIRIR DA HONRA.....	83
FIGURA 13 - MENU INICIAL DO SISTEMA INFORMÁTICO DE GESTÃO E ANÁLISES DE GENEALOGIAS.....	85
FIGURA 14 - MENU DE ESCOLHA DAS OPÇÕES POSSÍVEIS PARA A GESTÃO DE GENEALOGIAS	86
FIGURA 15 - GESTÃO DE EGOS NO SISTEMA INFORMÁTICO DE GESTÃO E ANÁLISES DE GENEALOGIAS.....	86
FIGURA 16 - GESTÃO DE GENEALOGIAS NO SISTEMA INFORMÁTICO DE GESTÃO E ANÁLISES DE GENEALOGIAS	87
FIGURA 17 - GESTÃO DE LOCALIDADES NO SISTEMA INFORMÁTICO DE GESTÃO E ANÁLISES DE GENEALOGIAS.....	88
FIGURA 18 - MENU DE ESCOLHA DAS OPÇÕES POSSÍVEIS PARA ANÁLISE DE GENEALOGIAS	89
FIGURA 19 - EXEMPLO DE ERROS NA IMPORTAÇÃO DE FICHEIRO GEDCOM PARA O PERSONAL ANCESTRAL FILE ...	90
FIGURA 20 - LISTAGEM DE EGOS NO PERSONAL ANCESTRAL FILE	91
FIGURA 21 - EXEMPLO DE DADOS RELACIONADOS COM O EGO: FRANCISCO RODRIGUES HOMEM NO PERSONAL ANCESTRAL FILE	92
FIGURA 22 - EXEMPLO DE ANÁLISE DE UMA FAMÍLIA NO PERSONAL ANCESTRAL FILE	93
FIGURA 23 - EXEMPLO DE ÁRVORE GENEALÓGICA NO PERSONAL ANCESTRAL FILE	94

FIGURA 24 - EXEMPLO DE OPÇÕES AVANÇADAS NO PERSONAL ANCESTRAL FILE.....	95
FIGURA 25 - ECRÃ ESCOLHA DE OPÇÕES - ANÁLISE GOOGLE EARTH.....	97
FIGURA 26 - EXEMPLO DE VISUALIZAÇÃO DA GENEALOGIA DO EGO E1 NO GOOGLE EARTH.....	98
FIGURA 27- EXEMPLO DE VISUALIZAÇÃO DA FAMÍLIA (PAI, MÃE E IRMÃOS) DO EGO E35 NO GOOGLE EARTH.....	98
FIGURA 28 – EXEMPLO DE VISUALIZAÇÃO DOS LOCAIS DE NASCIMENTO E FALECIMENTO PARA O EGO: 240 NO GOOGLE EARTH.....	99
FIGURA 29 - VISÃO GERAL DOS LOCAIS DE NASCIMENTO POR EGOS NO GOOGLE EARTH.....	100
FIGURA 30 - VISÃO DETALHADA DOS LOCAIS DE NASCIMENTO POR EGOS NO GOOGLE EARTH	101
FIGURA 31 - VISÃO GERAL DAS LOCALIDADES DE FALECIMENTO - GOOGLE EARTH.....	102
FIGURA 32 - PROCESSO DE CRIAÇÃO DO FICHEIRO COM REPRESENTAÇÃO GEDCOM XML.....	103
FIGURA 33 - EXEMPLO DA VISÃO GERAL DA REPRESENTAÇÃO GEDCOM XML.....	104
FIGURA 34 - EXEMPLO DA DEFINIÇÃO DO EGO: E2623 - CATARINA MARTINS EM FORMATO GEDCOM XML	105
FIGURA 35 - EXEMPLO DA VISUALIZAÇÃO DO FICHEIRO GEDCOM XML NO BROWSER INTERNET EXPLORER.....	106

ÍNDICE

Capítulo I - Introdução.....	1
1.1 Objectivos do Estudo.....	3
1.2 Relevância do Estudo.....	4
1.3 Estrutura da Dissertação.....	6
Capítulo II - Gestão e Análise de Genealogias.....	8
2.1 Fundamentação e Aplicação.....	8
2.2 Definições e Conceitos.....	10
2.2.1 Metadados.....	10
2.2.2 GEDCOM.....	12
2.2.2.1 A Gramática.....	14
2.2.3 XML.....	18
2.2.4 Google Earth e KML.....	21
2.2.5 GEDCOM XML.....	26
2.3 Tecnologias – Estado da Arte.....	29
2.3.1 GEDCOM.....	29
2.3.2 GEDCOM XML.....	44
Capítulo III - Sistema Informático de Gestão e Análise de Genealogias.....	48
3.1 Conceito.....	48
3.2 Importância.....	49
3.3 Origem da Informação.....	51
3.4 Processamento.....	53
3.4.1 Processamento - GEDCOM.....	53
3.4.1.1 Processamento - GEDCOM – Fase 1 - Header.....	54
3.4.1.2 Processamento - GEDCOM – Fase 2 – Informação Individual.....	55
3.4.1.3 Processamento - GEDCOM – Fase 3 – Informação das Relações.....	56
3.4.2 Processamento - XML.....	58
3.4.2.1 Processamento - XML – Fase 1 - Header.....	59
3.4.2.2 Processamento - XML – Fase 2 – Transformação GEDCOM - XML.....	60

3.4.3 Processamento – Google Earth.....	63
3.4.3.1 Processamento - Google Earth – Fase 1 - Header.....	63
3.4.3.2 Processamento - Google Earth – Fase 2 – Criação KML.....	63
3.5 Exportação e Portabilidade	65
Capítulo IV - Caso de Estudo	68
4.1 Descrição do Estudo	69
4.2 Recolha e tratamento de dados: Base de Dados Prosopográfica	70
4.3 Exploração de dados	78
4.4 Portabilidade	82
Capítulo V - Apresentação de Resultados	84
5.1 O Sistema Informático de Gestão e Análise de Genealogias aplicado à Genealogia do Santo Ofício e das Ordens Militares (1570-1773).....	85
5.1.1 Personal Ancestral File 5.2.....	90
5.1.2 Google Earth.....	97
5.1.3 GEDCOM XML	103
Capítulo VI - Conclusão.....	107
6.1 Perspectivas Futuras.....	110
Referências.....	113
Anexos	121
Anexo I – Document Type Definition – GEDCOM XML	121

CAPÍTULO I - INTRODUÇÃO

Em primeiro lugar, importa entender qual é a área que se aplica este estudo. A Genealogia é o ramo da História que se dedica ao estudo das famílias, à sua origem e evolução, descrevendo as gerações em cadeia (em sentido ascendente ou descendente) e traçando, sempre que possível, as biografias dos respectivos membros (Barbosa, 2000).

Desde muito cedo, o homem procurou deixar vestígios de si próprio. Existem vários exemplos capazes de comprovar esta teoria, desde as pinturas rupestres passando pelos hieróglifos e até mesmo os pictogramas, sempre com a finalidade de fazer lembrar algo. Referenciada desde os tempos bíblicos, existem várias alusões para a sua origem (Family Chronicle, 1999), sendo que, para nós o que importa neste estudo, é perceber a partir de que momento esta ciência começa a ser “organizada”, isto é, qual o momento em que se começa a perceber que existem métodos e procedimentos para organizar informação genealógica.

Remontando a história desta ciência a tempos provectos, foi definida por *Otto Forst de Battaglia*, famoso diplomata e genealógico australiano, como “a ciência das relações de parentesco entre os homens”. Para começar pelo ponto em que a Genealogia começa a ser aliada a metodologias de trabalho é necessário registar um dos mais significativos representantes alemães do Iluminismo, *Johann Christoph Gatterer* (1727-99), juntamente com o famoso histórico e genealogista austríaco, *Ottokar Lorenz* (1832-1904) que avançaram como sendo os autores do primeiro processo “organizado” da Genealogia, tendo ambos definido a “teoria das gerações” que acabou por cair em vão, já que apenas mais tarde este avanço veria as suas repercussões (Goffart, 2003).

Até ao início do século XX várias foram as comunidades genealógicas criadas sendo este século (séc. XX) o ponto de partida na aplicação de auxiliares informáticos à fascinante área da Genealogia (About.com: Genealogy, 2009).

Foram várias as soluções encontradas e com o avançar do tempo, a tecnologia também inovou, existindo nos dias de hoje, soluções genealógicas capazes de responder a elevadas exigências.

Este estudo procura apresentar uma proposta concreta para realizar uma gestão cuidada e criteriosa de dados genealógicos assim como permitir uma futura análise desses mesmos dados, seja essa análise feita numa vertente ligada ao estudo biográfico dos indivíduos e respectivas árvores genealógicas, ou análises feitas recorrendo a novas tecnologias e novas formas de caracterização de dados genealógicos, por exemplo, dados genealógicos georreferenciados, isto é, a capacidade de analisar a informação genealógica referenciada no mapa-mundo.

Esta dissertação inclui uma secção dedicada a recursos genealógicos existentes no mercado das aplicações informáticas, sendo que, existe uma panóplia de ofertas geralmente em consonância com as necessidades dos utilizadores.

Escrever um livro sobre a genealogia de uma família, seja em formato papel ou electrónico, requer muito trabalho de investigação, muita perseverança e bastante vontade na procura de informação (Amaral, 2008), já que, qualquer trabalho relacionado com a genealogia, envolve sempre variadíssimas áreas, não sendo por isso um trabalho simples (Correia, 2006). Isto porque a Genealogia raramente se debruça apenas nas simples relações de parentesco, abarcando muitas vezes informações prosopográficas ¹ referindo particularidades e características externas ao indivíduo e que podem caracterizar um espaço ou local (Barbosa, 2000). Muitas vezes aplicada a círculos mais amplos do que a família, aldeias; cidades; países, continentes, etc., a Genealogia alarga-se em demografia,

¹ Nome feminino

- Descrição de caracteres fisionómicos
- Esboço de uma figura

(Do gr. prósopon, «face» +gráphein, «descrever» +-ia)

Dicionário Porto Editora

difícilmente se separando da História no geral, mostrando assim a importância extrema que esta área da História apresenta para o conhecimento de todos nós (Amaral, 2008).

O sistema concebido resultante deste estudo não tem como objectivo “criar a roda”, isto é, não pretende ser mais um sistema que, à sua maneira, efectue a gestão e permita análises “internamente”. O objectivo é conceber um sistema capaz de juntar o melhor de cada tecnologia, desde a gestão à análise, dando o controlo ao utilizador para que seja ele o elemento que decide como, quando e de que forma pretende analisar os dados. É por esta razão que este sistema utiliza várias tecnologias, desde a tecnologia para gerir os dados, *MySQL*, passando pela tecnologia base de análise de dados genealógicos, *GEDCOM* e permitindo novos formatos de forma a garantir o presente e o futuro como o *XML*, *KML* ou a junção das representações *GEDCOM* e *XML*.

Assim, seja como ciência autónoma ou como método de estudo, a Genealogia define-se nos dias de hoje como um dos saberes de maior potencial para o estudo de toda a sociedade, nas diferentes vertentes, podendo cooperar no sentido favorável com a descoberta do núcleo da sociedade em estudo. É assim fundamental, construir aplicações capazes de ajudar os curiosos, os profissionais, ou qualquer interessado nesta tão importante área, isto porque, o que se ganha por elaborar uma história familiar, não é só descobrir de onde viemos, mas essencialmente descobrir quem somos (Correia, 2006).

1.1 OBJECTIVOS DO ESTUDO

Face ao que foi supramencionado, o presente estudo define como objectivo geral, conceber um sistema informático capaz de efectuar por um lado uma gestão eficiente de dados genealógicos, e por outro, permitir aos utilizadores, a análise desses mesmos dados.

No entanto, este estudo apresenta como objectivos específicos:

- Caracterizar a informação genealógica e qual a sua essência.
- Planear e desenvolver uma base de dados prosopográfica.

- Identificar exemplos de soluções já existentes que efectuem a gestão e/ou a análise de dados genealógicos.
- Reflectir sobre a utilização e utilidade de uma ferramenta informática ao serviço da gestão e análise de genealogias.
- Garantir a expansão no acondicionamento de dados e respectiva partilha utilizando soluções e recursos que possam ser expandidos num futuro a curto, médio e longo prazo.
- Permitir a interoperabilidade entre aplicações de gestão de dados, e aplicações de análise de dados, independentes do seu tipo e forma de operar, através de formatos “universais” entendidos pelas aplicações.

1.2 RELEVÂNCIA DO ESTUDO

O presente estudo assume a sua importância em várias vertentes das quais a mais significativa será a apresentação de uma metodologia capaz de preencher lacunas dentro desta área de estudo, sobretudo ao nível da ligação dos Sistemas de Gestão aos Sistemas de Análises de dados genealógicos.

Tal como referenciado anteriormente, na definição dos Objectivos do Estudo, acredita-se que este estudo abra novas portas e horizontes no estudo das relações, isto é, confio que com o resultado deste estudo será possível aplicar um Sistema de Gestão e Análise a outras áreas para além da Genealogia com as devidas alterações a cada área em específico. Este facto constitui assim uma relevância muito importante para este estudo já que não se pretende construir uma ferramenta para um caso específico, mas sim que possa ser utilizada quer para outros projectos semelhantes, quer para novas áreas como é meu desejo.

Falando em projectos semelhantes, durante a realização do estudo surgiram casos concretos para adoptar este Sistema Informático de Gestão e Análise de Genealogias, tal como o Projecto *PTDC/HAH/66107/2006 - Pequena Nobreza e "Nobreza da Terra" na Construção do Império - Os Arquipélagos Atlânticos*, que o objectivo dos membros deste projecto era por um lado ter um sistema de gestão para os dados genealógicos melhor que

o suporte papel, e por outro lado, com essa funcionalidade (a gestão dos dados em formato electrónico) proporcionar novos mundos de análise dos dados quer ao nível da família, quer ao nível de dos dados genealógicos georreferenciados (FCT - Fundação para a Ciência e Tecnologia, 2006), provando assim que este estudo constituiu uma mais-valia quer para o projecto em concreto para o qual foi concebido, quer para projectos semelhantes ou até mesmo de outras áreas.

Mantenho também uma expectativa em relação a este estudo já que acredito que este trabalho desenvolvido poderá potenciar novos trabalhos nesta área, acreditando assim que este trabalho possa funcionar como um elemento potenciador para o aparecimento de trabalhos do mesmo género mas aplicados a outros conteúdos.

Apesar de o estudo ter sido feito no âmbito de um determinado projecto, o resultado final não se restringe apenas ao projecto, isto é, não existe assim um público-alvo (Jacobus, 1968) definido para a utilização do Sistema Informático de Gestão e Análise de Genealogias, pois a utilização pode ser feita quer por um simples curioso na área da Genealogia para registar a história da sua família, ou pode vir mesmo a ser utilizado como aplicação educacional já que permitirá alargar novos horizontes e ultrapassar barreiras que até aqui não eram ultrapassadas ou podiam ser mas com relativa dificuldade (Jacobus, 1968).

A integração de tecnologias recentes como é o caso da representação XML constitui igualmente um ponto forte e uma relevância extrema já que a adaptação ao mundo da Internet fica a um simples passo, proporcionando assim uma divulgação acessível e de forma gratuita de todos os recursos criados. A possibilidade de análise dos dados genealógicos sem ter limites espaço-temporais, sem depender do sistema de gestão de base de dados e independente da plataforma do sistema operativo é ao nível da interoperabilidade o cume de uma longa escalada na montanha da integração de aplicações, já que posso realizar a análise dos dados genealógicos utilizando por exemplo um PDA, ou um computador com um Sistema Operativo Windows® ou um computador com o Sistema Mac OS X®.

Desta forma, conclui-se que quando se liga o universo da Genealogia ao universo dos Sistemas de Informação, contribui-se para o avançar da Genealogia naquela que é a sua essência, o estudo das famílias, promovendo a troca de informação e partilha de dados genealógicos. No entanto, sabendo da existência das aplicações informáticas já existentes na área da Genealogia, procurar realizar uma integração que retire o melhor de cada parte para juntar num todo, é de facto o grande objectivo que alicerça este estudo, podendo permitir assim, alargar novos horizontes quer no universo da Genealogia quer mesmo a outras áreas diferentes.

1.3 ESTRUTURA DA DISSERTAÇÃO

A dissertação está dividida em seis capítulos.

O primeiro, referente à Introdução, contextualiza a temática do presente trabalho, apresenta o problema estudado, os objectivos a atingir, assim como a relevância que este estudo apresenta, apontando ainda a estrutura da dissertação, actual ponto de leitura.

No capítulo seguinte, o segundo, abordam-se as questões relacionadas com a Gestão e a Análise de Genealogias. É fundamentada a necessidade de gestão e análise de genealogias, assim como a definição de conceitos importantes para uma total compreensão desta problemática, e do objecto de estudo desta dissertação. Está ainda presente neste capítulo, um estudo sobre o estado da arte, ponto este que se destina a exemplificar ou a expor o que é actualmente feito no campo em estudo, a que se acrescenta uma história compacta acompanhando a evolução das soluções tecnológicas existentes.

Por seu lado, o terceiro capítulo é o trabalho desenvolvido com base neste estudo. Assim, é neste ponto que é caracterizado todo o sistema informático desenvolvido de forma a permitir a gestão e a respectiva análise de dados genealógicos, passando por todas as fases que arquitectam este sistema informático.

O Caso de Estudo é apresentado no quarto capítulo ao procedendo-se aqui, à sua descrição e mostrando as etapas necessárias à construção do sistema informático aplicado ao caso

concreto de um projecto financiado pela Fundação para a Ciência e a Tecnologia, denominado por FCOMP-01-0124-FEDER-007360.

O quinto capítulo é totalmente dedicado à Apresentação e Discussão de Resultados. Desta forma, começa-se por identificar quais os resultados gerais obtidos da aplicação deste sistema ao projecto supracitado. Seguidamente é efectuada a análise e a caracterização das genealogias obtidas, sendo este ponto dedicado à análise dos resultados gerais obtidos. No último ponto deste capítulo é mostrada visualmente a forma como se aplica este sistema informático à Genealogia do Santo Ofício e das Ordens Militares (1570 - 1773) exemplificando formas de gestão e de análise utilizando diferentes representações.

Fechando, no capítulo da Conclusão, o sexto, apresentam-se as conclusões deste estudo, apontando-se ainda, perspectivas para trabalhos futuros.

CAPÍTULO II - GESTÃO E ANÁLISE DE GENEALOGIAS

2.1 FUNDAMENTAÇÃO E APLICAÇÃO

Desde os tempos mais remotos que existe a ideia de homenagear os antepassados e a prova está que ao longo dos tempos, as pessoas tem honrado os seus antepassados de muitas formas diferentes. Recordá-los e preservar a sua história de vida é apenas uma das muitas formas (Doane, 1992).

A importância dos nossos antepassados é de facto muito grande já que bastaria a minha Heptavó (sétima geração) não conhecesse o meu Heptavô e toda a história se alterava questionando até a minha existência caso isso se confirmasse (Hearn, 2006).

Mais do que honrar a história da família e dos antepassados, proporcionar a gestão e análise de genealogias pode ser simplesmente fascinante porque permite não só analisar a informação individual de cada pessoa, isto é, verificar a mudança espaço-temporal que essa pessoa teve ao longo da vida, mas também olhar para os seus antepassados e analisar de diferentes formas toda a informação proveniente dos ancestrais (Jacobus, 1968).

A história e a própria genealogia não podem ficar restrita a livros didáticos. É necessário inovar para que o utilizar participe nas reconstruções históricas, partilhe o gosto pelos antepassados ou simplesmente que conheça algo que não conhece e que tem curiosidade (Doane, 1992).

O contexto histórico é bastante importante já que se o entendermos, iremos ter uma melhor percepção no tipo de registos encontrados, e até mesmo o porque de alguns estarem em falta. Além disso, quanto mais se conhece sobre o contexto histórico, mais hipóteses temos de encontrar informações sobre o que estamos a analisar (Doane, 1992; Wright, 1995).

Aprender e conhecer a genealogia da família também pode ajudar na aprendizagem sobre as condições médicas que podem ser transmitidas geneticamente e com isso compilar um historial médico que possa ajudar e prevenir doenças hereditárias. Será que os meus antepassados tinham problemas de coração? Será que eram propensos a cancro? Tinham incidência de doenças auto-imunes? Conhecer bem a história dos antepassados pode ajudar a tomar uma atitude face a problemas de saúde (Hearn, 2006).

Muitas religiões espalhadas por todo o mundo também apresentam este carácter educativo, ensinando a apreciação e mesura para com os seus antepassados. Um exemplo é os membros da Igreja de Jesus Cristo dos Santos dos Últimos Dias, também conhecidos de mórmones, que desenham e constroem as árvores genealogias como uma convicção religiosa (A Igreja de Jesus Cristo dos Santos dos Últimos Dias, 2009).

Poderão ser várias as fontes da informação colecionada desde registos antigos dos seus antepassados, bibliotecas locais ou até mesmo o Arquivo Nacional da Torre do Tombo (ANTT) (Portal Oficial ANTT, s/d). A internet, obviamente, é também um local onde se pode encontrar e divulgar dados genealógicos, sendo assumido por muitos que na Internet podemos encontrar um tesouro de informações, quer através do acesso a materiais de pesquisa, quer através de fóruns de encontros de parentes passando por outras possibilidades que a Internet nos possibilita (Hearn, 2006).

É evidente que são apenas hipóteses, fundamentos e aplicações possíveis para gerir e analisar genealogias sendo que cada caso é específico com todas as suas particularidades e objectivos e o que se pretende é conseguir responder de forma correcta e coerente a todas as possibilidades e vertentes dentro da mesma área da Gestão e Análise de Genealogias.

2.2 DEFINIÇÕES E CONCEITOS

Não há dúvidas que a Informática veio dar outra dimensão à Genealogia. Se no passado o manuseamento de grandes arquivos de dados desmobilizava os investigadores, a actual rapidez e flexibilidade de utilização dos arquivos transformados em bases de dados, despertam novas hipóteses, novos caminhos, que anteriormente não seriam sequer idealizados (Dupâquier, 2009).

Assim é importante perceber algumas definições e conceitos importantes relacionados com a Gestão e Análise de Genealogias que passarei a explicar.

2.2.1 METADADOS

Os metadados são dados que descrevem completamente os dados (bases) que representam, permitindo ao utilizador decidir sobre a utilização desses dados da melhor forma possível. Assim, os metadados podem ser definidos como dados que permitem informar as pessoas sobre a existência de um conjunto de dados ligados às suas necessidades específicas (Almeida, 1999).

A essência da palavra metadados foi criada por *Jack Myres* no final da década de 1960, precisamente para denominar os dados que descreviam registos de ficheiros tradicionais (Howe, 1996).

Desta forma, este conjunto de dados que identificam os dados de um determinado documento e que podem fornecer informação sobre a forma como os dados estão descritos, administrados, como serão utilizados, parametrizados entre muitas outras funcionalidades, são chamados de metadados (Taylor, 1999) e actuam como facilitadores na organização, gestão e análise da informação (Dempsey & Heery, 1998).

As aplicações destes “dados sobre os dados” podem ser diversas, podendo estes fazerem parte de processos de procura de informação, organização da informação, entre outras vertentes. O comércio electrónico é uma área em que os metadados são aplicados como facilitadores já que muitas vezes são os metadados que viabilizam o próprio negócio servindo assim como uma possível plataforma de mudança para a publicação em formatos electrónicos (Day, 1998).

Sempre com o objectivo presente de documentar e organizar de forma estruturada os dados independentemente do seu tipo, os metadados apresentam igualmente como objectivo, minimizar o esforço e facilitar a própria manutenção dos dados (Almeida, 1999).

Assim, os metadados podem ser divididos em pelo menos duas categorias: Descritivos ou Estruturais (Dempsey & Heery, 1998). A componente descritiva é quando os metadados são implementados com o intuito da documentação sobre os próprios dados, isto é, quando os metadados descrevem e identificam recursos de informação. A sua implementação num Sistema de Gestão e Análise de Genealogias torna-se assim muito importante nesta área já que permite uma informação detalhada sobre os próprios dados, incluindo toda a parametrização da aplicação para que suporte a transição da área de gestão para uma outra e não ser necessário reprogramar todo o código de programação (Miller, Vandome & McBrewster, 2009). Ao definir-se os metadados, assume-se que toda a aplicação irá moldar-se aquilo que foi definido desde as tabelas até aos campos que constituem toda a área de gestão.

Na vertente estrutural, estamos perante uma abstracção maior na definição de metadados, isto porque, através desta vertente, todo o processo de navegação e apresentação de recursos electrónicos ficam mais facilitados com a introdução de informação sobre os próprios dados. Uma possível forma de definir metadados com este intuito é através da representação XML, de que este trabalho apresenta um exemplo no capítulo II, que encaixa-se assim como um elemento que sirva para estruturar e armazenar os metadados (Sousa, 2002).

Desta forma, os metadados procuram cumprir a função de informar sobre o documento digital em causa, promovendo assim a sua gestão e análise. Sem o recurso aos metadados, a interoperabilidade necessária para um avanço nos recursos de informação da Internet pode ficar comprometida. Assim, a sua existência num Sistema Informático de Gestão e Análise de Genealogias é desta forma vital já que não só permite uma informação adicional sobre os próprios dados, mas também uma portabilidade entre curiosos ou genealógicos, que procuram ter e partilhar o máximo de dados genealógicos possíveis.

2.2.2 GEDCOM

A representação GEDCOM, abreviatura para comunicação de dados genealógicos (em português), foi desenvolvida pelo departamento da História de Famílias da Igreja de Jesus Cristo dos Santos dos Últimos Dias com o objectivo primordial de providenciar um formato flexível, uniforme e eficaz para a troca de informação genealógica computadorizada (Family History Department - The Church of Jesus Christ of Latter-day Saints, 1996). A finalidade é fomentar o intercâmbio de dados genealógicos para que, aliado ao desenvolvimento de programas de análise de genealogias, ajudar genealogistas, historiadores, investigadores ou até mesmo curiosos sobre a Genealogia (About.com: Genealogy, s/d).

A representação GEDCOM pode ser vista como uma base de dados na forma de um conjunto sequencial de dados com registos relacionados entre si (Eastman, 2004a). Um registo é representado por uma sequência de três elementos. O primeiro elemento, trata-se de um elemento numérico representando o nível hierárquico. O segundo elemento, consiste na etiqueta que caracteriza esse registo. Por último, o terceiro elemento, nem sempre presente, consiste no texto caracterizador da etiqueta. No entanto, cada linha pode conter uma referência a um identificador ou apontador, utilizado por exemplo, para a definição das Genealogias (Family History Department - The Church of Jesus Christ of Latter-day Saints, 1996).

A etiqueta na representação de uma linha do ficheiro GEDCOM, considerando o contexto hierárquico, identifica as informações contidas na linha, no mesmo sentido em que os nomes dos campos são definidos nos Sistemas de Gestão de Bases de Dados (Eastman,

2004a). Isto significa que os dados são auto-definidos. A capacidade de uso de diferentes ou novas etiquetas no GEDCOM é algo que deve ser prestada peculiar atenção, uma vez que, esta introdução de elementos desconhecidos não causa incompatibilidade, isto porque, o sistema que analisa esta informação, ignora todas as etiquetas que desconhece, processando apenas a informação que compreende, não provocando mau funcionamento pela introdução de etiquetas desconhecidas (Family History Department - The Church of Jesus Christ of Latter-day Saints, 1996).

A introdução de etiquetas desconhecidas no ficheiro GEDCOM prende-se com o facto de a representação GEDCOM não ser o ponto de chegada na exportação e na análise de informação. Significa assim que são colocados elementos no ficheiro GEDCOM que um programa de leitura desta representação não irá conhecer, mas que na minha opinião, pode constituir uma vantagem enorme para futuros sistemas de análise. O caso utilizado neste Sistema Informático de Sistema de Gestão e Análise de Genealogias foi apenas a introdução de duas etiquetas desconhecidas na representação GEDCOM, caracterizadoras da Latitude (LATI) e Longitude (LONG) de um determinado local.

A introdução destas duas etiquetas desconhecidas apresenta-se como algo que pode permitir a exploração e a análise de dados numa perspectiva diferente, isto é, interligando os dados com Sistemas de Informação Geográficos. Sabendo que os Sistemas de Informação Geográficos, como por exemplo o *Google Earth*® necessitam apenas das coordenadas latitude e longitude, juntamente com a informação a apresentar, para colocar a informação desejada nos seus mapas (Crowder, 2007), foi com essa ideia que foram introduzidas estas duas etiquetas para referenciação da latitude e longitude de um local. Ambas as etiquetas seguem a mesma regra das restantes na representação GEDCOM surgindo sempre num nível hierárquico interno à definição de um local, isto é, quando é definido a etiqueta “2 PLAC Évora”, definindo que o local é Évora, abaixo seguem as duas instruções: “3 LATI 38.568533”; “3 LONG -7.908268” definido para a localização de Évora, a Latitude de 38.568533 e a Longitude de -7.908268. A introdução do nível 3 significa que as instruções são relativas ao local de Évora, nível hierárquico 2.

Tal como já referi, as relações hierárquicas são indicadas por um número que define o nível. As linhas subordinadas apresentam um número de nível mais elevado (About.com: Genealogy, s/d). A hierarquia dos níveis permite que uma linha tenha linhas subordinadas e que essas mesmas linhas subordinadas contenham outras subordinadas e assim sucessivamente. Isto permite que um conjunto de linhas e linhas subordinadas ostentem um conjunto de informação directamente ligada a um só objecto (Eastman, 2004a; Miller, Vandome & McBrewster, 2009). Desta forma, um registo é definido como um conjunto de uma ou mais linhas, sendo o início de um novo registo definido pela linha cujo número relativo ao nível corresponde ao número zero (0) (Family History Department - The Church of Jesus Christ of Latter-day Saints, 1996).

Para além das relações hierárquicas, a representação GEDCOM define igualmente relações entre registos, permitindo que um registo possa ser relacionado com outros registos sem a introdução de redundância da informação (Family History Department - The Church of Jesus Christ of Latter-day Saints, 1996). Este tipo de relações inter-registos só é possível através da introdução de uma referência, isto é, através da introdução do elemento de identificação do registo que é feito de forma universal no GEDCOM à semelhança de uma chave primária na terminologia utilizada nos Sistemas de Gestão de Bases de Dados relacionais (Eastman, 2004a; Miller, Vandome & McBrewster, 2009).

Sabendo que cada dialecto apresenta por vezes caracteres especiais, a representação GEDCOM necessita assim de acondicionar diferentes tipos de caracteres de modo a facilitar a partilha dos dados genealógicos em linguagens distintas. Com vista a minimizar o número de diferentes standards, na representação GEDCOM cada sistema converte o seu funcionamento para ANSEL, e eventualmente para UNICODE (Family History Department - The Church of Jesus Christ of Latter-day Saints, 1996; Miller, Vandome & McBrewster, 2009).

2.2.2.1 A Gramática

Uma gramática é um conjunto de regras que especificam os conjuntos de caracteres que são válidos para a criação de expressões em cada linha da representação GEDCOM (Family

History Department - The Church of Jesus Christ of Latter-day Saints, 1996). O conjunto de caracteres pode ser definido como várias combinações de elementos, sejam eles, elementos variáveis e/ou constantes (Miller, Vandome & McBrewster, 2009).

Na definição da gramática os elementos podem ser definidos com base noutros elementos ou recorrendo a elementos alternativos. Cada elemento na definição é separado pelo sinal de adição (+) significando assim que ambos os elementos são necessários. Caso exista a escolha de diferentes elementos que possam ser usados, então os elementos surgem numa lista entre parênteses rectos ([]) em que o elemento separador é a barra vertical (|). A leitura da gramática é feita substituindo todos os sub-elementos, isto é, substituindo todos os elementos que podem ser substituídos por elementos atómicos, até todos os sub-elementos estarem na forma atómica, isto é, até encontrarem um elemento que não tenha alternativa (Family History Department - The Church of Jesus Christ of Latter-day Saints, 1996).

Um ficheiro GEDCOM é representado por um conjunto de linhas que é caracterizada por um conjunto de caracteres. Cada linha da representação GEDCOM segue o seguinte princípio: Nível Etiqueta Texto; Ou seja, cada linha irá conter sempre um nível associado, uma etiqueta que irá retratar qual o assunto ou tema que estamos a abordar nessa linha e o texto caracterizador dessa etiqueta (Miller, Vandome & McBrewster, 2009).

Assim, existem um conjunto de regras que tem de ser respeitadas para que a representação esteja correcta. Uma das regras básicas da representação GEDCOM é que o início de um novo registo (entenda-se registo como a criação de um novo individuo, família, entidade, etc.) é sempre feito atribuindo ao nível de cada linha o valor 0 (zero) (Miller, Vandome & McBrewster, 2009). Ainda em relação aos níveis, olhando para duas linhas, o nível da linha seguinte, não pode ser superior ao valor do nível da linha anterior juntando uma unidade (Eastman, 2004a). Esta regra permite assim que um registo tenha a informação caracterizadora desse registo, identificada através de níveis cujo os valores irão ser seguidos, isto é, no nível 0 irá estar a informação da criação, no nível 1 irá surgir a identificação base do registo, no nível 2 informação complementar e assim sucessivamente (About.com: Genealogy, s/d).

Aquando da criação de um indivíduo, por exemplo, a estrutura da linha mantêm-se, isto é, a linha é separada em três partes, nível da linha, que neste caso será o 0 (zero), a identificação do indivíduo, que é caracterizada pela utilização da seguinte estrutura, @E1@, em que E1 é o identificador único deste indivíduo em questão, seguido da colocação de uma “etiqueta” INDI, que indica a criação de um indivíduo (Family History Department - The Church of Jesus Christ of Latter-day Saints, 1996). Assim define-se uma linha da representação GEDCOM pela seguinte forma:

linha_gedcom:= nível + espaço_em_branco + [referencia_id + espaço_em_branco +] etiqueta + [espaço_em_branco + valor_linha +] enter

ou

linha_gedcom:= nível + espaço_em_branco + opcional_referencia_id + etiqueta + espaço_em_branco + opcional_valor_linha + enter

Demonstrando uma aplicação prática de uma possível linha da representação GEDCOM:

1 NAME João Leitão

Em seguida apresentarei em detalhe todos os elementos da gramática, sendo algumas constantes definidas na forma hexadecimal que corresponderá a um determinado valor ASCII hexadecimal.

Quando surgir na definição da gramática constantes separadas pelo sinal de reticências (...) indica que se trata de um intervalo de valores possíveis sendo o primeiro elemento à esquerda do sinal e o ultimo elemento à direita do sinal.

Desta forma, segue então a descrição detalhada da gramática da representação GEDCOM (Family History Department - The Church of Jesus Christ of Latter-day Saints, 1996; Miller, Vandome & McBrewster, 2009):

alfa:=[A...Z | a...z | _]

A colocação do *underscore* (`_`) serve para ligar duas palavras sem um espaço em branco entre elas de forma a criar nomes de etiquetas ou rótulos.

alfanum:= [alfa | dígito]

continua_ate:= [alfa | dígito | outro | # | espaço_em_branco]

dígito:= [0...9]

espaço_em_branco:= [(0x20)] em que: (0x20) é o carácter que identifica um espaço em branco

etiqueta:= [alfanum | etiqueta + alfanum]

item_linha:= [qualquer_char | regra | item_linha + qualquer_char | item_linha + regra]

nível:= [dígito | nível + dígito]

null:= vazio

opcional_referencia_id:= referencia_id + espaço_em_branco

opcional_valor_linha:= valor_linha + espaço_em_branco

outro:= [! | " | \$ | % | & | ' | (|) | * | + | , | - | . | / | : | ; | < | = | > | ? | [| \ |] | ^ | ` | { | | } | ~ | @ | # | _ | espaço_em_branco]

ponteiro:= [@ + alfanum + ponteiro_string + @]

ponteiro_char:=[continua_ate]

ponteiro_string:= [null | ponteiro_char | ponteiro_string + ponteiro_char]

qualquer_char:= [alfa | dígito | outro |espaço_em_branco| @ | #]

referencia_id:= [ponteiro]

regra:=[@+ # + texto_regra + @ + #]

Este elemento trata-se de uma sequência de caracteres na gramática utilizado para especificar processamento especial, isto é, para a transformação de elementos que não pertençam à representação GEDCOM.

texto_regra:=[qualquer_char | texto_regra + qualquer_char]

É definido para satisfazer as necessidades exigidas no elemento “regra” para casos particulares de transformação de elementos fora da representação GEDCOM e que se deseja manter a “formatação” original.

valor_linha:= [ponteiro | item_linha]

O valor da linha delimita o domínio possível e permitido no contexto da etiqueta. A combinação da etiqueta, com o valor da linha e o contexto hierárquico das linhas GEDCOM providenciam uma melhor percepção dos valores em causa.

2.2.3 XML

O XML, eXtensible Markup Language, foi criado pela World Wide Web Consortium e surgiu da necessidade de gerar linguagens de marcação (W3C, 2003). Derivado do formato SGML (Standard Generalized Markup Language), o XML foi desenhado para lidar com os desafios de publicação de conteúdos electrónicos a larga escala, tendo nos dias de hoje, um papel fundamental na troca de informação pela Web (Flynn, 2010). A ideia inicial era a criação

de uma versão simplificada da linguagem SGML, mas voltada para aplicações Web, mantendo as vantagens ao nível da flexibilidade, estrutura e validação, mas com uma aprendizagem, uso e implementação mais facilitada (Walsh, 1998).

E é essa característica da linguagem XML, a flexibilidade, que lhe permite a criação de marcadores para um contexto específico, isto é, possibilita e permite que um profissional desenvolva para si ou para a sua Entidade, a sua própria “linguagem de marcadores”, isto porque uma das características da linguagem XML é o facto de os elementos serem auto-descritivos, isto mesmo tendo em conta que um ficheiro XML é essencialmente um ficheiro de texto. Esta característica, os elementos auto-descritivos, permite assim, em comparação com outros formatos de preservação de dados, uma melhor leitura pelo Homem, sendo uma das razões para a sua popularidade ter crescido nos últimos anos (W3C, 2003).

Desta forma, a principal característica da linguagem XML é a de criar uma infra-estrutura única para diversas linguagens. Baseada em vários princípios importantes, a representação XML é um formato que não depende das plataformas de hardware ou de software e que cria documentos a partir de dados organizados de forma hierárquica (Loureiro, 2008).

Tal como o HTML, o XML permite a definição, transmissão, validação e interpretação de informação entre diferentes plataformas de computação e aplicações. A interoperabilidade é assim então aprimorada pois permite-nos ter diferentes áreas de competências a desenvolver esquemas XML que definam a linguagem de marcação para a troca de informação especializada tendo em conta os critérios específicos de cada um (Sousa, 2002).

Assim, o XML combina a flexibilidade da SGML com a simplicidade HTML, permitindo que um criador estenda a linguagem dividindo novas etiquetas de forma a descrever e partilhar a informação em qualquer contexto especializado e desejado, com a única condição que as novas etiquetas sigam as regras de sintaxe XML definidas pelas especificações XML da W3C.

O XML não é uma linguagem de programação como o JAVA ou C, por exemplo, mas sim uma linguagem de marcação. Qualquer browser, ou outra aplicação específica para o efeito, pode ler um documento XML com o propósito de fazer alguma coisa com ele (Walsh, 1998).

Os elementos da informação na representação XML são definidos utilizando etiquetas representativas dos seus próprios elementos (W3C, 2003), isto é, numa abordagem muito minimalista e num contexto de explicação simples, uma pessoa tem 2 nomes (primeiro nome e ultimo nome). Uma possível representação XML para o efeito seria:

```
<Nome>  
    <Primeiro Nome> João </Primeiro Nome>  
    <Último Nome> Leitão </ Último Nome>  
</Nome>
```

A criação de um esquema é assim importante já que contem os nomes das etiquetas, o formato dos dados, e a forma como se podem relacionar. A este esquema dá-se o nome de Document Type Definition (DTD) (Refsnes, 1999a).

Um esquema é assim um simples ficheiro ou uma colecção de ficheiros com o objectivo de servir como base para a definição de dados, formato e respectiva estrutura de um ficheiro XML. Um esquema correctamente definido permite que seja possível validar ficheiros com representações XML, verificando se as regras impostas são satisfeitas. O esquema XML descreve assim o conteúdo e a estrutura que terá que ter um ficheiro XML, enquanto o ficheiro XML irá conter a informação propriamente dita (Refsnes, 1999a).

No entanto é ainda possível criar um documento que contenha a especificação com sintaxe e semântica para transformar um ficheiro XML noutro ficheiro XML mas com identificações diferentes. Esta definição encontra-se nos ficheiros XSLT, designada como parte de uma folha de estilos para as representações XML (W3C, 1999). A capacidade destes ficheiros XSLT serem totalmente independentes, assim como os anteriores DTD,

constitui uma mais-valia já que não é necessário estar tudo definido no mesmo ficheiro XML (Refsnes, 1999b).

Desta forma, espera-se que o XML ocupe no futuro, um papel cada vez mais aplicado e utilizado em diversas vertentes, pela capacidade de incluir comportamentos dinâmicos a um texto e de ligar diversos conteúdos entre si.

2.2.4 GOOGLE EARTH E KML

O Google Earth® e trata-se de uma aplicação de mapeamento interactivo que permite aos utilizadores navegarem por todo o globo, analisando imagens satélites com a descrição de estradas, edifícios, características geográficas e inúmeras especificações detalhadas de determinados locais (EDUCASE Learning Initiative, 2006). A possibilidade do utilizador introduzir os seus próprios pontos de interesse aliado à sua partilha (Gibin, Singleton, Milton, Mateos & Longley, 2008), é apenas uma das razões que ajudou na decisão por este software.

O Google Earth conecta-se através da Internet carregando desta forma conteúdos online quando conectados com determinados lugares. Para alguns locais, a aplicação cria representações a três dimensões (3D) (EDUCASE Learning Initiative, 2006). No entanto este tipo de representação topográfica não está disponível para grande parte do território português, com excepção de pontos como Lisboa, Porto, entre outros. Os utilizadores podem assim escolher as *layers*² que desejam mostrar ou esconder, podendo combina-las através de qualquer combinação (Google Earth Official Webpage, s/d). Por exemplo, se o objectivo for estudar o rasto de um determinado objecto, provavelmente as *layers* que interessam ao utilizador serão as estradas, e não os centros comerciais entre outros elementos que não estão ligados à sua procura.

² Nível de Profundidade.

Considerada como uma das aplicações favoritas para pessoas que necessitem de ver ou mostrar localizações facilmente, são várias as aplicações deste sistema de informação geográfico, como o caso da sua utilização nas escolas com o objectivo de ajudar na compreensão de temas desde as Ciências até às Artes (EDUCASE Learning Initiative, 2006). A área da Geologia é uma das áreas que utiliza este tipo de ferramenta já que permite aos seus alunos visualizarem virtualmente as áreas a estudar mostrando toda a topografia e áreas circundantes de forma bastante rápida permitindo aproximações e afastamentos através da funcionalidade *zoom* (Google Earth Official Webpage, s/d). Em suma, não existe uma utilização predefinida para esta ferramenta já que a mesma se adapta às mais variadas exigências podendo servir como simples elemento de lazer, ou indo até ao extremo e servir como uma forte e potente ferramenta de análise de dados.

O Google Earth pode ser descarregado da internet e instalado de forma gratuita. No entanto, caso se pretenda, é possível adquirir uma versão profissional paga. No nosso caso, utilizou-se a versão gratuita já que cumpria e satisfazia todas as necessidades exigidas. O Google Earth permite viajar até qualquer localização do mundo quer seja pela introdução de moradas, de nomes, de pontos de interesse ou de coordenadas geográficas como a latitude e longitude (EDUCASE Learning Initiative, 2006). Foi através deste princípio que se pensou ser possível efectuar a ligação entre o Sistema de Gestão de Bases de Dados e o Google Earth pois tal como vimos anteriormente, para cada local, registou-se as coordenadas latitude e longitude.

No entanto, os utilizadores podem ainda procurar uma região específica através de palavras-chave. À medida que a ferramenta *zoom* aumenta para a área em questão, maior é o detalhe aumentando a resolução (Google Earth Official Webpage, s/d). Algumas localizações do mapa da ferramenta Google Earth são pontos de interacção com o utilizador, isto é, caso o utilizador clique com o botão esquerdo do rato em alguns pontos, pode surgir uma janela com informação sobre esse local e eventuais ligações externas para mais informações (GoogleCode, 2007). Os elementos multimédia, como imagens, também estão fortemente presentes, já que também é possível, quando o utilizador clica num local, surgir uma janela com fotografias sobre esse local (EDUCASE Learning Initiative, 2006).

Tal como foi dito anteriormente, a capacidade de personalização da entrada dos dados foi sem dúvida o ponto que marcou a diferença e fez apostar e seguir esta via de conhecimento. Apresentar um relatório em formato papel com tabelas ou com informação puramente textual, ou representar a mesma pelo mapa carregando consigo toda a interactividade da aplicação Google Earth, o utilizador prefere geralmente o elemento interactivo, desde que ambas as abordagens respondam às mesmas perguntas (EDUCASE Learning Initiative, 2006).

Mais do que um mapa, o Google Earth permite aos utilizadores criarem e partilharem as suas investigações pessoais, isto já que a informação que é possível carregar para o Google Earth pode ser escolhida pelo utilizador sem qualquer limitação ao nível do conteúdo (GoogleCode, 2007).

Todavia, é importante também perceber as limitações desta ferramenta. O facto de o Google Earth mostrar imagens baseadas em imagens capturadas por satélite, limita o utilizador já que aquilo que este vê é uma fotografia num determinado momento do tempo. Caso um edifício seja construído recentemente, essas alterações não terão impacto nem serão reflectidas na aplicação (EDUCASE Learning Initiative, 2006).

Outro ponto considerado como limitação é o facto de nem todos os locais apresentarem a mesma resolução, isto é, para algumas cidades podemos efectuar um zoom tão elevado que se consegue visualizar as pessoas a andarem nos passeios, ao invés de outros locais que a resolução máxima permitida pode não ser suficiente para mostrar de forma eficiente alguns edifícios (eHow, s/d).

Por se tratar de uma aplicação exigente, os requisitos exigidos para que esta aplicação funcione de forma correcta são, possuir pelo menos 1Gbyte de memória RAM, uma boa placa gráfica e uma conexão à internet de banda larga de forma a poder tirar o mínimo de rendimento desta aplicação (EDUCASE Learning Initiative, 2006).

Não só por aquilo que esta ferramenta representa actualmente, mas também por tudo aquilo que pode proporcionar no futuro, esta aplicação integra a lista de aplicações a

utilizar para a análise de dados genealógicos no Caso de Estudo, sendo que é expectável que no futuro, esta aplicação torne-se mais sofisticada, com ferramentas adicionais e aumentando a cobertura de imagens em alta resolução.

Assim, o Google Earth junta-se à lista de aplicações que podem transportar as ideias dos livros para a imaginação dos utilizadores, mostrando ser uma ferramenta que pode ajudar a desenvolver o contexto espacial e cultural por todo o mundo.

O KML por seu lado, é uma gramática XML usada para codificar e transportar representações de informação geográfica para sistemas de informação geográficos (Wilson, 2008). Um ficheiro KML é processado da mesma forma que um ficheiro em HTML (ou XML) é processado pelos browsers (Crowder, 2007).

Tal como o HTML, KML é uma estrutura baseada em etiquetas, com nomes e atributos utilizados para um propósito específico que se trata da visualização de dados (Crowder, 2007). Os sistemas de informação geográficas, por exemplo o Google Earth, apresentam o mesmo papel que os browsers para a linguagem HTML. Assim, os *geobrowsers*, são os sistemas que interpretam a linguagem KML como por exemplo o Google Earth, Google Maps, entre outros (Butler, 2006).

A utilização do KML pode ser aplicada a diferentes razões, no entanto está sempre associada à caracterização de determinados locais geográficos. Especificar ícones e rótulos para identificar localizações no globo, ou criar diferentes tipos de posições da câmara são apenas alguns exemplos do que é possível fazer através desta representação (GoogleCode, 2007).

Originalmente esta representação foi criada como um formato de ficheiros para o programa Keyhole's Earth Viewer, programa este que mais tarde iria emergir como a aplicação Google Earth (Ratliff, 2007). Permitia assim, que os utilizadores carregassem o

seu próprio conteúdo mostrando essa informação nos próprios mapas baseados em imagens satélites.

Assim, o KML é centralizado na visualização de informação geográfica em todo o globo. Esta linguagem contém igualmente controlos para a navegação do utilizador, agregando assim para onde ir e para onde olhar (Wilson, 2008).

Olhando para a estrutura do KML, tal como todas as gramáticas XML, o KML é uma estrutura baseada em etiquetas com elementos e atributos hierárquicos. Cada bloco é chamado de elemento, enquanto a etiqueta (*tag*) representa a forma como o elemento é representado em código KML (Crowder, 2007).

Assim, mostra-se um exemplo concreto de um ficheiro KML, mostrando a localização exacta do Departamento de Informática da Universidade de Évora.

```
<?xml version="1.0" encoding="UTF-8"?>
<kml xmlns="http://earth.google.com/kml/2.2">
  <Placemark>
    <name> Departamento de Informática - UÉvora </name>
    <description> O Departamento de Informática é uma unidade
científico-pedagógica da Universidade de Évora que tem como missão a
realização de tarefas de investigação e ensino, de extensão
universitária e de prestação de serviços especializados à comunidade
nos domínios da Informática.
    </description>
    <Point>
      <coordinates> -7.910553, 38.568166
      </coordinates>
    </Point>
  </Placemark>
</kml>
```

Neste exemplo da Figura 1 temos que o Departamento de Informática da Universidade de Évora localizado nas coordenadas Longitude: -7.910553 e Latitude: 38.568166. A colocação de um marcador (Placemark) nessas coordenadas terá o nome “Departamento de Informática – Uévora” e caso o utilizador clique sobre esse marcador verá a informação, ou descrição, “O Departamento de Informática é uma unidade científico-pedagógica da Universidade de Évora que tem como missão a realização de tarefas de investigação e ensino, de extensão universitária e de prestação de serviços especializados à comunidade

nos domínios da Informática”. O resultado utilizando um Sistema de Informação Geográfica, neste caso o Google Earth, terá o seguinte aspecto como mostra a Figura 1.



Figura 1 - Exemplo da Localização do Departamento de Informática da Universidade de Évora no Google Earth

2.2.5 GEDCOM XML

Tal como verificado anteriormente, a representação GEDCOM foi desenvolvida para permitir uniformidade e flexibilidade na troca de informação de dados genealógicos. No entanto é conhecido por todos que a Internet ganha todos os dias cada vez mais utilizadores e com o uso generalizado da Internet, os próprios utilizadores reconhecem e necessitam de novas formas de comunicar e de interagir. Foi assim que a *World Wide Web Consortium* (W3C) desenvolveu a representação XML tal como já verificado no tópico respectivo. Pode-se considerar semelhante ao GEDCOM na medida em que ambos utilizam etiquetas (tags) e estruturam a sua informação num nível hierárquico. No entanto, estas duas representações diferem já que o XML é um reconhecido padrão mais fácil de entender, já que os utilizadores da representação GEDCOM tiveram que compreender a peculiar sintaxe antes de entender o próprio conteúdo.

Pegando na primeira palavra da abreviatura XML, o “X” refere-se à extensibilidade. Esta capacidade permite-nos assim, especificar um vocabulário próprio de XML, definindo um conjunto de etiquetas e de uma estrutura adequada para um determinado tipo específico de aplicação (W3C, 2003). Assim a especificação GEDCOM XML é um vocabulário XML para dados genealógicos.

Dois principais objectivos da representação XML é facilitar a troca de informação num formato adequado e permitir o processamento e publicação da informação através de ferramentas WEB. Por seu lado, a representação GEDCOM foi concebida essencialmente assente no primeiro objectivo. Conseguir integrar o segundo objectivo da linguagem XML na representação GEDCOM requer aplicações e conversões destinadas a essa finalidade, tornando-se este o objectivo desta integração da XML com a GEDCOM (Family History Department - The Church of Jesus Christ of Latter-day Saints, 1996).

A complexidade e subtileza dos dados genealógicos é elevada o que faz salientar que o objectivo de integrar a XML e a GEDCOM não é que a análise dos dados seja feita utilizando um browser por exemplo, mas sim criar novos mecanismos de troca de informação deixando a análise e apresentação dos dados genealógicos na WEB para ferramentas específicas (A Igreja de Jesus Cristo dos Santos dos Últimos Dias, 2001).

Sabendo da evolução que a representação GEDCOM sofreu ao longo dos tempos, isto provocou que a tecnologia, uso e a compreensão dos dados genealógicos mudassem também neste intervalo de tempo.

De notar que apesar de ainda não existir um formato GEDCOM XML oficial penso que não irá demorar muito tempo até esse dado seja consumado, já que começam a surgir cada vez soluções integrando esta tecnologia, muitas delas de organismos com elevado renome no mercado das aplicações genealógicas. Assim, o objectivo da definição ou da integração destas duas representações é a de efectuar uma “tradução 1-1” da representação GEDCOM para a representação XML, isto é, cada elemento da representação GEDCOM corresponde a um elemento XML e desta forma garantir o mesmo conteúdo do ficheiro GEDCOM original, mas num formato mais limpo, mais perceptível e numa estrutura mais robusta (A Igreja de Jesus Cristo dos Santos dos Últimos Dias, 2001).

Desta forma, o objectivo passa por conseguir juntar o melhor dos dois mundos, permitindo que os utilizadores entendam não só a forma, mas também o conteúdo. Juntando ainda o facto de o XML ser amplamente aceite, esta representação XML traz consigo o facto de ser possível integrar múltiplas opções quer ao nível de interfaces para a internet, quer a nível de *parsers* e muitas outras ferramentas disponíveis para operar sobre XML proporcionando assim novas opções quer sobre a forma, quer sobre o conteúdo de dados (W3C, 2003).

Assim, a representação XML traz efectivamente várias vantagens a juntar à representação GEDCOM tradicional. Unir esforços no sentido de integrar estas duas vertentes procura ser o futuro da análise de dados genealógicos e caminhar na direcção correcta.

Após uma breve explicação sobre estas duas representações, segue-se um exemplo da transformação de uma representação GEDCOM para a representação XML com GEDCOM integrado.

GEDCOM

0 @I1@ INDI	'Criação de um novo individuo com o código I1
1 NAME João Leitão	'Identificação do nome do indivíduo I1
1 BIRT	'Identificação da informação de nascimento
2 PLAC Évora	'Identificação do local de nascimento

GEDCOM + XML

```
<INDI>
  <ID>
    @I1@
  </ID>
  <NAME>
    João Leitão
  </NAME>
  <BIRT>
    <PLAC>
      Évora
    </PLAC>
  </BIRT>
</INDI>
```

Assim, para que a transformação fosse possível foi necessária a criação de um elemento que se trata do ID em XML já que este elemento faz parte do sistema padrão de ligação de elementos e é equivalente ao sistema de ligação de elementos na representação GEDCOM.

2.3 TECNOLOGIAS – ESTADO DA ARTE

No presente capítulo analisam-se as tecnologias, as ferramentas e os processos actuais para deles poder extrair o importante a utilizar no desenvolvimento do Sistema Informático de Gestão e Análise de Genealogias.

O Estado da Arte é uma das partes mais importantes de todo este estudo, uma vez que é mencionado o que já existe sobre o tema em questão, sendo o objectivo conhecer o maior número de ferramentas ligadas ao tema em estudo para que se possa retirar o melhor das arquitecturas de cada uma delas e construir um sistema diferente capaz de responder às exigências do universo da Genealogia.

Apesar de se tratar de um sistema informático que abarca diversas áreas, desde sistemas de bases de dados, até sistemas de informação geográfica, a essência prende-se na gestão e análise de dados genealógicos, pelo que o estado da arte é elaborado nesse sentido. Assim este ponto encontra-se separado em duas partes: uma primeira em que retrata o estado da arte da tecnologia GEDCOM e respectivos programas associados; e uma segunda parte mostrando a tecnologia GEDCOM integrada com a representação XML, formando, tal como definido, a representação GEDCOM XML e mostrando desta forma, que ferramentas já existem no mercado dispondo desta tecnologia.

2.3.1 GEDCOM

Na sua essência, qualquer programa de Genealogias ajuda a traçar a história da família com auxílio electrónico (Beecher, 2010). É no final do século XX, que começaram a surgir vários programas genealógicos computadorizados. Os softwares informáticos considerados como sendo os pioneiros na transformação para a era electrónica de dados genealógicos incluem nomes como o software *Roots* e o *Personal Ancestral File*. O mais curioso é que ainda hoje estes programas existem, em versões melhoradas, mas carregando consigo as insígnias de terem sido os pioneiros de um tão importante progresso genealógico (About.com: Genealogy, 2009).

Neste momento, são vários os programas existentes no mercado das aplicações genealógicas, sendo que, independentemente do factor monetário, isto é, ser um software pago ou gratuito, todos eles dispõem de funcionalidades base que passam pela criação visual da árvore genealógica e do respectivo preenchimento e manutenção.

No entanto, surgem programas com maior destaque do que outros por diversos factores. O que importa reter desde cedo é que não existe um melhor software genealógico já que cada caso é um caso particular, e como tal é necessário encontrar um software que satisfaça os requisitos pelo que não se trata de um método infalível a escolha da aplicação genealógica a utilizar.

Em primeiro lugar, começo por enumerar os softwares que no mercado das aplicações genealógicas não apresentam um elevado renome, sendo o principal objectivo, conhecer e encontrar para cada um destes softwares razões que justifiquem esse factor.

Assim, começo por enumerar o software genealógico *DoroTree 2.1*³. A justificação para não alcançar patamares elevados de renome no mercado das aplicações genealógicas acaba por ser simples já que este software foi concebido para o estudo da história de famílias Judaicas (DoroTree Technologies Ltd., 2005). Uma vez que não era objectivo deste trabalho realizar a história de famílias Judaicas, acaba por ser factor para perder aceitação pelo simples facto de se restringir a uma comunidade. No entanto, é reconhecido o mérito, apesar da pouca flexibilidade que apresenta para restantes comunidades, trata-se de um software com uma interface muito atractiva, amigável ao utilizador, sendo esse o principal atractivo deste programa tal como demonstra a Figura 2. Olhando para as funcionalidades que este programa permite, como a criação de relatórios de Bar Mitzvah, suporte para calendário Judaico entre outros, chega-se à conclusão que este programa, para essa finalidade, é relativamente bom.

³ http://www.dorotree.com/eng_chul_order.html

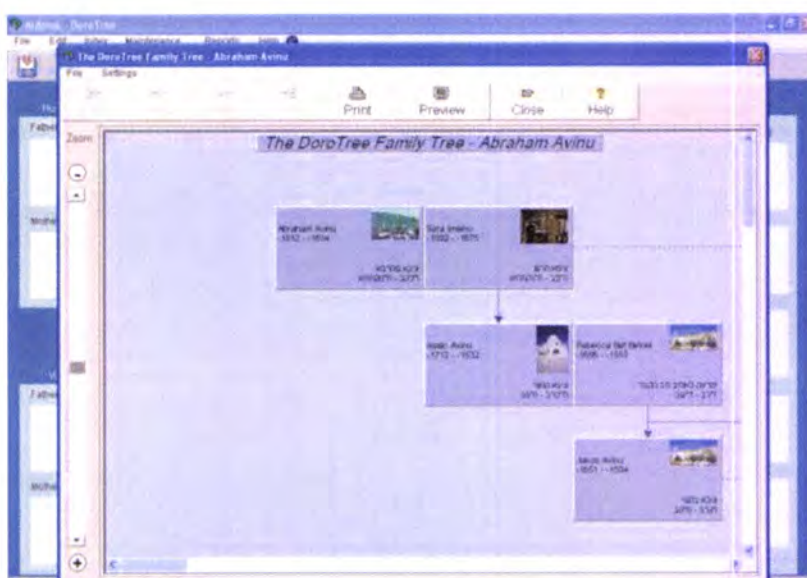


Figura 2 - Exemplo da aplicação DoroTree 2.1

À semelhança do software anterior, no que confere ao pouco renome no universo das aplicações genealógicas, surge o software genealógico *Famtree 4.43*⁴. A razão pela qual esta aplicação é conotada com baixas prestações é totalmente diferente das razões do software *DoroTree*, já que mesmo quando se acede à página oficial deste programa para efectuar a transferência do instalador, percebe-se que trata de uma estética desactualizada, dada a pouca preocupação com a imagem tal como demonstra a Figura 3.

Contudo, nem só a imagem prejudica este software, já que o seu funcionamento também deixa muito a desejar. A associação de indivíduos a famílias de forma manual ao invés de outros softwares cuja associação é feita automaticamente, é de facto um problema grave desde software, incluindo a não identificação automática para os formatos da data já que as datas constituem um elemento muito importante na análise de dados genealógicos, pelo que é necessário manter a integridade das datas (se é remota, aproximada ou exacta) sendo este um factor que se pretenda que a aplicação apresente (No1Reviews.com, 2009).

Assim, são argumentos que justificam a posição que este programa ocupa no mercado das aplicações genealógicas. (TopTenReviews, 2010).

⁴ <http://www.famtreesoftware.com/>

Figura 3 - Exemplo da aplicação FamTree 4.3

Com aspecto gráfico totalmente oposto ao programa anterior (*FamTree*) é o caso do software genealógico *WinFamily 2009*⁵. No entanto, a nível de funcionalidade falha em pontos semelhantes ao software genealógico *FamTree*. O preço de aproximadamente 95 euros pode constituir logo a partida um entrave, dependendo das situações, podendo ser ultrapassado pelo grau de apelo visual muito elevado, isto caso as preferências do utilizador recaiam sobre aplicações com um nível estético elevado tal como demonstra a Figura 4. Fica a sensação de que esta ferramenta apresenta um potencial enorme pelo que se as funcionalidades acompanharem o ambiente gráfico, podemos estar perante um possível ferramenta de topo, sendo que no actual momento, ainda não o é (Beecher, 2010).

⁵ <http://www.winfamily.com/>

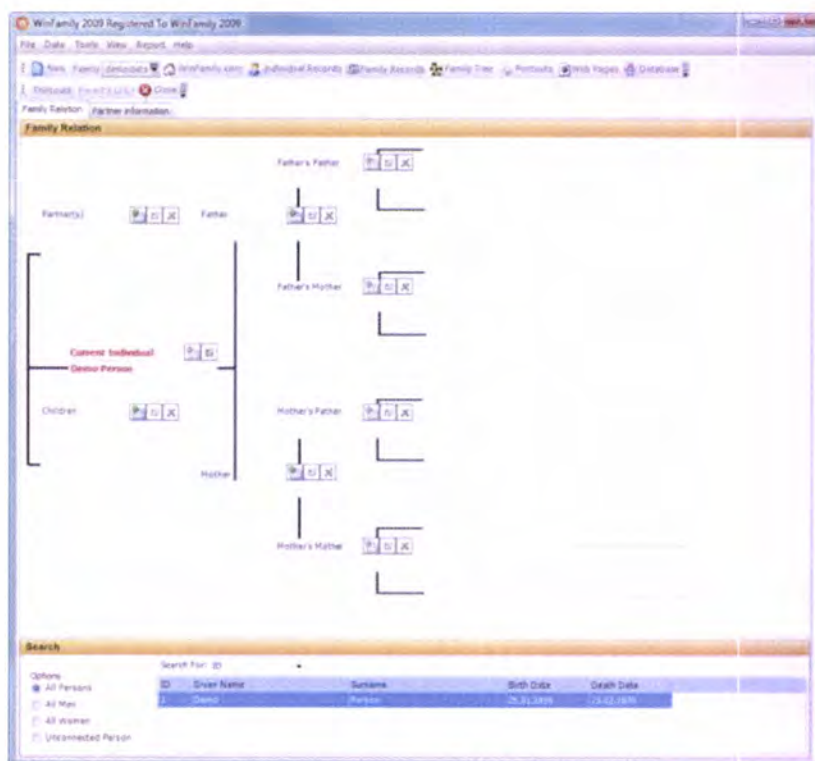


Figura 4 - Exemplo da aplicação WinFamily2009

O software genealógico *Reunion 9* trata-se de uma aplicação específica para uma plataforma em concreto, isto é, a plataforma de operatividade ser a plataforma Mac OS X 10.3.9⁶ ou versões superiores. Assim, por ser um programa aplicado a um sistema operativo em concreto, que ao contrário das aplicações genealógicas já apresentadas, não permite a integração de plataformas de sistemas operativos distintas, pode constituir à partida motivo pelo qual este programa não é reconhecido como solução de topo na comunidade em geral. Todavia, é por muitos considerado como o melhor programa de genealogias para Macintosh⁷ (ConsumerSearch, 2010) mesmo sabendo que a nível monetário, pode tratar-se de um software dispendioso com o preço desta aplicação a rondar os 80 euros. Considerado como um programa de fácil uso e bastante fácil de aprender tal como demonstra a Figura 5, apresenta características fundamentais para os amantes da tecnologia produzida pela empresa Apple Inc®, proporcionando a transferência de informação para um iPod®⁸ (Battersby, 2007), apresenta como pontos

⁶ Denominação do Sistema Operativo padrão dos computadores Macintosh

⁷ Computadores produzidos pela empresa Apple Inc®

⁸ Marca Registrada da Apple Inc® e refere-se a um dispositivo de áudio digital criado pela própria Apple®

fracos precisamente o facto de apenas operar na plataforma Mac OS e ainda a necessidade de compreender a linguagem HTML e ser conhecedor da linguagem para publicar a genealogia num website, algo que é feito por muitos programas apenas carregando num botão (Crume, 2010).



Figura 5 - Exemplo da aplicação Reunion 9

Outro software genealógico bastante interessante é a aplicação genealógica denominada por *The Master Genealogist*⁹. Com uma primeira impressão de ser bastante difícil de operar, tal como demonstra a Figura 6, é considerado um programa de elite para profissionais do estudo da genealogia, já que apresenta sofisticadíssimos métodos de publicação (TopTenReviews, 2010). O preço a pagar de aproximadamente 50 euros por este software pode constituir um entrave para certos casos, existindo alternativas no mercado que se podem comparar a este software a preços inferiores.

⁹ <http://www.whollygenes.com/Merchant2/merchant.mvc?screen=TMG>

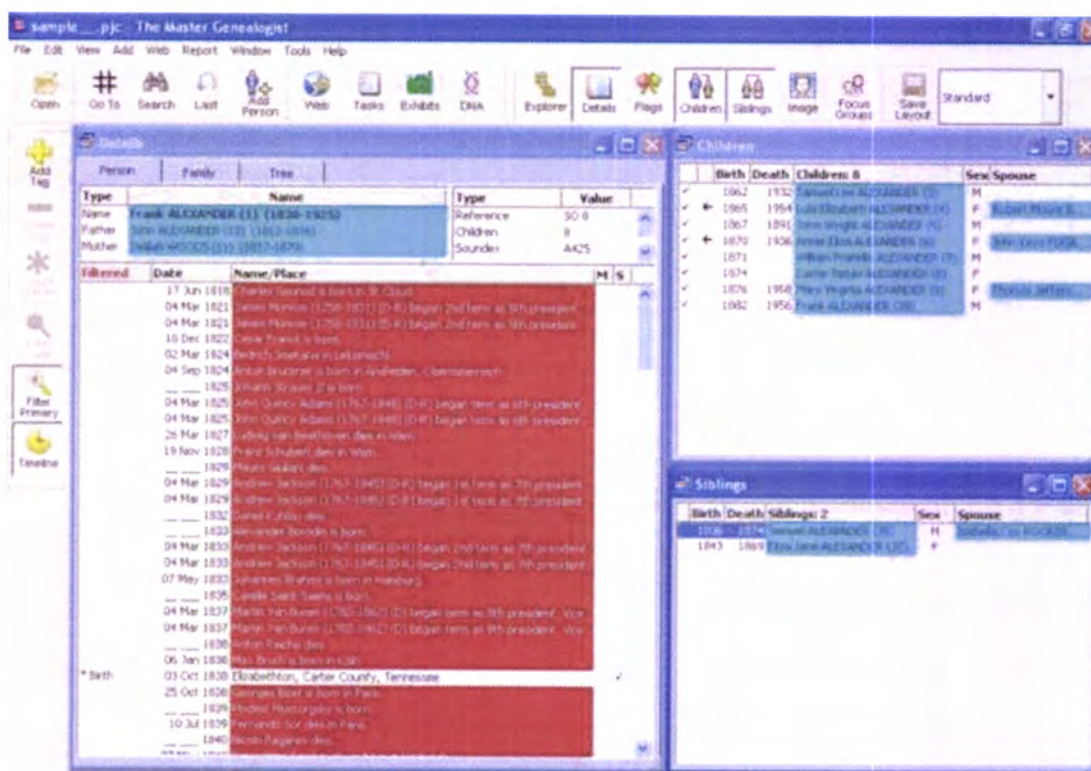


Figura 6 - Exemplo da aplicação The Master Genealogist 7

Num patamar muito semelhante ao programa anterior (*The Master Genealogist 7*) está este programa: *Family Historian 4*¹⁰. A facilidade de utilização demonstrada pela Figura 7 e premiada com várias distinções em 2010, faz desta aplicação uma ferramenta que dispõe de recursos online que podem ser muito úteis para quem procura na Internet uma fonte de informação genealógica (Family Historian Official Webpage, 2010). A criação de gráficos sólidos com a possibilidade de recorrer a imagens de cada indivíduo, podendo facilitar a sua leitura, em casos que assim se justifiquem, constitui assim um ponto forte desta aplicação. Como pontos fracos surgem apenas algumas falhas na apresentação de gráficos e relatórios que em alguns casos podem dificultar a visualização de dados (GenSoftReviews, 2010a).

¹⁰ <http://www.family-historian.co.uk/>

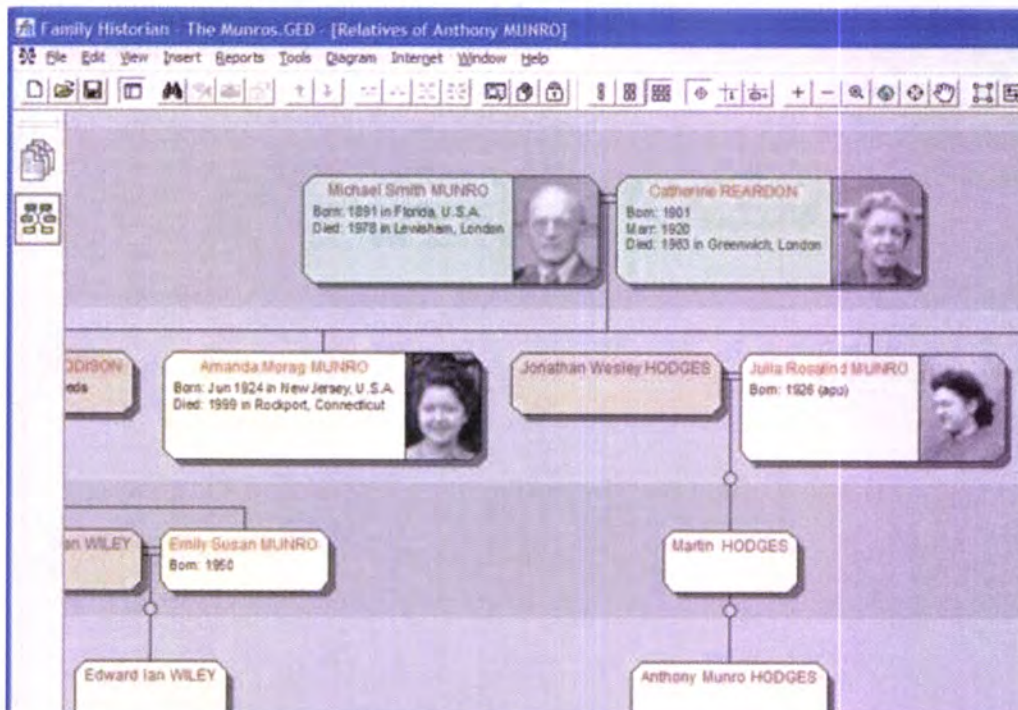


Figura 7 - Exemplo da aplicação Family Historian 4

O software genealógico *Legacy Family Tree 7*¹¹, trata-se, sem dúvida, de um programa muito bom, fácil de usar e com boas perspectivas futuras utilizando um design muito apelativo tal como demonstra a Figura 8 (Beecher, 2010). No entanto, verifica-se que este software pode ser aplicado mas num determinado contexto, voltado muito para situações em que a parte gráfica é o centro da questão, não procurando por isso uma solução que ao nível da genealogia e da análise seja poderosa. Um exemplo concreto é o caso da análise de uma família com um elevado número de gerações antepassadas que pode levar a uma análise confusa, dado o elevado grafismo empenhado por este software aquando do momento de análise (Crume, 2010).

Para alguns autores o *Legacy Family Tree 7* é considerado como o melhor programa de genealogias (Beecher, 2010; TopTenReviews, 2010). É necessário ser conhecedor do mundo da Genealogia para que se consiga interpretar de forma rápida e intuitiva as opções deste programa. As suas mais-valias são sem dúvida o apelo visual, a documentação, já que disponibiliza inclusive um vídeo a explicar alguns pontos a realizar e

¹¹ <http://www.legacyfamilytree.com/>

mesmo a parte da publicação e criação de relatórios é francamente positiva (Beecher, 2010). No entanto é na criação e visualização de gerações de antepassados que existem alguns problemas com este software (ConsumerSearch, 2010). A constante procura de um elevado grau visual para o utilizador trai esta aplicação quando é confrontada com grandes volumes de dados a apresentar ao utilizador sendo este o principal ponto fraco da aplicação (Crume, 2010; ConsumerSearch, 2010). O preço apesar de não ser nenhum impeditivo, 27 euros, pode ser considerado como um elemento a pensar, já que, caso exista um software gratuito com as mesmas funcionalidades, ou até melhores, a escolha poderia recair sobre o software gratuito, desde que desse as mesmas garantias.

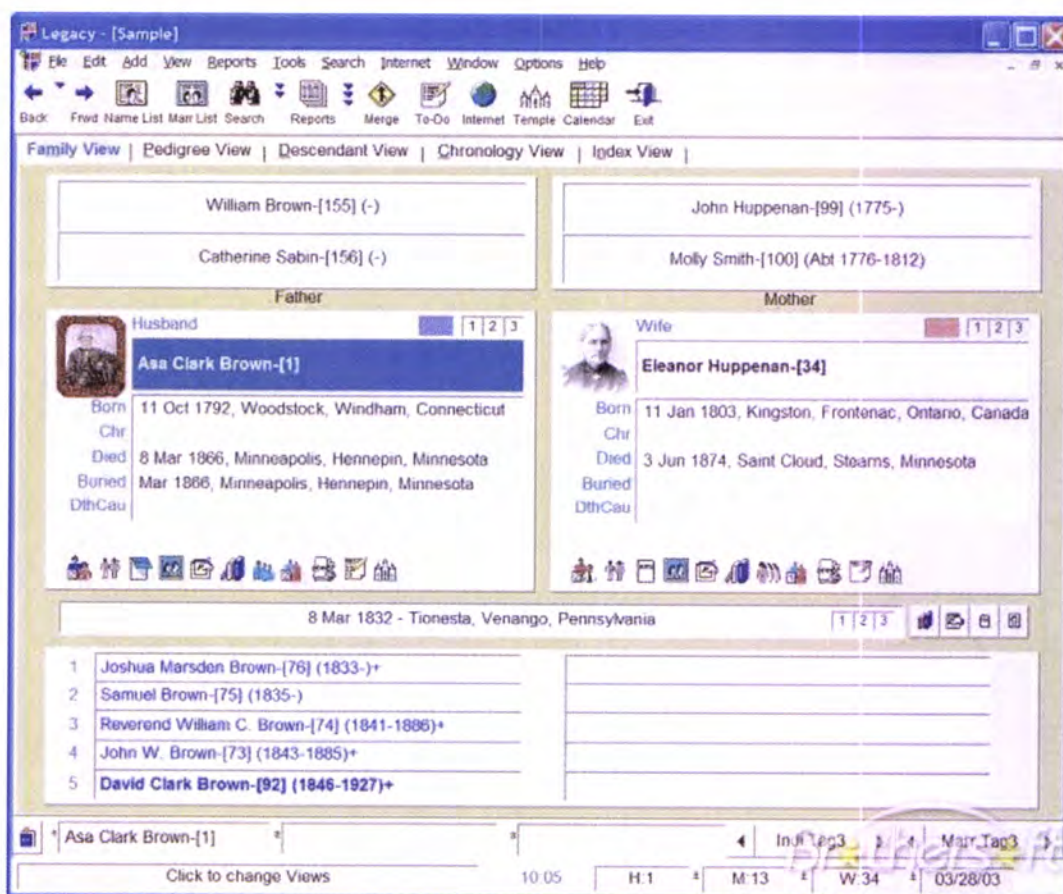


Figura 8 - Exemplo da aplicação Legacy Family Tree 7

A aplicação genealógica *Roots Magic 4*¹² é considerado como um dos potenciadores da era de programas genealógicos e ainda hoje continua a ser considerado como um programa

¹² <http://www.rootsmagic.com/>

muito fácil de usar tal como demonstra a Figura 9. Apresenta a característica de ser dos poucos programas que reúne um consenso entre vários autores já que, neste caso, as opiniões convergem no sentido favorável desta aplicação, afirmando na generalidade dos autores, que se está perante um software muito poderoso quer ao nível da usabilidade, quer a nível da publicação e criação de relatórios (Eastman, 2010). Mesmo nos pontos fracos existe uma grande unanimidade ao definir que a falta de suporte para plataformas Macintosh e algumas falhas ao nível da pesquisa são os pontos fracos desta aplicação (Beecher, 2010; ConsumerSearch, 2010; Crume, 2010; TopTenReviews, 2010) que não fosse, tal como a anterior, o facto de ser um software pago, poderia apresentar um número ainda maior de utilizadores no mercado das aplicações genealógicas.

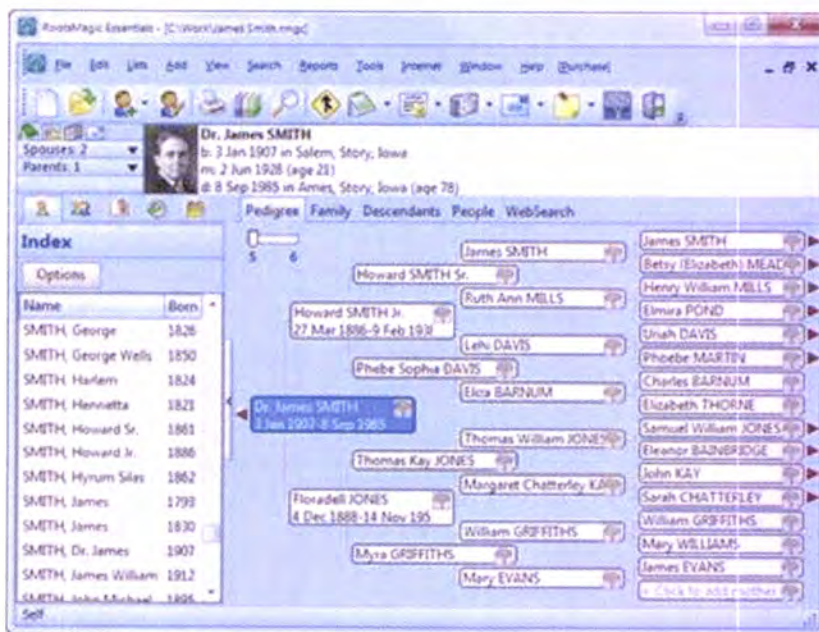


Figura 9 - Exemplo da aplicação Roots Magic 4

Os softwares genealógicos Ancestral Quest¹³ e Personal Ancestral File¹⁴ (PAF) são programas muito semelhantes pelo que o principal factor que os distingue é o simples facto de o programa Ancestral Quest ser um software pago e o Personal Ancestral File ser software gratuito. Ambos apresentam uma estrutura de menus, de grafismo, de opções

¹³ <http://www.ancquest.com/index.htm>

¹⁴ <http://www.familysearch.org/eng/paf/>

muito semelhantes pelo que existem autores que defendem que o Ancestral Quest é a versão paga do Personal Ancestral File (TopTenReviews, 2010; Crume, 2010). A Figura 10 demonstra essa mesma comparação de grafismo podendo-se confirmar aquilo que é afirmado.

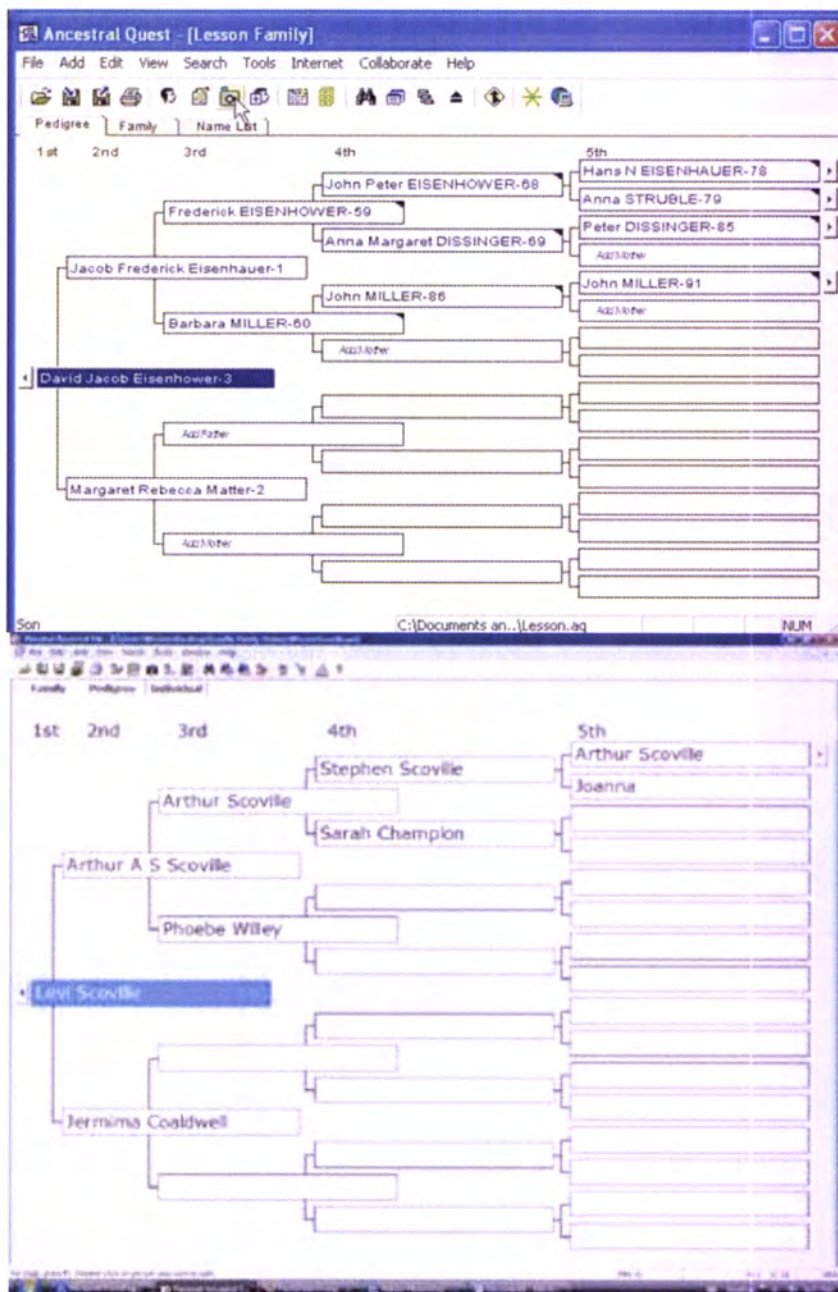


Figura 10 - Comparação de Grafismo entre as aplicações Ancestral Quest e Personal Ancestral File

À semelhança dos dois softwares anteriores, acontece o mesmo caso com os programas Family Tree Maker 2010¹⁵ e o Family Tree Builder¹⁶. A situação anterior repete-se sendo neste caso o Family Tree Maker considerada como a versão paga do Family Tree Builder (TopTenReviews, 2010; Crume, 2010). As semelhanças entre estes softwares começam desde logo no nome de ambos os programas, mostrando a Figura 11 mais pontos de comparação entre estes dois softwares genealógicos. O grafismo inovador e voltado para o futuro, constantes opções de publicação na Web faz com que estes softwares sejam voltados para o futuro da genealogia enquanto o primeiro conjunto são programas fortemente conhecidos, não pelo seu apelo visual, mas sim pela robustez e consistência que apresentam quer na gestão, quer na análise dos dados genealógicos (ConsumerSearch, 2010; Powell, s/d).

Por se verificar que o Ancestral Quest se trata de uma versão em tudo semelhante ao Personal Ancestral File e que o Family Tree Maker se trata de uma versão em tudo semelhante ao Family Tree Builder diferenciados apenas no factor monetário, centra-se atenções a partir de agora apenas nos dois softwares gratuitos: Personal Ancestral File 5.2 e o Family Tree Builder 4.0.

Por um lado o Personal Ancestral File é definido como um dos pilares do mundo das aplicações genealógicas (About.com: Genealogy, 2009; GenSoftReviews, 2010c; TopTenReviews, 2010), e por se tratar de um programa com história, chega a ser mesmo considerado como o mais popular programas de genealogia criado pela Igreja de Jesus Cristo dos Santos dos Últimos Dias (Mórmones) que carrega consigo todas as funcionalidades necessárias aliado a um design simples mas não simplista capaz de responder às mais variadas exigências (Powell, s/d). Por outro lado, tem-se o Family Tree Builder que impressiona logo à primeira vista com um design inovador com funcionalidades claramente apelativas para genealogias mais recentes (GenSoftReviews, 2010b). Assim, à primeira vista, a escolha é difícil, pelo que é necessário efectuar um estudo cuidadoso de cada um deles, e consoante as necessidades, escolher o software a utilizar.

¹⁵ <http://www.familytreemaker.com/>

¹⁶ <http://www.myheritage.com.pt/family-tree-builder>

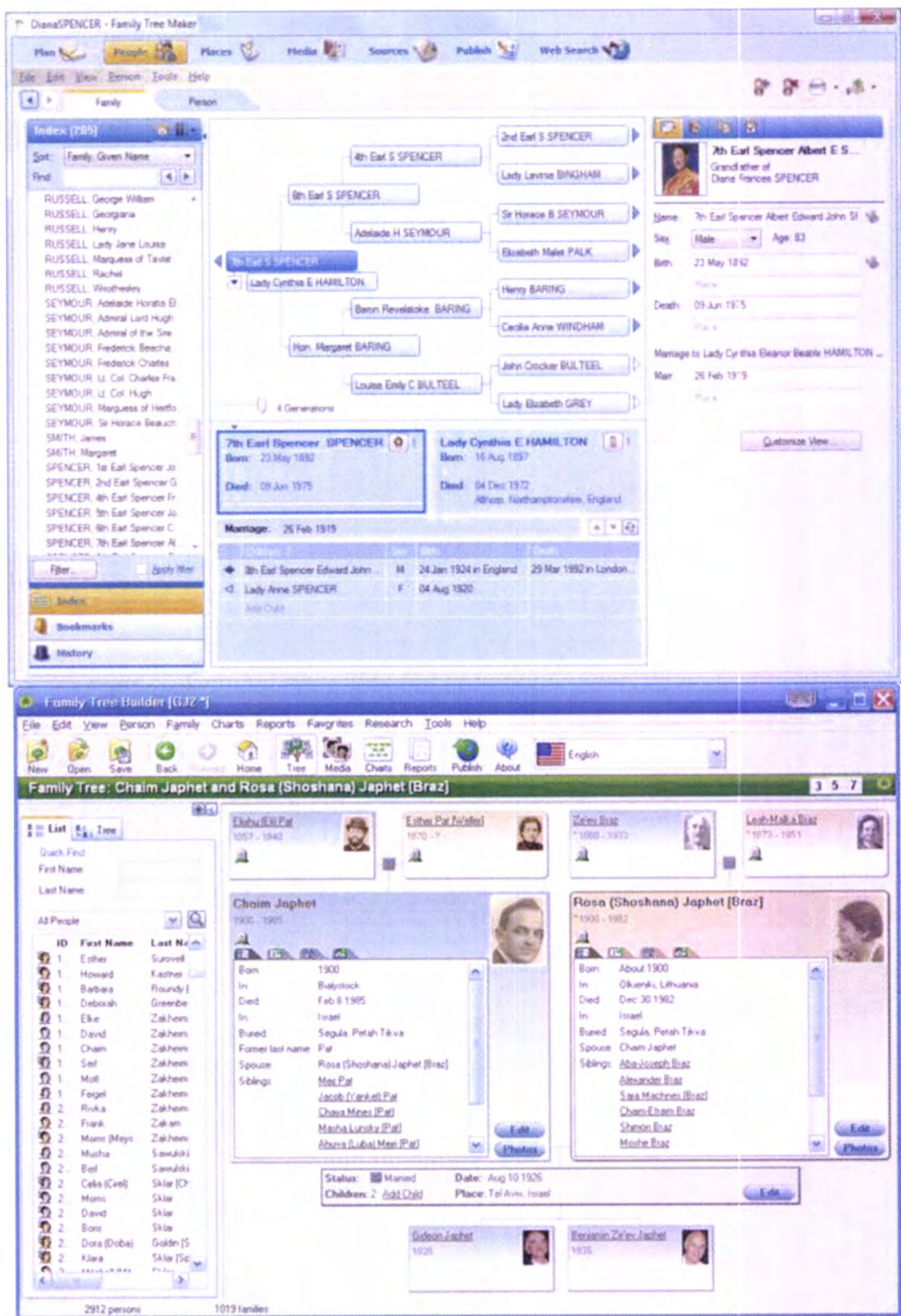


Figura 11 - Comparação de Grafismo entre as aplicações Family Tree Maker e Family Tree Builder

O Family Tree Builder é um programa tal como dito anteriormente com um design bastante atractivo. A sua capacidade de interagir com o utilizador de forma bastante amigável é sem dúvida um dos muitos pontos a favor deste software (Smith, 2010).

Apresentando um suporte de 35 idiomas e tecnologias inovadoras para a publicação de fotografias associadas a cada pessoa e a cada membro, integrando inclusive um mecanismo único nos programas de genealogia com reconhecimento fácil, permitindo através de uma foto de família retirar as devidas fotografias individuais (MyHeritage, 2010).

A vantagem de uma fácil expansão de dados genealógicos neste programa, juntando a fácil instalação são pontos atractivos deste software que joga com pontos fracos o facto de apresentar alguma publicidade, isto é, é necessário ter algumas cautelas aquando da instalação do programa já que caso não se tenha atenção corre-se o risco de se instalar barras de ferramentas para browsers¹⁷, mudar a página de entrada do browser entre outros tipos de publicidade (GenSoftReviews, 2010b).

Na apresentação do produto no site oficial aparece ainda a possibilidade de ter dados genealógicos georreferenciados, isto é, descobrir de onde vieram os antepassados através de um mapa-mundo (MyHeritage, 2010). No entanto a desilusão surge quando é constatado que essa funcionalidade apenas está disponível caso se empregue aproximadamente 60 euros para a actualização de uma versão mais sofisticada (chamada de Premium) e ainda o facto de não permitir um número de indivíduos acima dos 250, consideração que pode levar certos autores a não optar por esta ferramenta, já que facilmente se ultrapassa esse número de indivíduos em estudo.

A escolha pode assim não recair sobre este programa já que, por exemplo, dadas as circunstâncias do Caso de Estudo, em que não existe registo de fotografias individuais, não existe recursos para elementos multimédia, o número de indivíduos a estudar é superior a 250, não é desejável a publicação dos dados na Internet nem dar a conhecer ao público os dados armazenados, podem ser, razões e fundamentos que levem a escolha a não recair sobre este produto. No entanto, o idioma de funcionamento do programa em Português e a possibilidade (paga) de visualizar dados genealógicos georreferenciados possam pesar na altura de escolha de qual o software genealógico a adoptar.

¹⁷ Programa utilizado para “navegar” na Internet

Por seu lado, o Personal Ancestral File 5.2 assume-se como uma aplicação genealógica utilizado por curiosos, iniciantes ou até mesmo profissionais do ramo da História, sendo definido como um programa com funcionalidades muito poderosas, com uma interface amigável não muito preocupada com elevados aspectos visuais (ConsumerSearch, 2010).

A facilidade de introdução de informação é notória, quer seja pela via manual (introdução directa do utilizador) que seja pela via automática (importação através de um ficheiro GEDCOM) tendo sido aquele que melhores mecanismos apresentava na importação de ficheiros GEDCOM (Powell, s/d).

As limitações deste software começam pelo facto de a publicação ao nível da Internet ser imperfeita quando comparada com outros softwares muito vocacionados para esta área (TopTenReviews, 2010; Morán, s/d). No entanto, para casos em que não constituía este ponto como objectivo, esta limitação rapidamente é superada.

Outras limitações são ao nível da multimédia e o facto de a última actualização ser em 2002 (TopTenReviews, 2010). Apenas uma fotografia é suportada por pessoa e este programa apenas oferece suporte a elementos básicos de multimédia (Powell, s/d; ConsumerSearch, 2010). Pelas mesmas razões apresentadas no parágrafo anterior, também estas limitações podem ser superadas, se estas características não fizerem parte dos objectivos primordiais dos casos concretos de estudo.

Assim, o Personal Ancestral File trata-se por um lado como um pilar na construção e desenvolvimento de softwares genealógicos, e por outro de uma das aplicações genealógicas mais utilizadas em versões electrónicas (Powell, s/d). Ajuda os utilizadores a organizarem os registos genealógicos, permitindo a produção de relatórios, de gráficos, em formato papel ou no ecrã capaz de satisfazer os requisitos quer de curiosos desta área, quer de profissionais como é o caso concreto deste estudo (Morán, s/d; Powell, s/d).

2.3.2 GEDCOM XML

No entanto, nenhuma das aplicações até aqui apresentadas surgem aliadas à representação XML. Por este facto, surge a necessidade de exemplificar o “mercado” das aplicações XML ligadas à genealogia.

Conhecendo a capacidade que a representação XML tem aliada à sua imagem de ser uma ferramenta que apresenta um futuro promissor sempre com constantes melhoramentos, aliar uma tecnologia tão consolidada como a representação GEDCOM à representação XML parece constituir uma enorme mais-valia na disponibilização de informação genealógica de forma coerente, permitindo assim uma partilha de dados entre os utilizadores a larga escala (Skonnard, 2004).

Apesar de a representação GEDCOM ser, a par do XML, uma representação hierárquica, não apresenta na sua estrutura qualquer elemento baseado na representação XML (Skonnard, 2004).

Sabendo as potencialidades de cada uma das representações, aliar ambas é considerado como o futuro dos dados genealógicos já que é possível manter um enorme volume de dados genealógicos com perspectivas tecnológicas de ser possível efectuar trocas ou simplesmente publicações desses mesmos dados a nível global (Eastman, 2001).

A não existência de um formato oficial, promove que vários autores já tenham definido normas denominadas por GEDXML, GEDCOM 6.0, GEDCOMXML entre outras, sendo que nenhuma é ainda aceite como a norma oficial do formato GEDCOM XML (A Igreja de Jesus Cristo dos Santos dos Últimos Dias, 2001).

Existem assim, várias normas, sendo no entanto as mais significativas e com maior peso na sociedade tecnológica, o caso do “rascunho” que a Igreja de Jesus Cristo dos Santos dos Últimos Dias lançou, em 2001, propondo a adopção do novo padrão para a troca de

informação, denominada por versão 6.0, mantendo o nome GEDCOM. Esta versão GEDCOM 6.0 passaria a especificar uma representação com uma flexibilidade maior que a representação GEDCOM, e que fazendo uso da internet e utilizando todas as suas potencialidades, esta representação GEDCOM XML, poderia adaptar-se à expansão da Internet a larga escala de uma forma muito mais rápida e eficiente do que a representação GEDCOM já que as limitações conhecidas (A Igreja de Jesus Cristo dos Santos dos Últimos Dias, 2001).

Sendo esta uma tradução integral entre a representação GEDCOM e a representação XML, este facto permite que a tradução seja feita “peça a peça” ou seja, existe uma transformação elemento a elemento de uma representação para outra (A Igreja de Jesus Cristo dos Santos dos Últimos Dias, 2001).

Quer esta representação da Igreja de Jesus Cristo, mais conhecida pela comunidade dos mórmones, quer a representação apresentada pela Microsoft na voz de Aaron Skonnard, nenhuma altera o conteúdo da representação GEDCOM original, inclusive as suas etiquetas. Quer isto dizer que estas representações são uma tradução na íntegra do que é a origem da representação GEDCOM visando garantir apenas alterações ao nível da forma e não do conteúdo (A Igreja de Jesus Cristo dos Santos dos Últimos Dias, 2001; Skonnard, 2004).

Verificando que, quer a comunidade criadora da representação de dados genealógicos, GEDCOM, quer uma das mais famosas empresas informáticas a nível mundial, Microsoft, assumem este compromisso e procuram ser “aliadas” aquando da escolha de uma representação que integre a representação GEDCOM e a representação XML acredito que quando chegar a altura de assumir um *standard* oficial, que seja o que estes “dois pesos pesados” defendem.

No entanto irei mostrar que outras representações existem com estas duas tecnologias e o que diferem da representação defendida pelas duas entidades anteriores.

As diferenças surgem essencialmente ao nível do resultado produzido, sendo as transformações realizadas de forma diferente, sendo as diferenças ao nível de nomes de elementos e atributos da representação final.

Um exemplo do parágrafo supracitado é o caso da representação GeniML, definida como, *Genealogical Information Markup Language*, concebida por Jerry Fitzpatrick, e que as mudanças são essencialmente na referência a indivíduos, isto é, enquanto nos casos do padrão criado pelos mórmones, apoiado pela Microsoft, as referências são feitas através das características da representação XML, denominado por ID e IDREF, nesta representação GeniML é denominado pela palavra SOUR efectuando assim a referencia a indivíduos. No entanto, por ser uma representação ainda incompleta, não apresenta grande renome no mercado, constituindo assim problemas que talvez no futuro possam ser ultrapassados (Eastman, 2004b).

Outro formato reconhecido por garantir uma integração entre as tecnologias GEDCOM e XML é o formato GenXML, com algum renome já nos mercados das aplicações XML integradoras da tecnologia GEDCOM que garante que este formato é um formato de ficheiros com o objectivo de permitir a troca de informação entre aplicações genealógicas. As suas premissas principais começam pela simplicidade que outros programas apresentam na leitura desde formato, muito associado à forma como pode ser manipulado estes ficheiros em XML. As restantes vão de encontro às próprias potencialidades do XML, garantindo por exemplo a flexibilização na informação, a partilha e uma gestão em jeito de “base de dados XML” dos próprios dados genealógicos (Owe, 2010).

Por ser relativamente fácil criar esta plataforma de entendimento entre as representações XML e GEDCOM, qualquer autor pode desenvolver um sistema, sabendo ainda que a maioria do código das restantes aplicações é *open source*, pelo que é difícil descrever todos os sistemas existentes. Surge assim um outro nome, Ged2X, em que a principal diferença para com os restantes é a dispensável instalação de eventual software para efectuar a transformação da representação GEDCOM para a representação em XML. Desta forma, é efectuado o carregamento através do website oficial desta plataforma, e em seguida é visualizada através de páginas Web o seu conteúdo. Este é um claro exemplo, em que a

aplicação não serve apenas para efectuar a transformação, mas possibilita igualmente a visualização em páginas Web, isto apesar de o grafismo não ser de excelência (Ged2X, 2006).

Com o constante aparecimento de novas aplicações capazes de efectuar a transformação para o formato XML, existem ainda, consórcios destinados a promover a partilha de informação genealógica, tentando assim encontrar um local onde vários autores possam mostrar, partilhar e promover os seus trabalhos, com o objectivo de criar novas soluções no mercado das aplicações genealógicas. Um exemplo claro é o *XGenML*, que se trata de um consórcio global desenvolvido através da tecnologia XML com o objectivo de partilhar dados genealógicos. O objectivo é assim reunir autores, informações e todo o tipo de conteúdos que possam ser interessantes para a avanço da era electrónica da genealogia (XBI Software Inc. , 2006).

Constata-se assim que em Portugal, não existe nenhum software criado para este propósito mas já existem estudos no sentido de integrar o XML para anotações de registos paroquiais (Félix, 2002), para outras áreas da História (Ferreira, 2004) ou até mesmo para análise de cartografias (Oliveira, 2008) procurando constituir assim, mais um passo importante, rumo ao avanço tecnológico, de forma a oferecer ao utilizador um maior número de soluções.

CAPÍTULO III - SISTEMA INFORMÁTICO DE GESTÃO E ANÁLISE DE GENEALOGIAS

3.1 CONCEITO

A informação é o elemento principal da comunicação. No entanto esta palavra apresenta uma riqueza de conteúdo fazendo-a ser aplicada e escolhida para caracterizar diversas situações (Shannon, 2003). Quer isto dizer a informação pode definir múltiplas situações, sendo a escolhida para este estudo, a definição de Informação como sendo o resultado da organização de dados e respectivo processamento, representando assim uma modificação nos mesmos de forma a alcançar informação perceptível quer por máquinas, quer por seres humanos (Floridi, 2005).

Por seu lado, a observação de um objecto independentemente do seu tipo, fomentado por vários componentes ligados entre si formando um todo com funções cujos resultados pertencem ao conjunto dos componentes, define a primeira palavra das cinco que compõe esta ferramenta: Sistema. Assim, um Sistema, trata-se de um conjunto de partes que formam um todo (Gouveia, 1996).

A particularidade de ser um Sistema Informático, trata-se pelo simples facto de se estar perante uma ferramenta que faz uso das tecnologias de informação, nomeadamente o computador e todas as funcionalidades que este permite (Gouveia, 1996).

Uma vez que o Sistema faz uso das tecnologias de informação, trata-se assim de um Sistema de Informação, isto é, um sistema que oriente a tomada de decisão do utilizador, mas pede-se também a um sistema deste género que seja capaz de receber dados, processá-los, e transmitir informação para outros componentes do próprio sistema ou para o exterior (Nunes, 2009).

Sabendo da importância vital que a informação apresenta na nossa sociedade, tirar conclusões e tomar decisões requer normalmente o acesso a informação fidedigna e em grande escala, para que com base nessa informação sejam tomadas as devidas decisões (Gouveia, 1996).

Gerir e analisar Genealogias surge como a razão pela qual foi concebido este sistema de informação, ou seja, esta ferramenta tem de ser capaz de responder às exigências dos utilizadores genealógicos para alojar os dados que estes recolhem e posteriormente permitir a sua investigação.

Este sistema informático destina-se não só a especialistas da área da Genealogia, mas também a curiosos que pretendam estudar a história da família pelas mais diversas razões. Seja num âmbito profissional ou pessoal, o que verdadeiramente interessa é a aplicação conseguir satisfazer todos os graus de exigência, apresentando linguagens claras, concretas, para que nem seja demasiado simples para um especialista, nem demasiado complexa para um curioso desta área.

Desta forma, o Sistema Informático de Gestão e Análise de Genealogias define-se como sendo um sistema informático que realiza não só a gestão de dados genealógicos mas também promove a sua análise em duas vertentes: uma intrinsecamente ligada à Genealogia e ao estudo directo da história das famílias, das gerações, dos próprios indivíduos, e outra vertente ligada aos dados georreferenciados, isto é, a possibilidade de analisar dados ligados a lugares do mapa-mundo.

3.2 IMPORTÂNCIA

Qualquer sistema que apoie o utilizador na tomada de decisões independente do tipo, apresenta sempre uma importância acentuada, isto porque em teoria, através da informação, consegue-se ajudar o utilizador a tomar a decisão mais acertada (Nunes, 2009).

A ideia de concepção deste sistema informático surgiu no âmbito do projecto FCOMP-01-0124-FEDER-007360: Inquirir da Honra - Comissários do Santo Ofício e das Ordens Militares em Portugal (1570-1773), em que o objectivo central deste projecto financiado pela Fundação para a Ciência e a Tecnologia¹⁸ era precisamente comparar as redes de comissários, a actuação e as afinidades/discrepâncias entre as duas principais instituições que procediam ao apuramento da honra em Portugal (o Santo Ofício e a Mesa da Consciência e ordens), quer a nível local, quer no centro político (Portal Oficial Projecto FCT - PTDC/HAH/64160/2006, s/d).

Com esse grande objectivo definido, surgem assim necessidades de conhecer não só os indivíduos mas também assegurar e associar as suas genealogias para responder a diversas exigências.

No entanto, desde cedo se definiu que não se estava perante um sistema desenvolvido para o projecto, mas sim, um sistema que conseguisse ser aplicado ao projecto. Quer isto dizer que o projecto, assumiu ser o ponto de partida, mas não o ponto de chegada.

Uma vez que este projecto foi desenvolvido no âmbito do projecto supramencionado, foi necessário o recurso às tecnologias disponibilizadas pelo mesmo. Desta forma, a ferramenta escolhida foi o Microsoft Access 2007® para assegurar a transição entre a base de dados alojada em sistema *MySQL* e o utilizador.

Tal como foi visto anteriormente, um sistema de informação tem de disponibilizar recurso à análise dos dados nele armazenado. Desta forma, este sistema disponibiliza a análise de dados em três representações distintas: GEDCOM, para análise directa das Genealogias; KML, para análise de dados genealógicos georreferenciados; XML, para gestão e partilha de informação.

¹⁸ <http://alfa.fct.mctes.pt/>

É precisamente nesta última representação, XML, que reside o futuro deste sistema de informação. Aliar uma tecnologia com um futuro tão próspero e com imensas aplicabilidades, assegura desde logo que a construção de um sistema deste tipo não visa apenas assegurar o presente, mas também o futuro.

No entanto, e porque já foi definido, a informação é o elemento base de todo este sistema. Desta forma, passo a explicar qual a origem da mesma, seguindo-se a explicação aos dois restantes níveis de um sistema de informação: processamento e análise.

3.3 ORIGEM DA INFORMAÇÃO

Deste modo, a informação em forma de conjunto de dados, aparece como o primeiro elemento deste Sistema Informático de Gestão e Análise de Dados. Podemos estar perante um poderosíssimo sistema de gestão e análise, mas se os dados forem precários ou inexistentes, o sistema não consegue realizar qualquer operação.

Assim, o utilizador tem um papel chave na introdução de dados neste sistema. Essa mesma introdução é separada em dois momentos distintos: Informação individual e informação familiar.

Quando o utilizador pretende fazer uso do sistema de gestão para guardar informação genealógica, tem de primeiramente “criar” os indivíduos que o utilizador vai posteriormente estudar. Essa criação é o processo em que o utilizador irá inserir todos os elementos que dispõe e que se enquadram com a situação, sobre os determinados indivíduos e assim criar respectivos identificadores que serão denominados por um código unívoco dentro do universo dos dados em causa.

Após os indivíduos existirem, é possível efectuar associações entre eles, isto é, é possível associar um indivíduo a outro denominando essa relação de progenitor, por exemplo. No entanto, não faria qualquer sentido ter de rescrever toda a informação para associar esses

dois indivíduos. Como tal, e para tornar todo o processo mais rápido, este apenas necessita de saber utilizar os códigos disponibilizados pelo sistema.

Note-se que esta situação é um passo importantíssimo que pode comprometer toda a análise de dados caso a introdução for mal sucedida. A criação de mecanismos para minimizar os erros dos utilizadores foi pensada, sendo o caso das relações entre indivíduos um exemplo deste planeamento, isto é, quando é feita uma associação entre dois indivíduos, quando se coloca o código de ambos, automaticamente surgem os respectivos nomes podendo assim o utilizador notar que está a introduzir erradamente os dados, ou que por lapso surgiu uma distração que o levou a escolher uma associação errada.

Sabendo que a Genealogia é uma área que trabalha com um grande número de informações a opção recaiu sobre uma interface de rápida utilização, não deixando de ser por isso amigável ao utilizador. No entanto, a primeira opção era construir essencialmente uma ferramenta eficaz, funcional, que o utilizador se sentisse bem a trabalhar mas deixando de parte todo o grafismo e operações complexas de menus que a tecnologia de hoje nos permite.

A escolha da ferramenta Microsoft Access 2007® prende-se com o facto de ser uma solução bastante poderosa e flexível. Com um sistema gráfico bastante conhecido dos seus utilizadores (Groh, Stockman, Powell, Prague, Irwin & Reardon, 2007), foi uma das considerações que levou à escolha deste software para servir de plataforma de entendimento entre o utilizador e todo o Sistema de Gestão e Análise de Genealogias.

No entanto, a principal razão da escolha do Microsoft Access 2007® foi a possibilidade que nos oferece ao nível da transição e da flexibilidade que apresenta quando surgem mudanças no que respeita à origem da informação. Com um menu bastante rápido de ligações a bases de dados externas, este programa permite assim uma comunicação e uma interoperabilidade com outros sistemas de gestão dos melhores que existem no mercado (Chapple, 2010).

Assim o principal facto é colocar este sistema de gestão a operar com qualquer tecnologia, isto porque se o sistema opera através do *Microsoft Access 2007®* e se este, consegue interligar-se com sistemas de gestão como o *Oracle, MySQL, Postgres, SQLServer* entre muitos outros sistemas de gestão, então este Sistema de Gestão e Análise de Genealogias consegue assegurar igualmente qualquer tecnologia para a área de gestão desde que a mesma consiga interligar-se ao *Microsoft Access 2007®*, constituindo assim um passo importante na abstracção e na transparência ao nível do uso da aplicação por parte do utilizador, isto é, o utilizador não se irá aperceber caso a aplicação mude de tecnologia ao nível da gestão, isto porque a interface com o utilizador será sempre assegurada pelo *Microsoft Access 2007®*.

3.4 PROCESSAMENTO

O processamento neste Sistema Informático de Gestão e Análise de Genealogias é separado em 3 processos distintos. Isto porque o universo de análise também assim o pode ser, isto é, o meu objectivo pode ser apenas realizar posteriormente uma simples análise genealógica, ou gerar uma representação XML ou efectuar um processamento com o objectivo de integrar os dados genealógicos com um Sistema de Informação Geográfica.

Desta forma este ponto será dissecado em três sub-tópicos para uma melhor explicação e compreensão desta área complexa que é o Processamento no Sistema Informático de Gestão e Análise de Genealogias.

3.4.1 PROCESSAMENTO - GEDCOM

Com o objectivo de analisar dados genealógicos na sua essência, a ideia tal como descrita anteriormente foi de encontrar uma plataforma de interoperabilidade entre este Sistema Informático de Gestão e Análise de Genealogias e um programa de análise de genealogias. Tal como vimos também no ponto Tecnologias, o Personal Ancestral File (PAF) correspondeu as todas as exigências preliminares pelo que não existia assim necessidade de criar um programa semelhante a este para efectuar o mesmo tipo de análises. Em vez

disso, optou-se por alargar horizontes no universo da genealogia remetendo a explicação para o tópicio Processamento.

A plataforma de entendimento entre o nosso Sistema Informático e o PAF foi a representação GEDCOM, detalhada anteriormente, e que através da mesma, era possível transportar para o PAF todo o universo de análise que a parte de Gestão do nosso Sistema Informático continha, permitindo assim todas as possíveis análises que o próprio PAF oferece ao utilizador.

O objectivo de integrar o PAF no nosso Sistema Informático é apenas para situação de análise, ficando descorada qualquer utilização do PAF como gestão dos dados genealógicos, isto é, quando se analisa os dados no PAF se existir a necessidade de acrescentar um indivíduo, não poderá ser feito no próprio programa, uma vez que a integração é unidireccional, do nosso Sistema Informático para o PAF.

Entrando na área mais específica que é a programação do processamento para este caso propriamente dito, esta divide-se em três momentos.

3.4.1.1 Processamento - GEDCOM – Fase 1 - Header

Num primeiro momento, e para que a integração com o PAF seja bem sucedida, é necessário que a representação GEDCOM contenha um *header*¹⁹ com a informação do programa de destino. No entanto, para que seja possível escrever um ficheiro com toda a representação GEDCOM, é necessário editar-se esse ficheiro colocando-o apto para receber dados. Todo este processo é feito numa primeira fase, atribuindo um descritor para cada ficheiro de destino, sendo que a partir deste momento, todo o processamento efectuado e necessário à escrita do ficheiro, ficaria associada a esse mesmo descritor. O

¹⁹ Cabeçalho

cabeçalho com a informação relativa ao programa de destino (PAF) é estático já que o programa de análise utilizado é sempre o mesmo.

3.4.1.2 Processamento - GEDCOM – Fase 2 – Informação Individual

Após a criação do cabeçalho e respectiva escrita para o ficheiro com a representação GEDCOM, começa-se então a realizar o transporte da informação da parte da gestão de dados genealógicos do nosso sistema informático. Neste segundo momento, são processadas as informações individuais de cada pessoa, tendo sido inicialmente realizada uma consulta ao sistema de gestão “perguntando” que indivíduos existem no sistema para que possa ser procurada e posteriormente escrita a informação de cada um deles.

A informação a registar para cada um deles consiste nos dados, Código de Ego, Nome, Sexo, Ocupação, Data de Nascimento, Local de Nascimento, Data de Falecimento, Local de Falecimento e Notas Gerais. Toda esta informação é escrita para o ficheiro de destino com a representação GEDCOM através de uma função criada com o nome *EscreveCampos(codigoDeEgo, filename)* que tem como argumento da função o código do ego para que possam ser escritos todos os campos relacionados com esse Código de Ego e o *filename* que é o descritor do ficheiro de destino, para que depois de procurada a informação possa ser imediatamente escrita para o ficheiro de destino com a representação GEDCOM.

É importante salientar alguns casos particulares ainda na parte da informação individual. Tal como foi citado previamente, e por força das circunstâncias deste projecto, existiu a necessidade de registar os parentes desconhecidos. Como não é conhecida a relação que “une” dois indivíduos, e não querendo perder essa informação quando transportada para o PAF a solução passou por agregar às notas gerais essa informação quando o ego apresenta parentes desconhecidos.

A procura dos parentes desconhecidos através da função com o nome *EscrevePdesconhecidos(codigoDeEgo, filename)* que tal como a função anterior

(*EscreveCampos*) apresenta os mesmos argumentos. Esta função consiste na procura, para o Código de Ego em questão, se existe alguma associação entre esse mesmo Código de Ego e uma tia desconhecida (código da relação = "A"), um tio desconhecido (código da relação = "U"), um primo desconhecido (código da relação = "C") ou um parente desconhecido (código da relação = "U") quando apenas se sabe que essas duas pessoas estão relacionadas no entanto é desconhecida a forma como se relacionam.

Esta medida provou ser bastante útil já que posteriormente na fase de análise deu frutos ao conseguir encontrar-se em algumas situações a última peça do puzzle. Assim as notas gerais para cada indivíduo contem a informação dos parentes desconhecidos, da informação proveniente das cotas (Cota, Cota de Falecimento e Cota de Ocupação) e as informações gerais que porventura possam ser registadas para cada indivíduo.

3.4.1.3 Processamento - GEDCOM – Fase 3 – Informação das Relações

Após escritas as informações individuais para cada indivíduo, entramos na última fase do processamento para a criação do ficheiro com a representação GEDCOM. No entanto este passo é o mais complexo de toda a fase do processamento porque, tal como visto anteriormente, para este caso não se trata de "simples" consultas e respectivas escritas no ficheiro com a representação GEDCOM.

Para conseguir criar uma plataforma coerente na transformação dos dados armazenados no sistema de gestão para a representação GEDCOM, foi necessário analisar algumas situações. Desta forma, para cada indivíduo em primeiro lugar é escrita a família. Isto porque nesta parte do processamento, o elemento "mais atómico para cada indivíduo" é a sua família. A justificação passa pelo facto de evitarmos redundância de informação na construção do ficheiro GEDCOM para que por exemplo, a situação da Pessoa A ter como pai a Pessoa B seja considerada a mesma da Pessoa B ter como filho a Pessoa A. Assim, para cada indivíduo é "procurada" a sua família. A função utilizada tem como nome *EscreveFamilia*(codigoDeEgo, relacaoPai, relacaoMae, contador, filenumber) e como argumentos o Código de Ego do indivíduo para o qual se vai escrever a família, o código que está definido no sistema de gestão para a relação com o pai (o código utilizado foi o

“F”), o código para a relação com a mãe (código “M”), um contador, que se trata de uma variável criada, já que os egos são identificados univocamente pelo seu código, coisa que não acontece com as famílias e dessa forma, existiu a necessidade de criar um código numérico que fica associada a cada família. Esse contador será incrementado sempre que uma nova família for criada com vista a que cada família fique com o seu código e que esse código represente apenas uma família. O *filenumber*, tal como descrito anteriormente, é o descritor do ficheiro de destino com a representação GEDCOM para que possa ser escrito para o ficheiro à medida que a informação é processada.

Esta função começa por verificar quem é o pai e quem é a mãe do código de ego dado. Caso detecte que não existe nem pai nem mãe definidos para esse indivíduo coloca os pais como desconhecidos.

Após encontrar o pai e a mãe do ego em estudo, é imediatamente escrito para o ficheiro com a representação GEDCOM a constituição da família com, pelo menos, esses três elementos (dois no caso de se conhecer apenas a mãe ou pai). No entanto, a constituição da família não pode ficar por aqui. As associações dos irmãos a esta família constituem um elemento muito importante depois na fase de análise, pelo que se não se garantir a integridade e a coerência nos dados podemos incorrer mais tarde em problemas de análise derivados deste mau planeamento e execução.

Assim, nesta fase do processamento é procurada a existência de irmãos (sexo masculino e feminino) em todo o sistema de gestão para que possa ser escrita essa informação ainda na família do código de ego em estudo. Após a escrita de toda a família é escrita então informação adicional para essa família, isto é, é escrito para o ficheiro com a representação GEDCOM as informações sobre a data de casamento dos pais dessa determinada família, as notas dessa relação e a respectiva cota.

De notar que nesta fase do processamento podem existir duas formas de “casar” duas pessoas. A associação entre duas pessoas é feita ou por serem progenitores de um indivíduo e ai nesse caso, não é necessário criar um novo registo para registar a relação

entre os progenitores, pode ser feita indiciando especificamente que um individuo tem como esposa uma outra pessoa. No entanto, é importante perceber que nem todos os filhos nascem de duas pessoas casadas. A forma como se ultrapassa esse assunto nesta fase do processamento é através da data de casamento. Caso o campo data de casamento esteja preenchido é sinal que se está perante um matrimónio. Se não tiver preenchido, regista-se a associação da família mas não se constitui matrimónio entre os dois progenitores. Quando se quer registar que dois indivíduos que tiveram filho(s) mas não se sabe a data de casamento, pode-se sempre colocar uma data aproximada ou remota já que temos a certeza que a data de casamento é sempre posterior à data de nascimento do descendente mais novo.

Depois de registadas todas as famílias com filhos, entramos na última parte do processamento da representação GEDCOM, registando dois elementos que se relacionam entre si mas que não existe mais nada que os possa relacionar (no caso anterior era o facto de serem ambos progenitores do mesmo ego). Desta forma, é efectuada uma pesquisa apenas dos indivíduos que estão no sistema de gestão que não tem registado nenhum filho e que se relacionam com outros indivíduos, apenas pelo registo de uma associação entre eles (código da relação = "W").

Na constituição desta família é apenas necessário ter em conta qual o código de ego que corresponde ao "Esposo" e qual o código de ego que corresponde à "Esposa". Novamente o registo da informação adicional para esta família de dois elementos é transportado para o ficheiro GEDCOM incluindo a informação da data de casamento, eventuais notas da relação e a respectiva cota.

3.4.2 PROCESSAMENTO - XML

Tal como foi referido durante esta dissertação o aproveitamento das ferramentas já existentes ao nível da análise de genealogias possibilitou alargar os horizontes ao nível da análise das próprias genealogias. Desta forma, foi possível canalizar energias na procura de novas representações passíveis de gerir e de analisar genealogias.

Durante o ponto Tecnologias desta dissertação pôde-se constatar que apesar de ainda não ser uma meta atingida, a integração do XML e da representação GEDCOM está muito próxima de ser oficial, estando assim perante novos paradigmas na gestão e análise de genealogias.

Sabendo da utilidade que o XML apresenta nos dias de hoje, aliada à importância que pode vir a ter no futuro, foi sempre com esse objectivo que se optou por realizar também uma parte que conseguisse fazer o processamento dos dados do sistema de gestão para uma representação XML da representação GEDCOM apelidada de GEDCOM XML.

Esta representação GEDCOM XML consiste, tal como definida anteriormente, numa tradução 1 para 1, isto é, a cada elemento GEDCOM corresponde na generalidade um elemento GEDCOM XML.

3.4.2.1 Processamento - XML – Fase 1 - Header

Tal como a fase do processamento do ficheiro GEDCOM também o processamento para a criação de um ficheiro XML com a representação GEDCOM necessita de um cabeçalho. Após a abertura do ficheiro para escrita, é definido e criado esse *header*, que tratando-se de um elemento estático, contém a versão XML utilizada no ficheiro e a codificação utilizada.

Um ponto importante nesta fase é a definição de uma codificação que suporte caracteres da Língua Portuguesa, isto é, acentuação. Desta forma, para que seja possível utilizar-se acentuação no ficheiro XML é necessário que a codificação seja: ISO-8859-1 (w3schools, 2008).

3.4.2.2 Processamento - XML – Fase 2 – Transformação GEDCOM - XML

Após efectuada a escrita do cabeçalho, entra-se na parte da transformação da representação GEDCOM na representação XML. Sabendo que cada linha da representação GEDCOM apresenta sempre o mesmo formato separado em três partes (nível | etiqueta | dados) é necessário arranjar uma forma de dividir uma linha da representação GEDCOM nessas três partes para que depois com essas três peças construir uma linha da representação XML.

Sabendo que o nível é sempre o primeiro elemento e que no máximo só pode ter dois dígitos é retirado efectuando uma procura na *string* com a linha da representação GEDCOM retirando os primeiros (um ou dois) caracteres e assim ficamos com o registo do nível.

Retirado o nível fica a sobrar um conjunto de caracteres com a informação da etiqueta mais à esquerda e o primeiro espaço em branco a contar da esquerda para a direita separa a etiqueta da informação relacionada com essa etiqueta. Como não é possível controlar os dados que estão associados a cada etiqueta a única coisa que se consegue controlar é que cada etiqueta é apenas uma palavra o que significa que até encontrarmos o primeiro espaço em branco a contar da esquerda para a direita, todos os caracteres que estão para trás fazem parte da etiqueta, passando a registar tudo o que está para a direita desse mesmo espaço a informação relacionada com essa etiqueta.

Retiradas todas as peças, está na altura de “montar o puzzle”. O mesmo é dizer que agora que temos a informação separada do nível, da etiqueta e dos dados em separado é possível começar a construir a representação XML.

Os níveis constituem um ponto importante e necessitam de ser preservados para uma correcta transformação para o XML. Enquanto na representação GEDCOM uma etiqueta fica definida apenas numa linha, o mesmo não acontece na representação XML. Isto é, a etiqueta de fim do nível 1, por exemplo, só pode fechar depois de ser escrita toda a

informação do nível 1. O mesmo quer dizer que caso esteja num nível 4, por exemplo, e pretenda passar novamente para a criação de um indivíduo é preciso primeiro fechar todos os níveis “pendentes” que fiquem para trás.

Tendo este elemento em consideração basta ter cautelas apenas em mais duas situações para que a transformação da representação GEDCOM seja efectuada para o formato XML. A primeira situação depara-se com os elementos textuais. Sabendo que no texto o utilizador é livre de escrever o que entender é preciso garantir que o utilizador pode inserir caracteres especiais para a representação XML. Um caso concreto é o caso dos sinais de maior e menor que no XML são utilizados para definir a abertura ou fecho de uma etiqueta. No entanto, o utilizador num campo como as notas gerais pode se sentir livre de escrever algo do tipo: “A idade deste Ego é > que o Ego com o código 123”. Desta forma, exige-se a criação de um mecanismo para garantir que não se altera nenhum dado que esteja na representação GEDCOM.

O mecanismo utilizado passa pela introdução antes do texto de cada etiqueta do elemento `<![CDATA[Texto a Escrever]]>` em que o *Texto a Escrever* é a informação proveniente do ficheiro GEDCOM que corresponde à terceira peça de cada linha da representação GEDCOM. Tal como definido anteriormente, o elemento *CDATA* permite ao utilizador que a informação que está presente nesse elemento se mantenha intacta e que qualquer *XML Parser*²⁰ a ignore e não efectue nenhum processamento sobre essa informação, mostrando apenas essa informação.

Por último a situação que também carece de especiais cuidados é o tratamento dos códigos de cada indivíduo (Códigos de Egos). Esta é uma parte fundamental de todo o trabalho já que permitirá posteriormente referenciar através do código, cada indivíduo, isto é, quando numa associação entre duas pessoas é utilizado somente o código, se no XML for bem definida a gestão dos códigos, é possível posteriormente, através do mesmo,

²⁰ Funcionalidade que permite converter um ficheiro XML para outros formatos podendo posteriormente ser manipulado por outros programas

aceder a toda a informação sobre esse indivíduo proveniente da representação GEDCOM, que por sua vez é proveniente do sistema de gestão de bases de dados.

A garantia é dada através da introdução do elemento *ID* na representação XML já que, tal como descrito no ponto Tecnologias, este elemento (*ID*) permite a identificação unívoca de que não existe mais nenhum elemento daquele tipo e que todos os elementos que apareçam com aquele código vão ficar associados ao mesmo elemento.

Este elemento é utilizado, quer com elementos do tipo INDI (definição de indivíduos) quer com a definição de famílias (FAM) sendo que o mesmo é aplicado a seguir a identificação da etiqueta (exemplo: INDI ID = "E1"). Para que seja possível depois fazer a referenciação correcta é necessário utilizar tal como vimos o ID na definição dos indivíduos mas na sua "aplicação", isto é, sempre que se pretender referir a um determinado indivíduo usa-se o elemento REF para que a partir daquele momento aquele conjunto de caracteres passe a estar associado ao respectivo código e podendo "transportar" consigo toda a informação relativa a um determinado indivíduo.

Apesar de o código criado para permitir esta funcionalidade ao Sistema Informático de Gestão e Análise de Genealogias não ser muito extenso, apresenta no entanto uma importância vital para ampliar e permitir novas abordagens quer na vertente da gestão, quer na vertente da análise de dados genealógicos.

A constante procura de uma solução simples, mas não mais simples do que isso, permitiu no final obter um código compacto, optimizado já que são vários os casos de recursividade implementada na criação desta funcionalidade.

3.4.3 PROCESSAMENTO – GOOGLE EARTH

Os dados genealógicos georreferenciados constituem uma mais-valia na análise de genealogias. Uma das portas que a representação XML proporcionou foi precisamente a integração com um Sistema de Informação Geográfico, como por exemplo o Google Earth.

Sabendo que o input de dados para o Google Earth podia ser feito através de uma representação específica de XML (a representação KML), procurou-se assim permitir um novo paradigma na análise de dados genealógicos, como também um caso de aplicação da representação XML que comprovava assim as suas características de elemento bastante versátil capaz de se adaptar a várias situações.

3.4.3.1 Processamento - Google Earth – Fase 1 - Header

No seguimento das outras vertentes do processamento, também neste ponto é necessária a criação de um cabeçalho. Após a permissão de escrita ser dada ao ficheiro de destino, ficheiro esse que ficará com a representação KML, explicado no ponto “Definição e Conceitos”, o cabeçalho é escrito para o ficheiro. Uma vez que esta representação é um derivado da representação XML, o cabeçalho é igual em ambos os processamentos. Assim, o cabeçalho contém a informação da versão do XML utilizada, e ainda a informação sobre o verificador KML para que o ficheiro respeite as normas de criação de ficheiros em KML e por sua vez o Google Earth conseguir interpretar essa informação.

3.4.3.2 Processamento - Google Earth – Fase 2 – Criação KML

Depois de escrito o cabeçalho para o ficheiro KML, é descrito qual o estilo para as “marcas” apresentados no mapa do Google Earth, começando pela forma e terminando na coloração dos mesmos.

Depois de estar definido o estilo para cada “marca no mapa” é então colocada no ficheiro toda a informação pertencente aos dados a apresentar, isto é, será nesta etapa que os dados genealógicos serão transportados para o ficheiro KML.

Existem sempre três dados relativos aos dados genealógicos que, independentemente da opção escolhida para analisar no mapa, estão presentes, são eles: Código de Ego, Latitude e Longitude. A constante presença destes três elementos no ficheiro KML, seja relativo às datas de nascimento, às datas de falecimento, ou a qualquer outro assinto, é sempre feita já que o aparecimento das “marcas no mapa” do Google Earth tem como informação principal o código do Ego.

A razão do aparecimento dos três elementos justifica-se pelo facto da referenciação de cada local ser feita através da latitude e da longitude. Como tal a sua informação é imprescindível e necessária no ficheiro com a representação KML. O código de Ego, pela sua simplicidade ao nível da apresentação da informação foi a escolha para marcar cada “marca no mapa” do Google Earth.

A informação a detalhar cada local no mapa é feita consoante o caso específico de aplicação de análise, funcionando sempre este passo, na definição de local um a um, isto é, para cada local é definida a latitude, longitude, o nome a apresentar e a descrição dessa marca. Tal como vimos o nome a apresentar é sempre o código de ego associado a esse ponto e a descrição é um conjunto de informação que varia consoante o objecto de análise.

Toda a informação relativa quer à latitude, longitude ou aos dados sobre o individuo e/ou família são provenientes do sistema de gestão, sendo que após a criação do ficheiro com a representação, esse ficheiro passa a ser autónomo, isto é, a análise pode ser feita em qualquer computador já que tal como vimos um ficheiro XML (KML para este caso específico) pode não só funcionar como um sistema de análise mas também como um sistema de gestão já que transporta consigo toda a informação.

Após o carregamento de toda a informação para o ficheiro KML o ficheiro é fechado para escrita, sendo lançado automaticamente o Google Earth com a entrada de dados o ficheiro anteriormente criado.

3.5 EXPORTAÇÃO E PORTABILIDADE

Um dos principais objectivos deste sistema informático é utilizar os dados armazenados em múltiplas aplicações. Desta forma, a interoperabilidade de sistemas heterogéneos visa centralizar a atenção nesta aplicação, disponibilizando os mecanismos necessários para proceder à ligação entre as aplicações externas ao próprio sistema.

A implementação dos metadados neste Sistema Informático de Gestão e Análise de Genealogias é algo que se procurou garantir desde as suas primeiras fases. Conhecendo a importância extrema que os metadados podem apresentar em trabalhos deste tipo, a sua execução tinha de surgir, não só na componente descritiva, mas também na estrutural.

Na componente descritiva foi implementada ao nível da programação do próprio Sistema de Gestão e Análise de Genealogias, já que toda a aplicação pode ser parametrizada nos campos do sistema de gestão, isto é, é possível através dos metadados, transportar um sistema de gestão para outro, com nomes diferentes para os campos, bastando para isso apenas parametrizar todos os campos e respectivas tabelas. Esta medida visa constituir uma mais-valia deste sistema, pois sabe-se que as naturezas dos projectos são muito variáveis, assim como o próprio sistema de gestão de um projecto pode mudar ao longo do tempo de vida do próprio projecto, pelo que a concepção de aplicações capazes de se adaptarem às circunstâncias são necessárias e imprescindíveis.

A componente estrutural assume neste Sistema Informático uma plataforma de comunicação com outros sistemas que acredita-se ser um novo rumo na gestão e partilha de dados genealógicos. A capacidade de atingirmos uma representação capaz de descrever um documento electrónico nos quais o conteúdo e a descrição compõem um único ficheiro é uma meta que este Sistema Informático alcançou permitindo que essa estrutura possa

operar com linguagens como HTML por exemplo, operando com todas as funcionalidades que a própria linguagem HTML oferece, garantindo a interoperabilidade entre diferentes sistemas.

Esta componente é conseguida neste Sistema informático através da criação da representação XML, constitui um novo paradigma quer na gestão e análise de informação, mas igualmente na troca e partilha de conteúdos.

Neste Sistema Informático são assim suportadas as exportações dos dados armazenados no sistema de gestão para três representações já descrita no decurso desta dissertação: GEDCOM, XML e KML. A portabilidade é um elemento presente neste Sistema já que ao gerar uma representação GEDCOM para análise de genealogias não estamos a limitar de forma alguma, o utilizador na escolha do programa para analisar esses mesmos dados, isto porque, a representação GEDCOM é um formato de importação de dados da grande parte dos sistemas genealógicos existentes no mercado. Assim, ao possibilitar a sua geração, estamos não só a colaborar no sentido da partilha e troca de informação mas também a transportar para o utilizador os limites que ele pretende dar à aplicação. Assim, qualquer alteração no programa de análise de genealogias é assegurada sem qualquer tipo de problema, sendo que em raros casos seria apenas necessário mudar o cabeçalho do ficheiro com a representação GEDCOM.

A representação XML pela sua essência já está inserida na gama de tecnologias que permitem a partilha e interoperabilidade na troca de informações. Assim, procura de alguma forma arranjar novos métodos e soluções quer para a gestão, quer para a análise de dados genealógicos, permitindo ao utilizador novas soluções. Um caso concreto de uma possível aplicação prende-se com a representação: KML.

A representação KML pretende garantir aqui um papel muito importante numa outra vertente de análise de dados genealógicos, isto é, procura ser uma nova plataforma de análise já que toda a informação genealógica é georreferenciada, proporcionando ao utilizador novos universos de análise.

Qualquer uma das três representações são representações auto-suficientes, isto é, carregam consigo toda a informação necessária para que apenas com o ficheiro com a representação, seja tudo aquilo que é necessário para analisar a informação genealógica. Este facto visa constituir uma mais-valia já que qualquer pessoa pode fazer uma análise sem estar ligada ao sistema de gestão, podendo por exemplo, ser feita uma análise com a representação GEDCOM através de um ficheiro armazenado em qualquer dispositivo de armazenamento de pequeno porte e sem estar ligado à internet, nem ao sistema de gestão, conseguir contudo, visualizar toda a informação proveniente do sistema de gestão.

Ainda relacionado com a exportação a portabilidade, este Sistema Informático de Gestão e Análise de Genealogias permite ainda que a origem dos dados possa variar sem o utilizador se aperceber garantindo assim uma transparência ao nível dos serviços. Esta funcionalidade é conseguida através da implementação dos metadados que caso seja necessário mudar o sistema de gestão de uma tecnologia para outra, e por sua vez mudar o nome dos campos e das próprias tabelas a que estão associados, o sistema informático continua em actividade sendo apenas necessário parametrizar a aplicação, isto é, fazer referencia quais são agora as tabelas e os campos presentes no sistema de gestão.

Desta forma, e não descorando o objectivo de permitir ao utilizador o máximo de liberdade quer na escolha de análise quer na escolha de gestão, acredita-se estarem reunidas as condições para um uso fiável deste sistema e ao mesmo tempo não limitar a criatividade do utilizador, uma vez que na área da Genealogia limitar a capacidade do utilizador seria um passo atrás na utilização de plataformas electrónicas para a análise dos dados, isto porque o utilizador tem de ser o “dono” da análise que pretende fazer, e a função deste Sistema de Gestão e Análise é alargar novos horizontes ao utilizador não só na gestão mas essencialmente na análise de dados genealógicos.

CAPÍTULO IV - CASO DE ESTUDO

A ideia nuclear do projecto a que este Sistema Informático de Gestão e Análise de Genealogias foi aplicado, consiste em comparar as redes de comissários, a actuação e as afinidades/discrepâncias entre as duas principais instituições que procediam ao apuramento da honra em Portugal (o Santo Ofício e a Mesa da Consciência e ordens), quer a nível local, quer no centro político (Olival, 2008).

No que confere à genealogia, para satisfazer o objectivo principal deste projecto é então imprescindível conhecer o máximo de elementos possíveis de forma a averiguar o maior número de genealogias já que estas duas entidades faziam inquéritos nos locais de nascimento e moradas dos pretendentes, através de uma rede de comissários montada no último quartel do século XVI pelo Santo Ofício.

A data de 1773 impôs-se como limite por corresponder ao fim oficial da limpeza de sangue, o alicerce fundamental da honra durante o Antigo regime na Península (Olival, 2008).

Após a investigação prosopográfica por parte dos historiadores, e para que seja possível efectuar-se análises da própria rede de comissários, quer numa escala visando o individuo e a respectiva família, quer numa escala macroestrutural, procurando respostas para toda uma sociedade.

Por se tratar de um projecto que coabita com duas instituições que em toda a Península Ibérica tinham maior crédito em matéria de avaliação da honra individual e das parentelas, é assim de interesse máximo aplicar este Sistema Informático de Gestão e Análise, a um projecto com objectivos claros, a necessitar de respostas, após as perguntas emergentes.

4.1 DESCRIÇÃO DO ESTUDO

Com o intuito de atingir os objectivos estabelecidos no âmbito deste estudo fixou-se uma metodologia de trabalho que se traduz nas seguintes fases:

1. Conhecimento do projecto FCOMP-01-0124-FEDER-007360 nas diferentes áreas do mesmo.
2. Levantamento das necessidades para a gestão e análise da Genealogia através de reuniões com os membros do projecto.
3. Criação do sistema de gestão de dados genealógicos.
4. Carregamento de dados antigos para o novo sistema de gestão de genealogias para áreas temporárias.
5. Transformação dos dados importados para as áreas temporárias de forma a respeitar as regras do novo sistema de gestão e análise de dados genealógicos.
6. Exportação dos dados transformados para a base de trabalho real.
7. Criação de uma plataforma de conversão dos dados no sistema de gestão para um formato de análise de dados genealógicos.
8. Análise de dados genealógicos referentes à rede do Santo Ofício e das Ordens Militares (1570-1773).
9. Análise de dados genealógicos georreferenciados, utilizando um Sistema de Informação Geográfico.
10. Transformação do formato de análise de dados genealógicos para um novo formato de análise integrando a tecnologia XML.

4.2 RECOLHA E TRATAMENTO DE DADOS: BASE DE DADOS PROSOPOGRÁFICA

Todo o objecto de estudo se desenvolveu a partir da recolha de informação, feita por outros elementos do projecto, através da análise documental realizada no Arquivo Nacional da Torre do Tombo. Sendo este arquivo, um arquivo público, funciona como arquivo central do Estado Português e constitui-se numa moderna instituição aberta a pesquisadores e publico em geral (Portal Oficial ANTT, s/d).

Não é objectivo do Arquivo Nacional da Torre do Tombo efectuar, enquanto organismo, pesquisas nos fundos documentais para responder a investigadores ou entidades, mas sim proporcionar a assistência a utilizadores no acesso aos arquivos e possibilitar a investigação dos mesmos (Portal Oficial ANTT, s/d).

Constitui-se assim o primeiro passo na recolha da informação. Vários historiadores encontram-se reunidos no Arquivo Nacional da Torre do Tombo para recolher a informação genealógica e por sua vez introduzi-la no sistema informático criado para este efeito. O instrumento para a recolha e tratamento desses dados foi o sistema de gestão de dados genealógicos, desenvolvido e que irá ser detalhado ainda neste ponto. Sabendo o propósito de efectuar um estudo sobre a caracterização individual de cada indivíduo e a análise de genealogias dos dados recolhidos, define-se este sistema de gestão como uma base de dados prosopográfica (Jorge, Maria, Vilar, Henriques & Lopes, 2004), seguindo-se uma descrição deste sistema de gestão.

Base de dados prosopográfica

Em primeiro lugar é importante salientar que o desenho da estruturação de um sistema de gestão de genealogias é o primeiro passo importante na construção deste tipo de sistemas (Sousa, 2002) já que um mau planeamento pode originar um mau desempenho no próprio sistema de gestão (Caldeira, 2008; Loureiro, 2008).

No entanto, nem todos os criadores de sistemas de gestão de genealogias idealizam o mesmo desenho da estruturação destes sistemas, originando assim diferentes perspectivas e abordagens para a mesma problemática (Groh, Stockman, Powell, Prague, Irwin & Reardon, 2007). Desta forma, é possível separar as abordagens dos autores em dois grupos dissemelhantes: os que analisam e gerem a genealogia de indivíduos, detalhando para cada indivíduo, todas as suas relações até uma determinada profundidade na árvore genealógica (pais, avós, bisavós, etc.) e os autores que determinam que o cerne do problema da genealogia centra-se apenas na construção de famílias, isto é, para cada indivíduo é necessário apenas registar, os seus progenitores, e respectivos irmãos que apresentem o mesmo pai e mesma mãe do indivíduo em causa, sendo que, todo o cálculo da genealogia do indivíduo é efectuado posteriormente através de softwares de análise de dados genealógicos (Caldeira, 2007).

A escolha da abordagem a adoptar é um passo extremamente importante e decisivo para uma boa gestão e futura análise dos dados genealógicos (Damas, 2005). Assim, foi necessário efectuar um estudo introdutório sobre quais as reais necessidades para este projecto e a partir desse ponto, tomar uma decisão sobre a abordagem a perfilhar.

Após várias reuniões com os elementos do projecto, verificou-se que interessava essencialmente o estudo individual de cada ego e a sua genealogia até à quarta geração de ascendentes. No entanto, tal como referido por várias vezes nesta dissertação, não é aconselhável nem seria uma atitude recomendada, construir a aplicação fixada e limitada até à quarta geração de ascendentes e/ou descendentes. Assim, procurou-se uma solução que adoptasse um processo recursivo, para que a limitação passasse para o lado do utilizador, não impondo por isso limites no “avanço” ou “recuo” na análise de uma árvore genealógica. Esta solução adoptada apresenta como principal justificação, a possível expansão que este sistema informático pode vir a apresentar e como tal, desenvolver um sistema apenas e só a pensar nos problemas actuais, não é de todo aconselhável, porque a qualquer sugestão do utilizador, será necessário reprogramar toda a aplicação, pelo que se for possível deixar uma margem de progressão da própria aplicação, é de todo recomendado. A escolha do sistema de numeração foi baseada no Sistema de Numeração de Genealogias Dulluis/Stemmer que se trata de identificação de letras associados a cada

relação (Dullius, 2004). No entanto, foram efectuadas algumas alterações, que serão explicadas ainda no decorrer deste ponto.

Identificação de Indivíduos – Tabela Ego (1)

Neste ponto de caracterização de indivíduos é necessário registar um conjunto de informação relacionada com o individuo, que tal como o nome indicia, permite identificar, de forma detalhada, cada individuo analisado. Assim, são considerados os seguintes campos: Código de Ego (1.1), nome do Ego (1.2), o seu sexo (1.3), local onde nasceu (1.4), a sua data de nascimento (1.5), local onde faleceu (1.6), a data do seu falecimento (1.7) e a sua ocupação (1.8). São ainda considerados os campos: Notas relacionados com o Ego (1.9), a Cota (1.10), a Cota de Falecimento (1.11) e a Cota da Ocupação (1.12).

O Código de Ego trata-se de um campo numérico, que começa no número um, e caracteriza de forma única e inequívoca cada indivíduo, sendo por isso a chave primária desta tabela Ego (Sousa, 2002).

O nome do Ego trata-se de um campo do tipo texto no qual o nome de cada indivíduo é registado. Este campo assume grande importância neste tabela pois permite identificar de uma forma bastante rápida cada indivíduo através do nome e na fase de análise efectuar uma distinção mais rápida entre indivíduos através do nome.

O campo sexo é utilizado para registar o sexo de cada indivíduo, Masculino ou Feminino, tratando-se também de um campo importante já que permite realizar algumas regras de consistência na área da genealogia assim como mostrar dados estatísticos relevantes através deste campo.

O local de nascimento é considerado um campo com uma importância muito elevada, assim como o local de falecimento, já que são campos que estão ligados à tabela Localidade (detalhada no decurso deste tópico), sendo apenas códigos numéricos que transportam consigo toda a caracterização de um determinado lugar, apenas pelo código numérico,

constituindo assim uma chave estrangeira da tabela Ego (Damas, 2005; Loureiro, 2008). O facto de estes campos se encontrarem ligados à tabela Localidade permite, ao nível da análise de dados, efectuar-se uma análise muito mais detalhada e rigorosa, permitindo inclusive a integração com Sistemas de Informação Geográfica através dos campos da tabela Localidade, latitude e longitude.

As datas de nascimento e falecimento retratam a data de nascimento do ego em estudo e a data da sua morte, respectivamente. Estes campos são do tipo texto, pelo facto de muitas delas não serem datas exactas, sendo necessário a colocação de símbolos para a estimativa de datas.

Desta forma a colocação de um sinal de igual (“ = ”) numa data indica a certeza sobre o elemento anterior à colocação do sinal. Por seu lado, a colocação de um sinal de menor (“ < ”) numa data indica incerteza sobre o elemento anterior à colocação do sinal, referindo que apesar da incerteza, trata-se de uma data anterior à mesma. Assim, a colocação de um sinal de maior (“ > ”) numa data indica igualmente incerteza sobre o elemento anterior à colocação do sinal, referindo-se este a uma data posterior à indicada. Por último, a colocação do sinal de menor (“ - ”) numa data indica muita incerteza sobre o elemento anterior à colocação do sinal, referindo tratar-se de uma data vagamente aproximada (Cobra, 2001).

Segue-se apenas alguns exemplos para melhor compreensão deste problema: 1700=02=01, corresponde exactamente, ao dia 1 de Fevereiro de 1700; 1700<02<01, corresponde a uma data anterior ao dia 1 de Fevereiro de 1700; 1700<00<00, corresponde a uma data antes do ano de 1700; 1700>02>01, corresponde a uma data depois do dia 1 de Fevereiro de 1700; 1700=00<00, sabe-se que foi no ano de 1700 mas desconhece-se o mês e o dia e por último, a data de 1700-00-00 trata-se de uma data vagamente aproximada em torno de 1700.

Mesmo apesar de apresentar um elevado grau de incerteza em algumas datas, é importante a nível histórico, registar um referente cronológico, mesmo que seja vago ou aproximado em vez da sua não existência (Cobra, 2001).

Relativamente à ocupação, optou-se pelo seu registo já que permite no futuro analisar gerações de famílias e verificar se a ocupação se mantinha ao longo de gerações. É igualmente importante, pelo facto existirem várias pessoas com o mesmo nome, mas com ocupações diferentes, pelo que este campo também apresenta essa vantagem, efectuar a despistagem de Egos repetidos.

Sabendo que por vezes é necessário registar alguma informação extra sobre cada indivíduo, justifica-se a criação do campo Notas, permitindo assim adicionar algum detalhe ou comentário que possa vir a ser relevante mais tarde.

Por ultimo, o campo Cota, Cota de Falecimento e Cota de Ocupação são campos que se justificam numa abordagem mais profissional da gestão e análise de genealogias já que permitem assim o registo de onde a informação foi retirada para que caso se pretenda voltar a verificar a informação, por variadíssimas razões, sabe-se a referência de onde essa informação foi retirada.

Identificação do Tipo de Relações entre Egos – Tabela Relação de Parentesco (2)

Depois de criados os egos no sistema de gestão e para que seja possível atribuir relações entre os mesmos é necessário saber quais são as relações disponíveis pelo sistema de gestão para que a atribuição seja feita de forma coerente e correcta. Assim, a tabela Relação de Parentesco conta com os seguintes campos: Relação (2.1), descrição da relação em Língua Portuguesa (2.2) e a descrição da relação em Inglês (2.3).

O campo Relação consiste num campo do tipo texto, que suporta apenas um carácter no qual identifica de forma única e inequívoca cada indivíduo, sendo por isso a chave primária desta tabela Relação de Parentesco. Por seu lado, os campos descrição da relação

em Língua Portuguesa e em Inglês, registam a descrição da relação em causa em Português e Inglês respectivamente.

A escolha das relações a disponibilizar resulta da escolha da abordagem centrada em famílias. Assim, é apenas possível registar para cada indivíduo, as relações: Pai, Mãe, Irmão, Irmã, Esposa, Tia desconhecida, Tio desconhecido, Primo desconhecido e Parente desconhecido.

A não opção pela total definição das relações até à quarta geração, como era pedido, foi pelo simples facto de ser uma complexidade brutal para o utilizador na hora de introduzir e posteriormente na hora de analisar, definindo nomenclaturas e códigos para definir um trisavô paterno. Assim, a solução passa por uma solução recursiva, isto é, a definição é feita recursivamente pelo que posteriormente, o sistema de análise é que tem de apresentar a capacidade para “juntar as peças do puzzle”. Quer isto dizer que será o sistema de análise o responsável por efectuar a associação dos pais, avós, bisavós e assim sucessivamente, não apresentando por isso um limite estabelecido no próprio software.

Das relações possíveis, apresentam um especial destaque a existência apenas da relação Mulher, permitindo assim a inexistência de um mais um código que na prática teria o mesmo significado do que este (Esposa). Desta forma, sempre que se quiser registar um casamento, regista-se que um determinado homem casou com uma determinada mulher através da utilização da relação Esposa. As relações de parentesco desconhecidas constituem uma mais-valia no sistema de gestão já que permite assim, mesmo que de uma forma meramente informativa, registar parentesco que à partida desconhecemos a sua origem, como é o caso da tia desconhecida, do tio desconhecido e do primo desconhecido.

Por último, ainda dentro do assunto dos parentes desconhecidos, é possível utilizar o código de Parente desconhecido sempre que for necessário registar uma associação entre duas pessoas mas que não sabemos qual o grau. É importante salientar, que este tipo de associação de graus de parentesco desconhecido, tem como objectivo primordial a adição de informação extra, não podendo ser por isso considerados como elementos a utilizar no

cálculo e desenho de genealogias, relegando-os para o plano da informação extra de cada indivíduo, num patamar igual aos campos Notas e cotas, como explicado anteriormente.

Identificação das Relações entre Egos – Tabela Genealogia (3)

Para que seja possível registar a associação entre dois egos é necessário registar um conjunto de informação que se resume aos seguintes campos: Referência (2.1), Código do Ego 1 (2.2), Relação entre Egos (2.3), Código do Ego 2 (2.4), Data de Casamento (2.5), Local de Casamento (2.6), Notas sobre a relação (2.7) e a Cota da informação sobre a relação (2.8).

O campo Referência consiste num campo do tipo numérico, apenas necessário para garantir a integridade das relações no sistema de gestão já que identifica de forma única e inequívoca cada relação entre dois indivíduos, sendo por isso a chave primária desta tabela Genealogia. Este campo assume assim um papel importante, na despistagem de duplicações de relações, já que é possível, na fase de análise, verificar se a mesma relação foi inserida de forma duplicada.

Os códigos de Ego 1 e 2 são campos com uma importância extrema, já que são estes que vão ditar entre que pessoas, vai ser feita uma determinada associação. Para tal, estão ambos os códigos ligados à tabela Ego para que seja possível, apenas através do código, conhecer toda a informação de cada indivíduo. Os códigos de Ego 1 e 2 são campos numéricos, que pelo facto de ser chave primária na tabela Ego, são na tabela Genealogia, chave estrangeira.

O campo Relação entre Egos é tal como os códigos anteriores, um campo de elevado interesse e utilidade pois é neste campo que é caracterizada e especificada a relação que refere dois indivíduos. Para que seja possível manter a integridade na colocação dos tipos de relações é necessário colocar este campo como chave estrangeira da tabela Genealogia, permitindo assim que os tipos de relações sejam apenas os acima descritos (Sousa, 2002).

A data de casamento retrata a data em que dois egos casaram. Este campo está preenchido apenas para as relações que indicam que existiu um casamento, que pode ser o caso da relação Pai, Mãe e Esposa. Este campo é do tipo texto, tal como todas as datas deste sistema de gestão, pelo facto de também a data de casamento não ser uma data exacta.

O local de casamento é considerado um campo com uma importância muito elevada, assim como todos os campos denominados por locais, no sistema de gestão, já que, tal como explicado anteriormente, trata-se de códigos numéricos que transportam consigo toda a caracterização de um determinado lugar, constituindo assim uma chave estrangeira da tabela Genealogia o que permite realizar operações de análise mais avançadas tal como referido previamente.

A utilização de um campo Notas justifica-se com a possibilidade de inserção de informação extra sobre a relação que mereça particular destaque e que seja importante a sua preservação.

Por último, o campo Cota é um campo do tipo texto onde se regista a fonte de informação registada.

Identificação de Localidades – Tabela Localidade (4)

A tabela Localidade é criada uma vez que esta se trata de uma dimensão de análise de informação necessária para a compreensão de relações prosopográficas. Desta forma, os campos registados para cada lugar são os seguintes: Referência (4.1), Localidade (4.2), Província (4.3), Bispado (4.4), Concelho (4.5), Comarca (4.6), Freguesia (4.7), Termo (4.8), Latitude (4.9), Longitude (4.10), Notas (4.11) e Cota (4.12).

O campo Referência consiste num campo do tipo numérico, apenas necessário para garantir que cada localidade é identificada de forma única, sendo por isso a chave primária desta tabela Localidade.

O campo Localidade indica o lugar para o qual está a ser registada toda a informação. De notar que será este o campo que ficará associado aos locais de nascimento, de falecimento e de casamento, aquando da transformação para uma representação GEDCOM, já que a mesma implica a escrita de um local exacto, sendo este o melhor classificado para assumir tal posição.

A província, bispado, concelho, comarca, freguesia e termo, são todos elementos caracterizadores de uma localidade, mas que neste estudo apresenta uma importância diminuta. No entanto, tal como até aqui defendi, a não implementação de sistemas rígidos em benefício de sistemas flexíveis, maleáveis e que possam vir a integrar outros conteúdos, é essencialmente esse o fundamento para registar tais informações.

Os campos Latitude e Longitude registam coordenadas geográficas que se caracterizam pelo facto de serem únicas para cada local. Estes campos constituem-se como muito importantes já que serão eles, os responsáveis pela interligação dos locais com os sistemas de informação geográficos (Filho & Saraiva, 2010).

A utilização de um campo Notas surge pela possibilidade de inserção de informação extra relativa a cada localidade enquanto, o campo Cota é um campo do tipo texto onde se regista o local de onde a informação relacionada com esta localidade foi retirada.

4.3 EXPLORAÇÃO DE DADOS

A exploração e análise de dados apresentam como objectivo primordial, possibilitar ao utilizador, um maior proveito da informação armazenada num determinado sistema de gestão. Quando um determinado utilizador insere os dados no sistema de gestão, independentemente do seu tipo, é muito difícil para o utilizador efectuar a análise desses mesmos dados, isto porque, geralmente é necessário remontar a registos já inseridos que provavelmente o utilizador não consegue recordar.

Desta forma, a exploração dos dados, surge com o culminar da informação inserida no sistema de gestão, possibilitando assim maximizar o poder da informação, quer seja através de previsão de tendências e comportamentos futuros, quer seja através da descoberta de novos padrões de conhecimento adquiridos com base na informação introduzida.

Neste projecto, a exploração de dados surge como o produto final, onde se poderá visualizar e analisar toda a informação inserida, podendo detectar-se erros, ou por outro lado, responder aos objectivos traçados.

A exploração dos dados genealógicos pode ser feita através de diversos programas de estudo de genealogias, sendo no entanto o escolhido, o Personal Ancestral File (PAF) versão 5.2.18.0.

No entanto, antes de detalhar este programa de análise de genealogias é importante perceber como é feita a passagem do Sistema de Gestão para o Sistema de Análise.

Toda a mensagem trocada entre dois interlocutores carece de uma codificação, independentemente do seu tipo (Nunes, 2008). O mesmo acontece na comunicação entre a Gestão (Base de dados) e a Análise (PAF) do Sistema Informático implementado. A codificação utilizada foi a representação GEDCOM que se trata de um acrónimo para ***GE**nealogical **D**ata **CO**munication* tal como definido no Ponto Definição e Conceitos.

Depois de importados os dados para o software genealógico Personal Ancestral File, surge a primeira manifestação de todo o esforço desenvolvido, isto é, surge todo o manancial de informação que estava na base de dados de uma forma já ao estilo genealógico, organizada por famílias, por gerações ou se preferível de forma individual.

Neste programa são várias as hipóteses de análise, podendo ser separada em três secções diferentes. Começando pela parte da família (separador *Family* no PAF) é possível visualizar a informação “principal” sobre uma determinada família, isto é, após escolher um elemento, é apresentada a informação neste separador para os progenitores, filhos (se existir) e respectiva esposa ou esposo, no caso de estes existirem. Surge ainda como informação primária, isto é, toda a informação que surge ao utilizador de forma automática sem ele recorrer a nenhuma opção, para cada indivíduo a data de nascimento, de falecimento, local de nascimento e falecimento e ainda o local e data de casamento, se existir.

No separador seguinte, *Pedigree*, é então proporcionado ao utilizador visualizar a árvore genealógica do indivíduo em estudo, sendo mostrada sem realizar nenhuma operação, até à quarta geração (trisavós e trisavôs). No entanto, através deste programa podemos analisar até a nonagésima nona geração. O elemento mais à esquerda é o indivíduo escolhido, e toda a árvore para a direita é a definição dos pais, avós, bisavós, e assim sucessivamente.

Outra possibilidade de análise é o terceiro e último separador do PAF, *Individual*, que mostra a lista de todos os indivíduos carregados para o sistema. Na lista surgem os nomes dos elementos, género, local de nascimento e falecimento, data de nascimento e falecimento, podendo estas opções serem parametrizadas e escolhendo assim quais os campos de maior relevância para serem dispostos nesse lista de indivíduos.

Em qualquer dos três separadores é sempre possível aceder a toda a informação relativa ao próprio indivíduo, isto porque, se o programa mostrasse, por exemplo na árvore genealógica, toda a informação de cada pessoa ficaria um ecrã bastante confuso para o utilizador e desta forma, em todos os separadores é possível visualizar mais importante bastando apenas efectuar duplo clique sobre o nome do indivíduo que se pretende obter mais informações e visualizar essa mesma informação.

Ainda relativamente ao PAF é possível utilizar um conjunto de funcionalidades que este programa oferece ao utilizador, nomeadamente, a capacidade de analisar possíveis erros que existam nos dados que estamos a analisar. Quer isto dizer, que o programa irá analisar todos os dados importados do sistema de gestão e irá procurar situações do tipo: a data de nascimento é posterior à data de falecimento; a definição de um elemento do sexo feminino como pai e vice-versa; entre outros exemplos. Note-se que por ser uma função de lista de possíveis erros, pode acontecer em algumas situações não constituir um erro mas sim uma situação real mas que dadas as suas particularidades de excepção pode o programa assumi-la como um possível erro. Um exemplo concreto trata-se na situação em que um indivíduo é progenitor de outro, em que a idade do progenitor é inferior a 16 anos de idade. Para o programa isto constitui um alerta, já que o utilizador pode ter introduzido de forma incorrecta os dados para análise, ou então pode constituir um facto real pelo que deve ser ignorado pelo utilizador.

Contudo, a análise dos dados genealógicos não se fica por aqui. Com o objectivo de analisar os dados de forma cada vez mais diferenciada, optou-se por uma abordagem diferente: a genealogia georreferenciada, isto é, ser possível visualizar através de mapas e cartografias, as relações dos indivíduos, locais de nascimento, locais de falecimento entre outros tipos de análises relacionadas com as genealogias.

Para ser possível esta integração entre estes dois mundos, Genealogia e Georreferências, é necessário existir um Sistema de Informação Geográfico e uma linguagem que sirva de comunicação entre a Genealogia e o Sistema de Informação Geográfico.

O Sistema de Informação Geográfico escolhido foi o Google Earth® e tal como visto, trata-se de uma aplicação interactiva cujos mapas são compostos por imagens satélites que permite ao utilizador, de forma interactiva, navegar por todo o globo analisando diversas variáveis disponíveis nesta aplicação.

Ao mudar de programa de análise, muda-se igualmente de paradigma na análise de dados genealógicos. Se até aqui, o objectivo era analisar as genealogias de famílias, verificar a

ligação entre indivíduos e qual a relação entre eles, verificar a própria consistência dos dados introduzidos, quando se chega a esta parte, análise de informação utilizando o Google Earth o próprio universo de análise muda, já que as circunstâncias assim o exigem. Neste programa, o utilizador pode visualizar no mapa-mundo, os dados analisados. A possibilidade de visualizar no mapa os locais de nascimento, ou de falecimento de indivíduos procura assim alargar o leque de possibilidades de análise do utilizador no universo da Genealogia.

Desta forma, o utilizador começa a perceber que pode explorar novas formas de analisar o manancial de dados genealógicos que dispõe e, em vez de se restringir ao conjunto de árvores genealógicas, às informações familiares e individuais, pode começar a querer procurar respostas num patamar de abstracção mais elevado, isto é, procurar analisar a sociedade como um todo, por exemplo, verificar a distribuição de ocupações pelo território nacional e chegar à conclusão sobre se existe alguma região com determinado predominância. Estes patamares de análise até aqui não eram explorados já que os programas genealógicos não estavam preparados para responder a estas necessidades.

Assim, é possível analisar qualquer informação desde que esta esteja referenciada a um local, ou ponto de referência. O mesmo quer dizer que independentemente do conteúdo, desde que exista uma latitude e uma longitude para especificar um local, é possível colocar e associar a esse ponto, qualquer informação pretendida, sendo esta de um indivíduo, de uma família, ou de toda a sociedade.

4.4 PORTABILIDADE

O Sistema Informático de Gestão e Análise de Dados Genealógicos quando aplicado a este caso de estudo, proporcionou elementos de portabilidade interessantes, desde logo a análise de informação sem estar directamente ligado ao servidor onde estava guardada toda a informação. Esta medida proporcionou que elementos do projecto estivessem a milhares de quilómetros de distância do servidor, sem recurso à Internet, e mesmo assim analisar a informação que carregaram consigo e que se encontra armazenada no disco rígido do computador utilizado para analisar dados ou num suporte amovível de dados.

Desta forma, a única particularidade necessária é ter em sua posse o ficheiro com a respectiva representação para que seja possível analisá-lo mesmo estando isolado de tudo o resto.

Para minimizar e ao mesmo tempo otimizar o servidor, a geração das representações GEDCOM, XML e KML, são feitas por um grupo restrito de utilizadores, colocando estes, as representações geradas no portal oficial do projecto²¹ tal como mostra a Figura 12, tendo os restantes elementos que efectuar a transferência dos ficheiros para os respectivos computadores. Outra possibilidade é o ficheiro seguir em formato físico, seja em suporte amovível (USB Flash Drive) ou em CD-ROM e/ou DVD-ROM.

Desta forma pretende-se assim garantir que os utilizadores são o centro da gestão e da análise de informação e não a própria aplicação ou sistema informático. Acredito que esta mudança de paradigma no funcionamento de aplicações informáticas é o correcto sendo este tipo de gestão e análise (de genealogias) um caso muito específico com características muito próprias que necessitam assim, de sistemas adequados a todas essas necessidades.

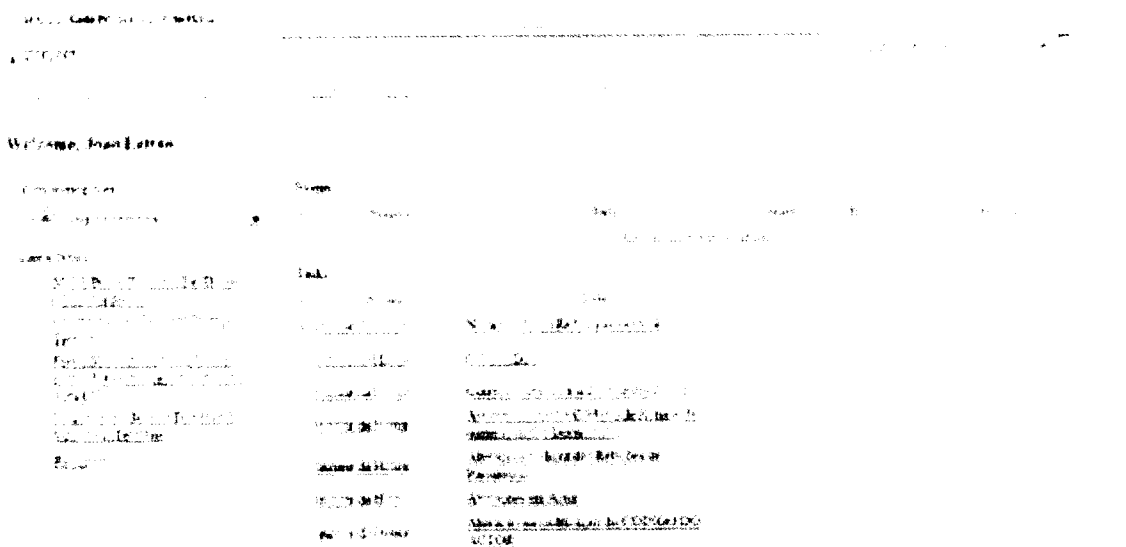


Figura 12 - Portal dos membros do projecto SPARES – Inquirir da Honra

²¹ http://www.santooficio.uevora.pt/index.php?option=com_user&view=login&Itemid=7

CAPÍTULO V - APRESENTAÇÃO DE RESULTADOS

Tenho consciência de que este trabalho, como toda obra humana, não é perfeito. No entanto, estou certo de ter contribuído para aumentar o leque de soluções genealógicas e proporcionar aos utilizadores a análise de diversas gerações e visualizarem em formatos amigáveis todo o seu trabalho de historiador. Tal como definido em tópicos próprios, os dados em estudo remontam à época entre 1570 e 1773, estando directamente relacionados com a Genealogia da Rede de Comissários do Santo Ofício e das Ordens Militares. Por se tratar de uma época remota, a apresentação dos dados e respectiva análise é em consonância com este facto, sendo que, existe certamente alterações quando comparando, por exemplo, o espaço que na época pertencia a Portugal, com os limites actuais do território luso.

A apresentação e discussão de resultados impõem como data limite o dia 01 de Setembro de 2010 por se tratar da data mais actual possível de modo a que seja possível a entrega da corrente dissertação no prazo legal. Assim, o resultado geral da aplicação do Sistema Informático de Gestão e Análise de Genealogias consiste na definição de 4886 indivíduos, com 4682 associações entre eles. Do total de indivíduos, 2208 consistem em indivíduos do sexo feminino sendo os restantes 2678 indivíduos do sexo masculino. Com 2268 casamentos definidos é possível verificar nos dados do sistema de gestão que nesta fase já apresentam um grau de coerência muito elevada algo que não acontecia nas primeiras versões. De notar que com a aplicação deste sistema informático foi possível melhorar substancialmente o sistema inicial que estava repleto de erros ao nível da coerência da informação desde a múltipla definição de progenitores para o mesmo indivíduo, passando por associações de indivíduo sexo masculino como “esposas” nas relações e vice-versa.

Por se tratar de um projecto em concreto com objectivos na área da História, são os historiadores membros do projecto responsáveis pelas conclusões dos dados apresentados já que são os mesmos que conhecem ao pormenor toda a vida de cada ego. Como tal, este espaço limita-se apenas à apresentação de resultados de forma a mostrar como é possível o Sistema Informático de Gestão e Análises de Genealogias ajudar os historiadores a obterem respostas às suas perguntas, encontrando também o caso de muitas vezes, ser a própria aplicação a suscitar novas dúvidas face ao que é electronicamente possível fazer através deste sistema.

5.1 O SISTEMA INFORMÁTICO DE GESTÃO E ANÁLISE DE GENEALOGIAS APLICADO À GENEALOGIA DO SANTO OFÍCIO E DAS ORDENS MILITARES (1570-1773)

Neste ponto será o resultado do trabalho desenvolvido aplicado ao Caso de Estudo em concreto. Este ponto será dividido em duas áreas; uma mais relacionada com a área da gestão dos dados e outro relacionado com a análise e portabilidade.

Esta separação também é feita logo no momento em que o utilizador inicia este sistema informático. Assim, é permitida ao utilizador a escolha por uma das vertentes: a da gestão de dados genealógicos ou a sua análise tal como demonstra a Figura 13.



Figura 13 - Menu Inicial do Sistema Informático de Gestão e Análises de Genealogias

No que diz respeito à gestão dos dados, a preferência foi sempre de encontrar a uma solução mais funcional do que propriamente estética, sem nunca descorar os princípios bases da construção de uma ferramenta amigável ao utilizador.

Assim o utilizador caso pretenda efectuar uma gestão dos dados genealógicos, introduzindo ou alterando dados já existentes depara-se numa primeira fase com o

seguinte menu que demonstra as possibilidades que o mesmo tem na área da gestão tal como demonstra a Figura 14.



Figura 14 - Menu de escolha das opções possíveis para a gestão de genealogias

Caso o utilizador pretenda efectuar uma gestão dos Egos será reportado para uma matriz de entrada de dados que permite ao utilizador introduzir de forma rápida uma grande quantidade de dados biográficos.

O utilizador pode a qualquer momento alterar os dados já existentes, assim como remover e inserir um novo indivíduo. A Figura 15 exemplifica este processo de gestão dos egos.

Ego: entrada múltipla									
Id	Código de Ego	Nome	Sexo	Data de Nascimento	Localidade de Nascimento	Fonte Nascimento	Data de Falecimento	Localidade de Falecimento	Fonte Falecimento
	1	Bartolomeu César de Andrade	M	1671+01-26	5	TT, HSO, Bartolomeu, Mç 3, doc. 52, 11	1742+02-00	2	
E	2	Manuel de Castro	M	1637+00+00	2	TT, HSO, Bartolomeu, Mç 3, doc. 62		4	
E	3	Maria César (Dona)	F	1637+00+00	2	TT, HSO, Bartolomeu, Mç 3, doc. 62		4	
E	4	Gonçalo Gomes de Castro	M	1617+00+00	7	TT, HSO, Bartolomeu, Mç 3, doc. 62		4	
E	5	Antónia Pais	F	1617+00+00	7	TT, HSO, Bartolomeu, Mç 3, doc. 62		4	
E	6	Melchior Berenguer de Andrade	M	1617+00+00	7	TT, HSO, Bartolomeu, Mç 3, doc. 62		4	
E	7	Mana da Ponte (Dona)	F	1617+00+00	7	TT, HSO, Bartolomeu, Mç 3, doc. 62		4	
E	8	Francisco de Castro Berenguer	M	1670+00+00	2	TT, HSO, Bartolomeu, Mç 3, doc. 62		4	
E	9	Joana Noronha e Esmeraldo (Dona)	F	1670+00+00	7	TT, HSO, Bartolomeu, Mç 3, doc. 62		4	
E	10	Manuel de Faria de Abreu (Padre)	M	1695-05-28	6	TT, HSO, Manuel, Mç 126, doc. 2228	1	4	
E	11	Pedro de Faria de Abreu	M	1690-02-18	6	TT, HSO, Manuel, Mç 126, doc. 2228	1	4	
E	12	Manana da Conceição	F	1690-03-25	5	TT, HSO, Pedro, Mç 14, doc. 300, 52	1704+07+16	7	TT, HSO, Pedro,
E	13	Josefa Mana de Bethencourt e Noronha	F	1689+00+00	7	TT, L.L.108, 357v		4	
E	14	Vitoriano de Vasconcelos e Bethencourt	M	1667+00+00	2	TT, HSO, Manuel, Mç 126, doc. 2228	1	4	
E	15	Pedro de Faria	M	1660+00+00	4	TT, HSO, Manuel, Mç 126, doc. 2228	2	1742+00+00	4
E	16	Mana das Neves	F	1658+08+07	6	TT, HSO, Manuel, Mç 126, doc. 2228	2	1742+00+00	4

Figura 15 - Gestão de Egos no Sistema Informático de Gestão e Análises de Genealogias

No entanto, o utilizador pode preferir efectuar a gestão das genealogias e aqui será reportado para outro ecrã. Neste ecrã tal como demonstra a figura, o utilizador fará as associações necessárias para registar a genealogia, sendo efectuada uma associação de cada vez. Note-se que não é necessário o utilizador conhecer os códigos previamente, já que através do nome é possível identificar qual é a pessoa que quer identificar com quem, tal como demonstra a Figura 16.

Id	Código de Ego 1	Nome de Ego 1	Relação	Sexo	Idade	Código de Ego 2	Nome de Ego 2	Data de Nascimento	Notas	Final
1	1	Bartolomeu César de Andrade	F	E		2	Manuel de Castro	1671-01-20		
2	2	Bartolomeu César de Andrade	E	E		3	Maria César (Dona)	1671-01-20		
3	3	Manuel de Castro	E	E		4	Gonçalo Gomes de Castro	1671-01-20		
4	4	Maria César (Dona)	E	E		5	Antónia Pais	1671-01-20		
5	5	Gonçalo Gomes de Castro	E	E		6	Manuel de Castro	1671-01-20		
6	6	Antónia Pais	E	E		7	Maria da Ponte (Dona)	1671-01-20		
7	7	Manuel de Castro	E	E		8	Francisco de Castro Berenguer	1671-01-20		
8	8	Maria da Ponte (Dona)	E	E		9	Josaria Noronha e Estremado (Dona)	1671-01-20		
9	9	Francisco de Castro Berenguer	E	E		10	Manuel de Faria e Abreu (Padre)	1671-01-20		
10	10	Josaria Noronha e Estremado (Dona)	E	E		11	Pedro de Faria de Abreu	1671-01-20		
11	11	Manuel de Faria e Abreu (Padre)	E	E		12	Maria das Neves	1671-01-20		
12	12	Pedro de Faria de Abreu	E	E		13	Manuel de Castro	1671-01-20		
13	13	Maria das Neves	E	E		14	Maria César (Dona)	1671-01-20		
14	14	Manuel de Castro	E	E		15	Francisco de Castro Berenguer	1671-01-20		
15	15	Maria César (Dona)	E	E		16	Josaria Noronha e Estremado (Dona)	1703-00-00	Freigeira de S. Pedro, Funchal (ARM, Casal)	TT. 1
16	16	Francisco de Castro Berenguer	W	E		17	Pedro de Faria de Abreu	1695-05-20		TT. 1
17	17	Josaria Noronha e Estremado (Dona)	B	E		18	Pedro de Faria de Abreu	1695-05-20		TT. 1
18	18	Pedro de Faria de Abreu	W	E	2	19	Josaria Maria de Bethencourt e Noronha	1706-00-00	Condição familiar devolvida à carta, expõe-se q	TT. 1
19	19	Josaria Maria de Bethencourt e Noronha	W	E	1	20	Mariana de Conceição	1701-00-00	Sé de Funchal (ARM, Casamentos de Freg. de	TT. 1
20	20	Mariana de Conceição	W	E	1	21	Pedro de Faria	1695-05-20	Pai = Mercador no exercício de baptismo desta	TT. 1
21	21	Pedro de Faria	F	E		22	Maria das Neves	1695-05-20		TT. 1
22	22	Maria das Neves	W	E		23	Manuel Fernandes	1695-05-20		TT. 1
23	23	Manuel Fernandes	W	E		24	Josaria de Freitas	1695-05-20		TT. 1
24	24	Josaria de Freitas	W	E		25	Francisco Rodrigues Noronha	1695-05-20		TT. 1
25	25	Francisco Rodrigues Noronha	W	E						

Figura 16 - Gestão de Genealogias no Sistema Informático de Gestão e Análises de Genealogias

Por último, o utilizador pode ainda efectuar uma gestão sobre as localidades existentes no sistema de gestão, pelo que deverá aceder pela opção, Localidades. A Figura 17 mostra o ecrã que o utilizador irá visualizar no acto de gerir uma localidade.

A opção recaiu sempre em ecrãs funcionais, sem demasiado grafismo para que o utilizador não se disperse e ao mesmo tempo tenha sempre no ecrã o máximo de informação possível, da forma que se consiga ter mais organizada.

Localidades	
Referencia:	50
Localidade:	Arcozelo
Provincia:	Entre Douro e Minho
Bispado:	Porto
Comarca:	Porto
Termo:	
Distrito:	
Concelho:	
Freguesia:	S. Miguel de Arcozelo
Capitania:	
Ilha:	
Lugar:	
Arquipelago:	
Latitude:	
Longitude:	
Ano:	
Fonte:	TT, HSO, José, Mç. 50, doc. 803, fl. 62v
Notas:	

Figura 17 - Gestão de Localidades no Sistema Informático de Gestão e Análises de Genealogias

Uma vez que a análise de dados se trata de um ponto fundamental nesta dissertação, existe a necessidade de separar em três subtemas, que correspondem às três vertentes de análise, sendo a última vertente (GEDCOM XML) ainda uma fase embrionária mas com um enormíssimo potencial tal como referido ao longo desta dissertação.

Assim, numa primeira fase serão exibidas exemplos de análises feitas no Personal Ancestral File exibindo assim algumas das possibilidades que o utilizador pode efectuar, sempre muito relacionadas com a análise de genealogias de forma padrão.

Numa segunda fase será dada ao utilizador a hipótese de analisar dados geograficamente, isto é, através do Google Earth. Serão apresentadas as opções assim como exemplos concretos do Caso de Estudo.

Por último o utilizador tem ainda a possibilidade de criar a representação GEDCOM XML dos seus dados genealógicos através da opção, GEDCOM XML situada no formulário entre as opções anteriores.

Desta forma o utilizador tem assim a possibilidade de escolha tal como mostra a Figura 18



Figura 18 - Menu de escolha das opções possíveis para análise de genealogias

5.1.1 PERSONAL ANCESTRAL FILE 5.2

Na eventualidade do utilizador escolher analisar dados genealógicos na forma mais elementar, escolhe esta opção. Assim, passam a ficar disponíveis ao utilizador duas opções nesta fase. Ou o utilizador pretende gerar apenas a representação GEDCOM para um ficheiro, ou pretende abrir o Personal Ancestral File para visualizar uma representação GEDCOM. A primeira opção, a de gerar um ficheiro, consiste apenas na sua criação, dando um nome por defeito, da data de criação do ficheiro com a extensão “.gedcom”. No entanto, isto acontece apenas para o caso em que o utilizador não altera o nome pelo que se assim o entender, pode alterar.

Na segunda opção, o utilizador abre o programa PAF previamente instalado ao qual terá de primeiro, importar os dados. Este facto acontece já que tal como vimos no decorrer desta dissertação, o GEDCOM é um formato universal e como tal qualquer programa de genealogia necessita primeiro de efectuar a sua importação. Nesta primeira fase, surge desde logo um indicador que poderá mostrar se os dados contêm ou não problemas. A Figura 19 mostra um exemplo disso mesmo, em que surgem seis erros, cinco dos quais estão relacionados com o limite de texto (256 caracteres) que um campo, provavelmente o campo Notas, trazia. O sexto erro, esse sim, é um erro consumado já que o PAF analisou na importação que um indivíduo do sexo feminino, com o código E3851, tinha sido definido como “marido” numa relação, o que constitui uma irregularidade.

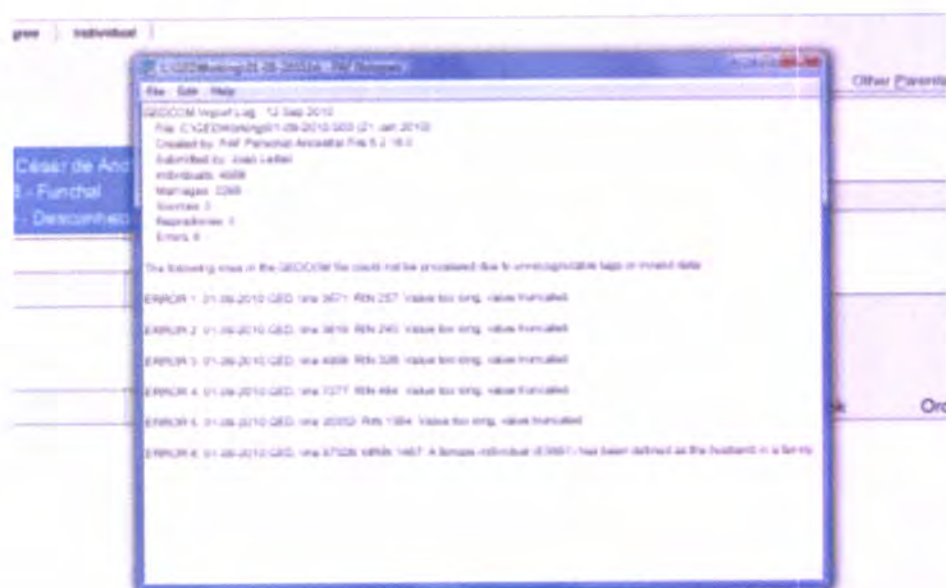


Figura 19 - Exemplo de Erros na importação de ficheiro GEDCOM para o Personal Ancestral File

Após a fase de importação o utilizador pode escolher uma de três opções: *Family*, *Pedigree* e *Individual*. A última opção mostra a lista de todos os indivíduos que estão agora presentes no ficheiro importado e que por sua vez estão no sistema de gestão, tal como demonstra a Figura 20.

ID#	Full Name	Sex	Birth - Date	Birth - Place	Death - Date	Death - Place	Sealing
1	Bethlehem Cabos de Avelãs	Male	<1871>	Funchal	<1871>	Desconhecida	
2	Manuel de Castro	Male	<1837>	Funchal	<1837>	Desconhecida	
3	Maria Cesar (Dona)	Female	<1837>	Funchal	<1837>	Desconhecida	
4	Gonçalo Gomes de Castro	Male	<1817>	Madeira	<1817>	Desconhecida	
5	Andria Pires	Female	<1817>	Madeira	<1817>	Desconhecida	
6	Melchior Beneguer de Andrade	Male	<1817>	Madeira	<1817>	Desconhecida	
7	Maria da Ponte (Dona)	Female	<1817>	Madeira	<1817>	Desconhecida	
8	Francisco de Castro Beneguer	Male	<1820>	Funchal	<1820>	Desconhecida	
9	Joana Noronha e Estevado (Dona)	Female	<1820>	Madeira	<1820>	Desconhecida	
10	Manuel de Faria e Abreu (Padre)	Male	1805-05-26	Funchal	<1805>	Desconhecida	
11	Pedro de Faria e Abreu	Male	1805-02-18	Funchal	<1805>	Desconhecida	
12	Mariana da Colação	Female	1800-02-25	Funchal	<1800>	Desconhecida	
13	Josefa Maria de Bethencun e Noronha (Dona)	Female	<1800>	Madeira	<1800>	Desconhecida	
14	Vitorino de Vasconcelos e Bethencun	Male	<1800>	Funchal	<1800>	Desconhecida	
15	Pedro de Faria	Male	<1800>	Desconhecida	<1782>	Desconhecida	
16	Maria das Neves	Female	1858-08-07	Funchal	<1827>	Desconhecida	
17	Manuel Fernandes	Male	<1827>	Funchal	<1827>	Desconhecida	
18	Joana de Freitas	Female	<1827>	Clérigos de Lobos	<1827>	Desconhecida	
19	Francisco Rodrigues Homem	Male	<1827>	Clérigos de Lobos	<24 Apr 1703>	Madeira	
20	Isabel Nunes	Female	<1827>	Funchal	<1827>	Desconhecida	
21	Catarina da Encarnação	Female	<1807>	Desconhecida	<1807>	Desconhecida	
22	Manuel Teixeira Brasil	Male	<1807>	Desconhecida	<1745>	Madeira	
23	Sesando Fernandes de Aguiar (Padre)	Male	1710-08-12	Funchal	<1807>	Desconhecida	
24	Sebastião Fernandes	Male	1800-01-25	Funchal	<1780>	Desconhecida	
25	Maria Vieira	Female	<1800>	Funchal	<1800>	Desconhecida	
26	Manuel Fernandes Pereira	Male	1818-03-26	Funchal	<25 Aug 1804>	Desconhecida	
27	Inácia de Aguiar	Female	1828-03-18	Funchal	<1828>	Desconhecida	
28	Armando Fernandes Pereira	Male	<1598>	Funchal	<12 Feb 1858>	Desconhecida	
29	Maria Fernandes	Female	<1598>	Desconhecida	<12 Feb 1858>	Desconhecida	
30	Manuel Pereira	Male	<1600>	Desconhecida	<1600>	Desconhecida	

Figura 20 - Listagem de Egos no Personal Ancestral File

Nesta lista pode-se ter à primeira vista uma percepção sobre cada um dos indivíduos, sendo apresentado o nome, género, data de nascimento e falecimento e respectivas localidades. Caso o utilizador pretenda conhecer um pouco mais sobre um determinado ego basta efectuar *duplo clique* com o botão esquerdo do rato sobre o nome que pretende conhecer mais. A Figura 21 demonstra esse processo, tendo escolhido para o efeito, o ego cujo nome é Francisco Rodrigues Homem.

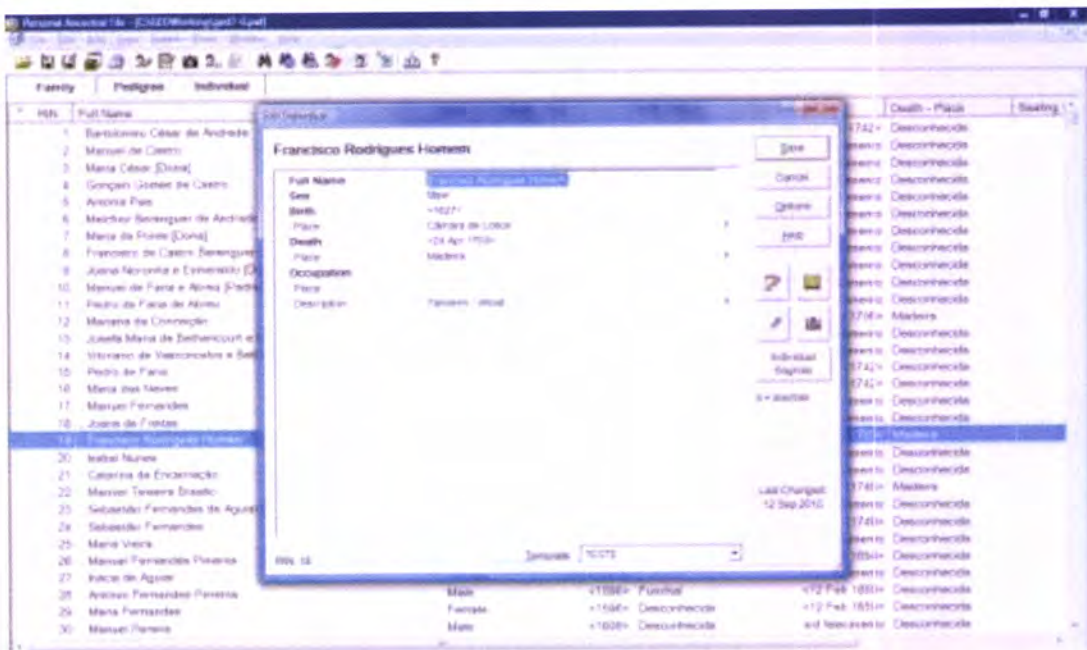


Figura 21 - Exemplo de dados relacionados com o Ego: Francisco Rodrigues Homem no Personal Ancestral File

É importante salientar que, ainda nesta opção de listagem de indivíduos é possível efectuar a ordenação por Nomes, o que ajuda bastante o utilizador caso este esteja à procura de um determinado nome em concreto ou até mesmo para ajudar a encontrar elementos duplicados, sendo que para este ultimo caso existe especificamente uma opção que será exibida mais à frente.

É no entanto, nos outros dois separadores, Family e Pedigree, que o utilizador irá visualizar e analisar toda a genealogia dos indivíduos. Começando pela opção mais limitadora no âmbito da análise genealógica, separador Family, este separador apenas retrata o caso de uma família, composta por “Pai, Mãe e Irmãos”. No entanto, é importante perceber como tudo isto se processa neste separador. Na opção Family é apresentada a informação sobre a indicação dos pais de um dos progenitores que pode ser escolhido a qualquer momento, assim como a informação de ambos os progenitores, com possível indicação da data de casamento e local, e posteriormente a indicação dos irmãos. No entanto, é o utilizador que tem de ser responsável por compreender os dados, isto é, a Figura 22 exemplifica o caso em que se analisa a família de Pedro de Faria casado com Maria das Neves resultando deste casamento três filhos, Pedro de Faria e Abreu, Catarina de Encarnação e o Padre Manuel de Faria e Abreu. Compreender que estes três nomes são

irmãos, deriva do facto de todos eles apresentarem o mesmo pai e a mesma mãe aquando da análise da família.

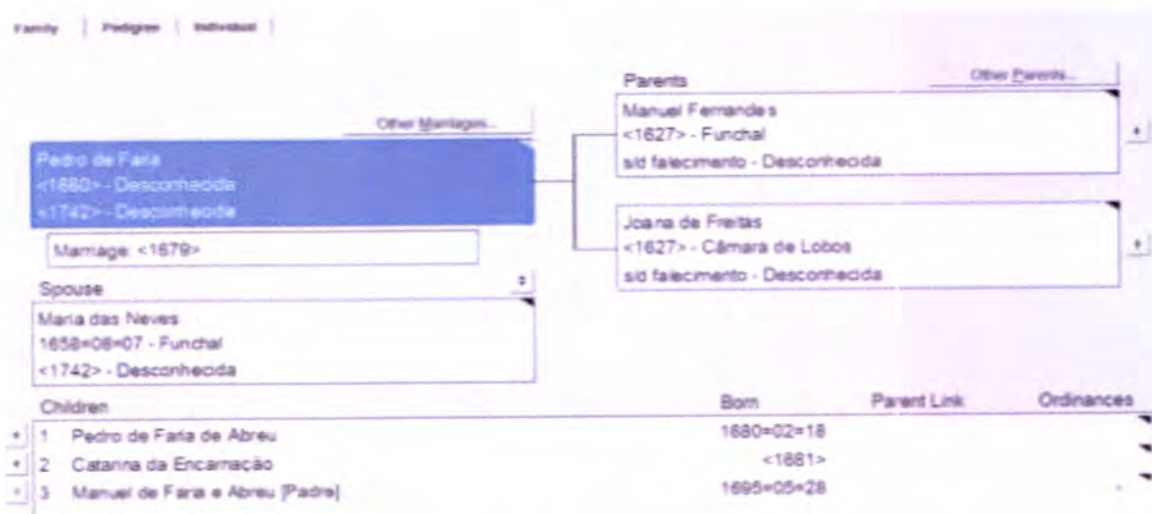


Figura 22 - Exemplo de análise de uma família no Personal Ancestral File

A qualquer momento pode ser feita a transformação do conceito de família, isto é, é possível através das setas localizadas à frente ou atrás do nome centrar toda a família nesse elemento escolhido, o que constitui uma mais-valia, já que é possível ir analisando família a família sem sair do ecrã.

A última das possíveis três opções dos separadores disponíveis, a opção *Pedigree*, mostra a árvore genealógica do ego escolhido. Neste caso, a Figura 23 mostra o exemplo do indivíduo Miguel de Aires Lobo de Carvalho do Vale cuja genealogia é dos casos mais complexos nestes dados, já que a mesma percorre até à nona geração de ascendentes, ou seja, é possível relacionar Miguel do Vale e os seus nonavôs, Cristóvão Vaz e Teresa Rodrigues. Apesar de numa primeira análise ser apenas possível visualizar cinco gerações, a introdução de setas à frente e atrás dos nomes permite por sua vez avançar ou recuar nas gerações. Estando recordado de o registo no sistema de gestão ser feito apenas entre pais e mães, conseguir visualizar toda a genealogia e chegar à nona geração (limitada pelo utilizador ao não introduzir mais dados) é de facto gratificante e ao mesmo tempo constitui uma forte componente de análise capaz de relacionar indivíduos que olhando apenas para os dados numa tabela era muito difícil de correlacionar.

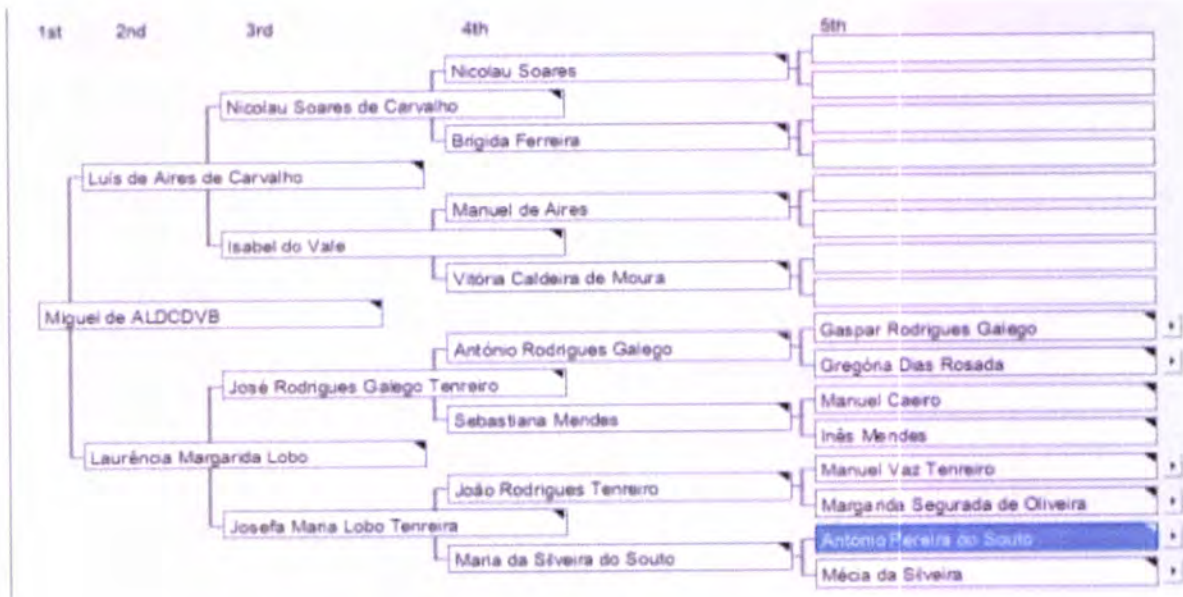


Figura 23 - Exemplo de Árvore Genealógica no Personal Ancestral File

Estes três separadores, por si só, já apresentavam uma vantagem na análise dos dados genealógicos muito grande porque permite responder às perguntas dos historiadores, verificando por exemplo, se um determinado ego apresentava “sangue limpo”. Quer isto dizer, que não podia ser descendente de mouros, raça negra nem qualquer outro tipo similar de raça. Com esta ferramenta é possível analisar isso, e abrir ainda o leque na análise de dados genealógicos.

No entanto, o PAF permite ainda um conjunto de ferramentas que para este projecto visaram ser fundamentais na medida em que proporcionaram descodificar e corrigir informações. Essas opções encontram-se no menu *File*, opção *Print Reports*.

Numa primeira análise pelo nome desta funcionalidade, será nesta opção que o utilizador consegue imprimir relatórios sobre os dados analisados. Estes relatórios tanto podem ser árvores genealógicas ascendentes ou descendentes sendo as opções parametrizadas à medida e ao gosto das preferências do utilizador.

No entanto, para além de tudo aquilo que é possível efectuar na construção e análises de gráficos nesta opção, saltam à vista algumas opções como, Duplicate Individuals e a opção *Possible Problems*, tal como demonstra a Figura 24

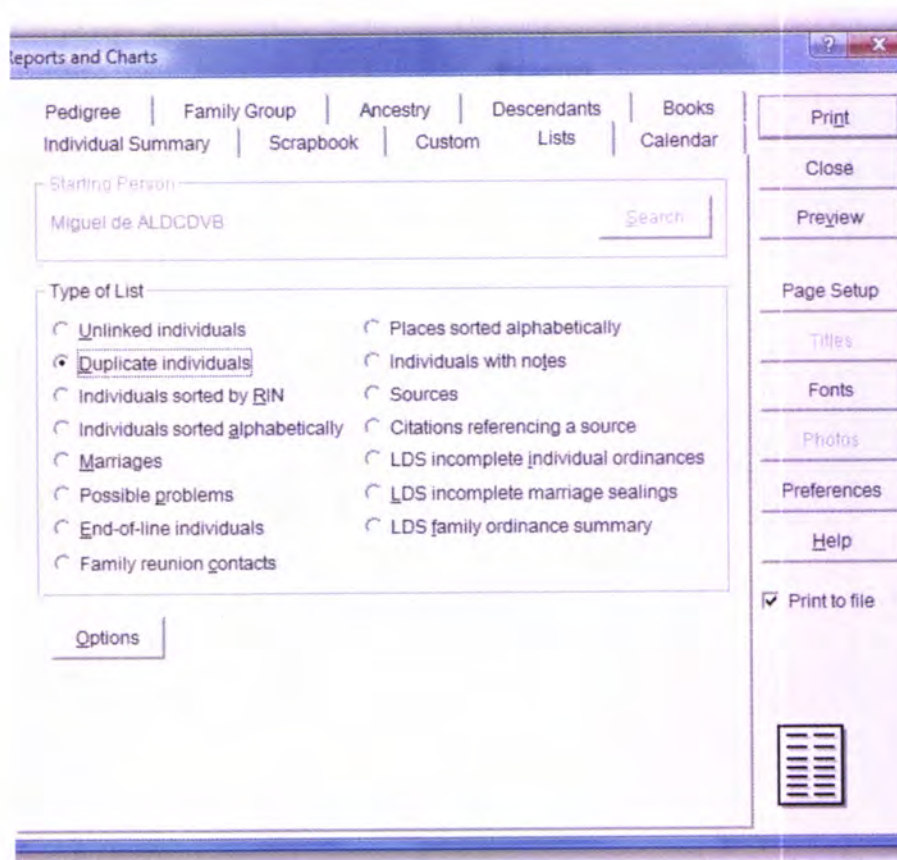


Figura 24 - Exemplo de opções avançadas no Personal Ancestral File

A primeira opção visa ser muito útil caso haja a incerteza de existirem dados duplicados no sistema. Desta forma, é criada uma lista com todos esses possíveis casos. A segunda opção é um ponto fundamental neste programa já que cria uma lista com possíveis problemas nos dados onde estão retratados alguns casos menos “coerentes” podendo os mesmos constituir uma situação de erro ou não. São situações que podem constituir erros os exemplos em que a data de nascimento de um ego é inferior à data de casamento dos progenitores. Esta situação pode não ser normal, mas acontece na realidade e pelo que pode em alguns casos constituir um erro e em outros casos não.

Um exemplo de um caso em que constitui um erro confirmado é quando por exemplo a data de falecimento é anterior à data de nascimento. Este passo chama assim à atenção

para esses casos e permitir que os mesmos sejam corrigidos de forma a garantir que os dados analisados se aproximam cada vez mais dos dados correctos.

São assim vários os casos em que regista uma ocorrência de possíveis erros, dependendo da situação pelo que uns podem ser assumidos como erros garantidos e outros que podem ou não dependendo da situação e do caso a analisar.

5.1.2 GOOGLE EARTH

Consumada a análise efectuada às possibilidades existentes na análise padrão de genealogias utilizando o Personal Ancestral File, o utilizador pode no entanto escolher o Google Earth para realizar análises de dados georreferenciados, sendo as opções disponíveis são as seguintes: Genealogia por Ego; Família por Ego (Pai, Mãe e Irmãos), História de Vida por Ego (Nascimento e Falecimento) e as opções Egos por Local de Nascimento e Falecimento, tal como demonstra a Figura 25.



Figura 25 - Ecrã escolha de opções - Análise Google Earth

A primeira opção que o utilizador dispõe é a visualização de toda a Genealogia por Código de Ego, isto é, esta opção permite percorrer a árvore genealógica no sentido, mostrando no mapa a informação sobre os pais, avós, bisavós, e assim sucessivamente. A Figura 26 exemplifica o caso do Ego1: Bartolomeu César de Andrade e toda a sua genealogia. Neste caso a diferenciação dos níveis é feita através da indicação do grau, por exemplo, avô materno, que indica ser pai da mãe do Ego1.

A opção Família por Ego permite ao utilizador visualizar a família de um ego, apenas para o primeiro nível, isto é, pai, mãe e irmãos. A Figura 27 mostra o exemplo da família de Hugo Maguiere (código de ego: E35), com destaque para o seu pai (Código de Ego: E36) e mãe (Código de Ego: E39). De notar que para este caso não existem irmãos pelo que não aparecem no mapa.



Figura 26 - Exemplo de visualização da genealogia do Ego E1 no Google Earth

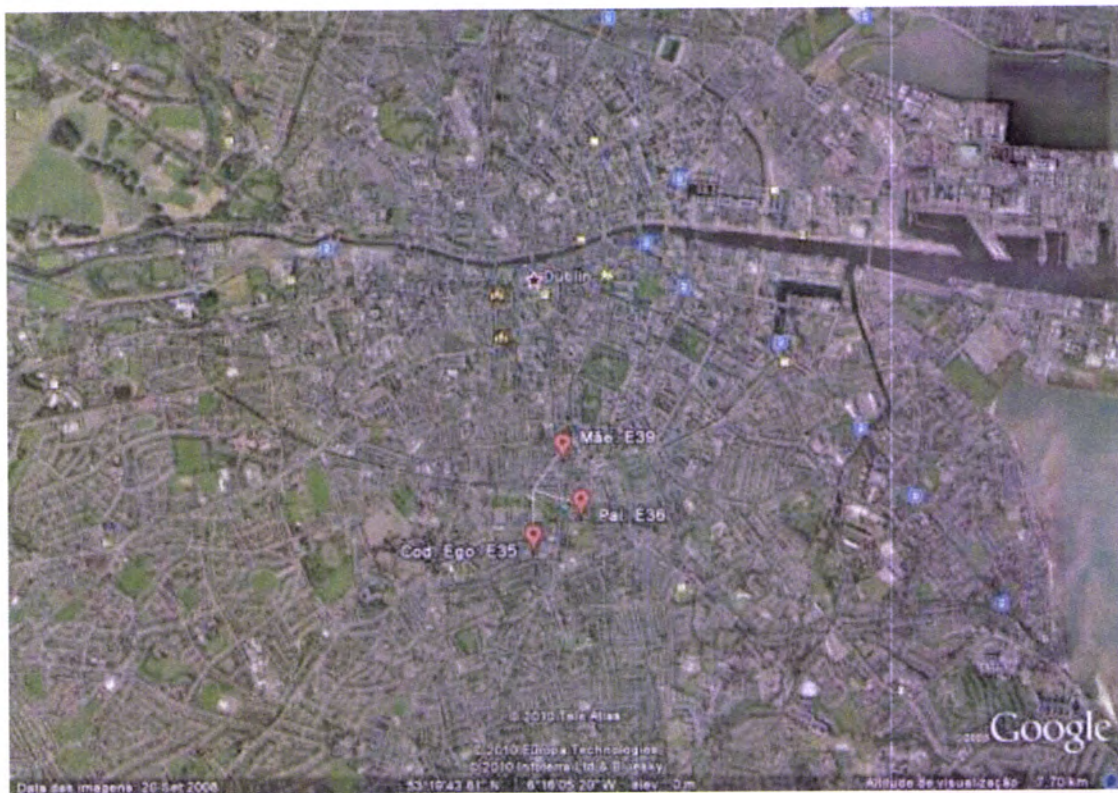


Figura 27- Exemplo de visualização da família (Pai, Mãe e Irmãos) do Ego E35 no Google Earth

Se o utilizador pretender visualizar apenas um código de ego conhecendo a “história de vida do mesmo”, isto é, o local de nascimento e de falecimento relativo apenas ao ego em questão, escolhe a opção, História de Vida por Ego, que requer tal como as opções anteriores, a escolha do código de ego a pesquisar.

Esta funcionalidade pode constituir uma mais-valia para verificar “por onde” passou o indivíduo. De notar, que se existem mais informações sobre a vida de um indivíduo para além da genealogia, essa informação poderia surgir aqui representada. No entanto, como tal não era objecto de estudo deste tema, esta opção retrata apenas os locais de nascimento e falecimento.

Um exemplo possível, trata-se do Ego Pedro da Costa, com o código 240, que nasceu numa localidade apelidada de Senrela, situada no concelho de Barcelos, distrito de Braga e acaba por falecer na localidade de Ossada situada igualmente no concelho de Barcelos, distrito de Braga. A Figura 28 exemplifica como surgiria no mapa estas informações mostrando o local de nascimento e local de falecimento com informações associadas ao ego, como o nome, pai, mãe, ocupação e as respectivas datas dos acontecimentos.



Figura 28 – Exemplo de visualização dos locais de nascimento e falecimento para o Ego: 240 no Google Earth

A escolha: Ego Por Local de Nascimento proporciona ao utilizador visualizar a informação dispersa pelo mapa-mundo de todos os egos existentes no sistema. Inevitavelmente, conhecendo os elevados números de indivíduos no presente sistema, a informação à primeira vista pode parecer confusa, isto é, no caso de não se efectuar uma aproximação ao navegador Google Earth, fica-se com a sensação de não se conseguir distinguir os dados, como demonstra a Figura 29.

De notar ainda que existem locais de nascimento espalhados por todo o mundo, sendo exemplos os casos de Portugal com as regiões autónomas da Madeira e Açores, Espanha, França, Reino Unido, Áustria e Macau. Por ser difícil mostrar numa imagem todos os locais descritos, escolheu-se a imagem onde surgissem uma maior quantidade de informação dispersa para que se possa ter em conta o que é explicado.



Figura 29 - Visão geral dos locais de nascimento por egos no Google Earth

De salientar ainda na Figura 29 que em Portugal Continental é a região onde existem maior incidência de pontos vermelhos, tendo sido este a principal fonte de arquivo e de estudo.

No entanto, realizando uma aproximação mais ao pormenor, percorrendo o mapa-mundo, chega-se à conclusão que é possível efectuar uma análise com detalhe, sendo que tem de forçosamente ser feita com um nível de detalhe muito elevado para que seja possível distinguir as localidades de nascimento de cada ego, como demonstra a Figura 30.

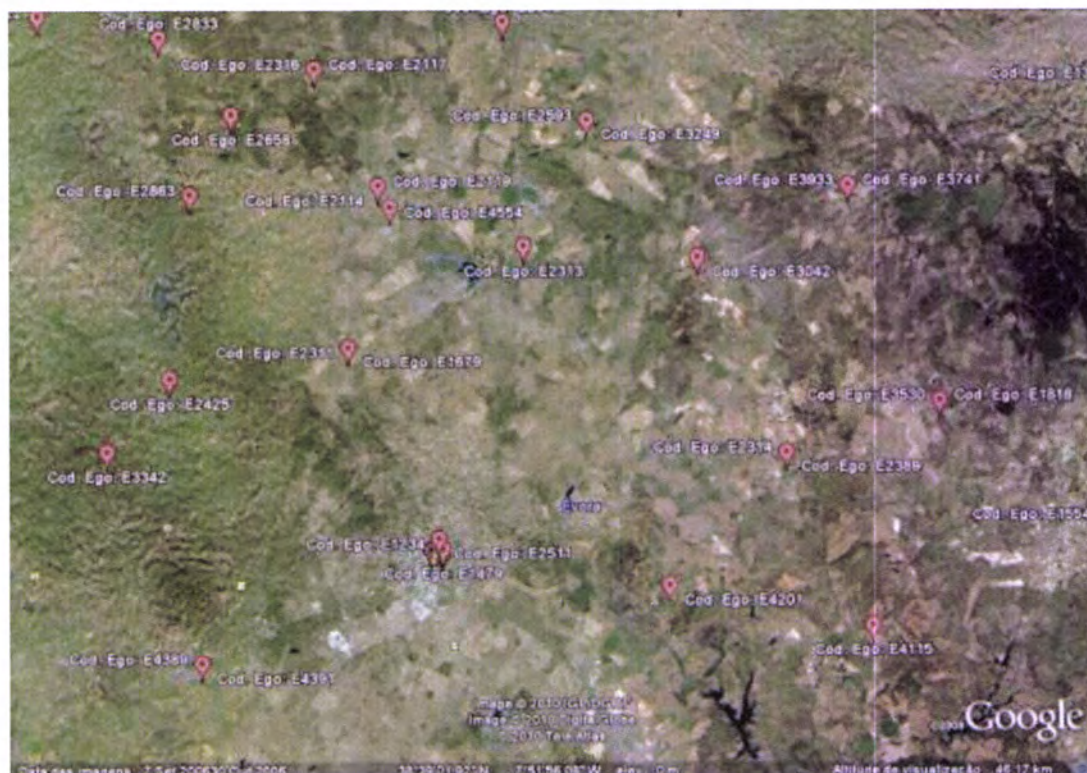


Figura 30 - Visão detalhada dos locais de nascimento por egos no Google Earth

Assim, é possível verificar que o nível de detalhe é maior que, chegando a visualizar por exemplo, na zona de Évora foram vários os egos que tiverem o seu nascimento registado. Note-se que em alguns casos, no mesmo ponto vermelho, surgem dois códigos de egos, o que significa que tiveram o mesmo local de nascimento.

Por último, a opção Egos por Local de Falecimento, é em tudo semelhante à opção anterior, pelo que a diferença reside no facto de os dados georreferenciados serem relativos aos locais de falecimento.

Todas as restantes considerações se mantêm, desde a falta de percepção dos dados numa visão mais geral e mais distante, passando pelo mesmo acontecimento dos dados

agrupados quando o local de falecimento é o mesmo. Um exemplo é a descrição dos locais de falecimento, olhando para o caso concreto da cidade de Coimbra, como demonstra a Figura 31.



Figura 31 - Visão geral das localidades de falecimento - Google Earth

O resultado da análise da Figura 31 chama a atenção pelo facto de os dados relativos ao falecimento estarem “agrupados” por quatro grandes zonas: Porto; Coimbra; Lisboa e Évora. As justificações possíveis podem ser várias, sendo os historiadores os melhores elementos para teorizarem sobre as mesmas. No entanto, a conclusão principal reside na origem de informação, já que são estes quatro pontos os principais elementos originadores de informação que sustenta este sistema informático.

5.1.3 GEDCOM XML

Se o utilizador pretender, por algum motivo, criar uma representação GEDCOM XML dos seus dados genealógicos, é esta a opção que tem de escolher. Nesta opção apenas está disponível a opção de criar um ficheiro com o formato GEDCOM pelo que no capítulo das Perspectivas Futuras, serão explicados alguns pontos que poderiam melhorar esta opção.

Assim, o utilizador pode gerar a representação GEDCOM XML para um ficheiro, escolhendo a única opção disponível que consiste apenas na sua criação de um ficheiro com o formato em causa, dando um nome por defeito, da data de criação do ficheiro com a extensão “.xml”. No entanto, isto acontece apenas para o caso em que o utilizador não altera o nome pelo que se assim o entender, pode alterar à semelhança do que acontecia quando gerava a representação GEDCOM.

A Figura 32 exemplifica todo este processo de escolha da criação da representação GEDCOM XML para um ficheiro com extensão XML.



Figura 32 - Processo de criação do ficheiro com representação GEDCOM XML

De forma a mostrar o formato GEDCOM XML dos dados no sistema de gestão surge a Figura 33, em que demonstra o resultado da transformação feita da representação GEDCOM para a representação agora denominada por GEDCOM XML.

```

<?xml version="1.0" encoding="UTF-8"?>
<GEDCOM>
<HEAD>
  <SOUR>
    PAF
    <NAME>
      <![CDATA[Personal Ancestral File]]>
    </NAME>
    <VERS>
      <![CDATA[5.2.18.0]]>
    </VERS>
    <CORE>
      <![CDATA[The Church of Jesus Christ of Latter-day Saints]]>
      <ADDR>
        50 East North Temple Street
        <CONT>
          Salt Lake City, UT 84150
        </CONT>
        <CONT>
          USA
        </CONT>
      </ADDR>
    </CORE>
  </SOUR>
  <DEST>
    Other
  </DEST>
  <DATE>
    21 Jan 2010
    <TIME>
      <![CDATA[15:37:08]]>
    </TIME>
  </DATE>
  <FILE>
    teste.ged
  </FILE>
  <GEDC>
    <VERS>
      <![CDATA[5.5]]>
    </VERS>
    <Form>
      <![CDATA[LINAGE-LINKED]]>
    </Form>
  </GEDC>
...

```

Figura 33 - Exemplo da visão geral da representação GEDCOM XML

É importante salientar que a indentação do documento XML só é feita para uma melhor compreensão por parte do utilizador pelo que não constitui obrigatoriedade tal como anunciado aquando da definição desta representação.

É ainda possível verificar o processo de atribuição da referenciação no ficheiro XML através da Figura 34 que demonstra a forma como é definida na linguagem GEDCOM XML, Catarina Martins com o código de ego E2623.

```
<INDI ID ="@E2623@">
  <NAME>
    Catarina Martins
  </NAME>
  <SEX>
    F
  </SEX>
  <OCUP>
    z/ Coup.
  </OCUP>
  <NOTE>
    Notas Gerais: /Data de nascimento grosseiramente estimada
  </NOTE>
  <NOTE>
    Cota: IT, HSO, José, Mç. 41, doc. 667, m31v
  </NOTE>
  <NOTE>
    Cota de Falecimento: IT, HSO, José, Mç. 41, doc. 667, m31v
  </NOTE>
  <NOTE>
    Cota de Ocupação: Cota de Ocupação não disponível
  </NOTE>
  <BIRT>
    <DATE>
      <![CDATA[1648<00<00]]>
    </DATE>
    <PLAC>
      <![CDATA[Desconhecida]]>
    </PLAC>
  </BIRT>
  <DEAT>
    <DATE>
      <![CDATA[1678<02<27]]>
    </DATE>
    <PLAC>
      <![CDATA[Desconhecida]]>
    </PLAC>
  </DEAT>
</INDI>
```

Figura 34 - Exemplo da definição do Ego: E2623 - Catarina Martins em formato GEDCOM XML

Para que a apresentação no browser fosse feita de forma correcta existiu a necessidade da introdução de um DTD (Document Type Definition) com as regras GEDCOM XML. Sabendo que a escolha recaiu sobre o formato adoptado pela Microsoft e pelos mórmones, a definição desse documento encontra-se listado no Anexo I.

Assim a Figura 35 mostra a forma como um browser interpreta o ficheiro GEDCOM XML validado pelo DTD definido pelos mórmones apoiado pela Microsoft.

```

<?xml version="1.0" encoding="ISO-8859-1" ?>
<!DOCTYPE GEDCOM (View Source for full doctype...)>
<?xml-stylesheet type="text/xsl" href="collapsed.xml"?>
- <GEDCOM>
+ <HEAD>
+ <SUBM ID="@SUB1@">
+ <INDI ID="@E1@">
+ <INDI ID="@E2@">
- <INDI ID="@E3@">
  <NAME>Maria César [Dona]</NAME>
  <SEX>F</SEX>
  <OCCU>s/ Ocup.</OCCU>
  <NOTE>Notas Gerais: Notas Gerais não disponíveis</NOTE>
  <NOTE>Cota: TT, HSO, Bartolomeu, Mç. 3, doc. 62</NOTE>
  <NOTE>Cota de Falecimento: Cota de Falecimento não disponível</NOTE>
  <NOTE>Cota de Ocupação: Cota de Ocupação não disponível</NOTE>
- <BIRT>
  - <DATE>
    <![CDATA[ 1637<00<00 ]]>
  </DATE>
  - <PLAC>
    <![CDATA[ Funchal ]]>
  </PLAC>
</BIRT>
- <DEAT>
  - <DATE>
    <![CDATA[ a/d falecimento ]]>
  </DATE>
  - <PLAC>
    <![CDATA[ Desconhecida ]]>
  </PLAC>
</DEAT>
</INDI>
+ <INDI ID="@E4@">
+ <INDI ID="@E5@">
+ <INDI ID="@E6@">

```

Figura 35 - Exemplo da visualização do ficheiro GEDCOM XML no browser Internet Explorer

A Figura 35 mostra assim uma forma diferente de visualização dos dados no formato GEDCOM XML e correctamente validados. É possível navegar por toda a informação, mostrando ou escondendo, opções consoante a necessidade de análise e vontade do utilizador.

CAPÍTULO VI - CONCLUSÃO

Chegado o final desta dissertação, acredito ter conseguido, falando de forma genérica, promover os recursos, metodologias e formas de divulgação ao dispor dos historiadores e público interessado em Genealogia e demografia histórica.

Sabendo de toda a importância que a Genealogia apresenta quer numa vertente profissional quer numa vertente mais lúdica, são muitos os casos em que este interesse surge por determinadas necessidades. Conseguir criar assim, um sistema que ajude e melhore a gestão e análise de dados genealógicos é de facto de extrema importância já que a exploração das raízes pode levar a uma consolidação de identidades, seja ao nível da família, seja ao nível da própria comunidade. Não vocacionada por todos, mas para muitos, a sedução de descobrir o passado, sabendo que a História faz parte de nós, é de facto importantíssima mesmo sabendo que o acesso a fontes fidedignas é complicado. No entanto, o aparecimento da Internet e de tecnologias como o XML visa garantir nova prosperidade na partilha e troca de dados genealógicos de forma cada vez mais rápida e eficaz. Para tal é necessário encontrarmos soluções capazes de dar resposta ao progresso que se impõe e que acompanhamos a cada dia que passa.

A construção deste Sistema Informático de Gestão e Análise de Genealogias procurou funcionar da mesma forma que o estudo da genealogia é feita, isto é, procurando sempre integrar um princípio, meio e fim. Num primeiro momento, verificou-se que a introdução de uma tecnologia como o *MySQL* e de um sistema de gestão sólido proporcionou corrigir erros do passado e ao mesmo tempo aumentar a capacidade de introdução de dados com formulários simples e eficazes para o utilizador. Porque sem a informação não é possível efectuar qualquer análise, esta primeira fase foi cuidadosamente planeada e definida para que não acontecessem erros, já que iriam depois ter repercussões sobre todo o sistema informático.

Permitir que a origem da informação possa ser transportada dos papéis ou da memória humana e transforma-la para um formato electrónico permanente exige muito da

aplicação e do sistema de gestão, já que uma má gestão pode transformar num autêntico caos não só a integridade dos dados em causa, como também futuras análises.

Numa segunda fase encontrou-se todo o núcleo deste sistema informático composto por procedimentos e funções capazes de gerar por um lado representações GEDCOM e por outro gerar representações XML e KML para que o utilizador tenha várias hipóteses de escolha e tentar impor-se o mínimo de rigidez à criatividade do próprio utilizador.

Desta forma, é então possível através deste sistema permitir a organização de dados genealógicos através da criação de árvores genealógicas, com recurso ao Personal Ancestral File (PAF), assim como organizar linhas cronologias de eventos cruzando informação à procura de respostas. Analisar dados no mapa-mundo que sejam passíveis de estar referenciados a um local é um “novo presente” do estudo da Genealogia já que começam a surgir como demonstrado no Estado da Arte, alguns programas genealógicos com recurso a esta tecnologia.

Assim, possibilitar a análise de dados quer num programa de genealogias bastante sólido e de renome como é o PAF e ao mesmo tempo a análise de dados georreferenciados no Sistema de Informação Geográfica de topo, Google Earth, visou ser uma mais-valia por responder às necessidades dos utilizadores e motivando-os a melhorar cada vez mais o seu trabalho à procura de novas respostas.

Apesar de disponibilizar apenas o ficheiro com a representação GEDCOM XML acredito ter sido um passo fundamental para futuros desenvolvimentos nesta área, já que é possível o utilizador ficar num só ficheiro, com toda a informação relativa. No entanto, o utilizador pode ainda visualizar essa informação através de um browser, ou partilhá-la pela internet, ou integrá-la num website, ou não fossem essas as inúmeras possibilidades que a linguagem XML oferece, acreditando claramente que este sistema será o futuro da gestão e análise de dados genealógicos. Por querer contribuir com o avanço da tecnologia, neste caso associado à Genealogia, foi essa a principal motivação para integrar esta tecnologia neste sistema informático já que por um lado permite converter os dados em formatos

GEDCOM para este novo formato já que se ninguém o fizer, este formato nunca chegará a ganhar relevância.

Do estado da arte fica a ideia de que as aplicações estão ainda um pouco paradas no que diz respeito a integração de tecnologias como o XML muito por culpa de não existir ainda um padrão oficial para partilha de dados genealógicos em formato XML. Apesar de existirem ferramentas muito poderosas da gestão e análise de genealogias, não se consegue encontrar um “tudo em um” que inclua funcionalidades de resposta quer para o presente, quer para o futuro.

Através deste estudo foi-me possível observar igualmente que para o formato GEDCOM XML ser aceite como padrão várias serão as situações que tem de mudar, isto porque o formato GEDCOM apresenta algumas limitações dadas as constantes exigências dos utilizadores desta nova sociedade global. Claro exemplo desta situação, são as situações de casamento, em que apenas é possível registar as duas pessoas que contraíram matrimónio. No entanto, era de todo conveniente registar padrinhos e madrinhas já que afinal, apresentam uma relação com um grau de importância maior que os restantes convidados.

Em suma, seja este Sistema Informático de Gestão e Análise de Genealogias cumpriu os objectivos a que foi destinado, acreditando sempre que se pode fazer melhor já que não existem obras perfeitas. Acredito ainda que este projecto possa ser aplicado a outros projectos, seja na área da Investigação de Genealogia ao nível profissional, pessoal ou até mesmo educativo, podendo este sistema estar integrado em escolas com educação primária, já que os alunos aprendem nesta fase da educação as relações de parentesco e com o auxílio do professor poderiam visualizar os dados electronicamente o que se tornaria mais apelativo e motivador para os próprios alunos, em escolas com educação secundária dada a possibilidade de aprender História e Geografia de Portugal ao mesmo tempo, passando pelas Dinastias dispostas electronicamente ou até mesmo num ensino superior, sendo um claro exemplo disto, este Caso de Estudo em concreto. É então por isso, um sem fim de opções que se tornam possíveis com a adopção de sistemas informáticos, como este Sistema Informático de Gestão e Análise de Genealogias, tratando-se assim de um sistema capaz de dar respostas desde as mais básicas até às mais exigentes necessidades do utilizador.

6.1 PERSPECTIVAS FUTURAS

No sentido da constante procura da perfeição, é com vontade que aguardo que outros "versados" do assunto completem o que consideram estar incompleto, esclareçam tudo aquilo que existir para esclarecer, melhorem o que não estiver bom. Todavia, a certeza de ter procurado fazer sempre um trabalho da melhor maneira e de forma o mais completa possível, independentemente do trabalho e esforço aplicado, faz-me acreditar que é sempre possível melhorar.

Tendo a vantagem de a base de informação disponibilizada ser de fonte bastante fidedigna, Arquivo Nacional da Torre do Tombo, todo e qualquer trabalho assente em informações fiéis torna-se muito mais fácil já que no final, o prazer e encanto de visualizar tudo aquilo que foi feito é recompensador.

No futuro, a perspectiva seria aplicar este Sistema de Informação não só a mais projectos desta área, Genealogia, mas como a outras áreas que fosse possível realizar o relacionamento entre entidades, sejam essas entidades pessoas, organizações, instituições ou qualquer outro tipo. Por ser um objectivo sem contornos bem definidos, não constou na lista de objectivos do estudo desta dissertação, sendo que, se o propósito é procurar cada vez mais e melhor, existe certamente esta necessidade, a de criar um Sistema de Informação capaz de responder a qualquer área que relacione entidades, seja essa área a Genealogia ou não.

Sabendo de todas as potencialidades de uma representação como é a representação XML, aliar a Genealogia a uma representação deste tipo, visa garantir essencialmente o futuro, já que no presente ainda não existe grandes respostas a nível de software genealógico capaz de responder às demais exigências de quem apresenta toda a sua genealogia em formato XML. Ao estarmos a garantir o futuro, estamos igualmente a permitir que outros desenvolvam sistemas capazes de interpretar representações genealógicas em XML, para que toda a genealogia informatizada dê o salto para o novo paradigma e para a nova abordagem que é a XML.

Tratando-se este ponto de perspectivas futuras, não podia deixar de apresentar referência ao Sistema de Informação Geográfico, Google Earth. A integração feita é uma amostra do parágrafo anterior, isto é, é a garantia do presente que a representação XML oferece aos utilizadores genealógicos, permitindo assim, novos paradigmas no estudo de dados genealógicos, passando a existir mais uma informação no Universo de análise, a informação sobre os locais. Possibilidades como a de efectuar todo o tipo de análise genealógica, espalhada pelo mapa-mundo, a de toda a aplicação estar consolidada num Sistema de Informação Geográfica, é de facto uma perspectiva futura, que penso estar a pouca distância do actual momento em que nos encontramos.

A definição de um vocabulário XML para a definição do dicionário da representação GEDCOM XML, assente na Língua Portuguesa, não tem como fundamento diferenciar de toda a comunidade mundial na troca e na partilha de dados genealógicos no formato GEDCOM XML, mas sim, promover a “nossa língua”, e fazê-la chegar o mais longe possível. Esta perspectiva futura permite assim, por um lado promover a Língua Portuguesa por todo o Mundo, e por outro lado, facilitar a análise e a leitura da própria representação GEDCOM XML por parte dos utilizadores lusos. Se juntarmos a capacidade de existir uma *Style Sheet* em XML que transforme a “linguagem” do documento com a representação GEDCOM XML em Português para uma outra língua, por exemplo, o Inglês, não estamos assim a colocar-nos de parte da comunidade genealógica e demais comunidades que realizam o estudo das relações. Estamos assim a garantir a interoperabilidade entre cada língua, cada dialecto, promovendo cada um à sua maneira, mas sempre chegando a um entendimento entre todos.

Assim, a grande perspectiva é a que este sistema permita a criação de outros sistemas que promovam um avanço tecnológico quer na gestão, quer na análise de dados genealógicos. Garantir a interoperabilidade de diversos sistemas, cada um com a sua tarefa, assim como diferentes linguagens e representações, é o supra-sumo dos objectivos e das perspectivas a longo prazo, já que, se for possível “cada um” expressar as suas ideias, os seus conteúdos, na forma que mais lhe convier, garantir que existe uma plataforma de entendimento entre “cada linguagem” visa garantir não só a continuidade na troca e partilha de informação, mas também, o angariar de novos curiosos, de novos profissionais, que vêm assim ser mais fácil expressar electronicamente, aquilo que expressam oralmente ou em formato

papel. O que verdadeiramente importa é descobrir os segredos do passado, compreender os do presente e revelar os do futuro.

Espero assim que este trabalho possa ajudar a chamar a atenção de outros investigadores para esta temática, contribuindo, deste modo, para um aprofundamento dos conhecimentos a este nível.

REFERÊNCIAS

A Igreja de Jesus Cristo dos Santos dos Últimos Dias. (2009). *Acerca da Igreja*. Obtido em 17 de Outubro de 2009, de <http://www.igreja-jesus-cristo.pt/acerca-da-igreja.html>

A Igreja de Jesus Cristo dos Santos dos Últimos Dias. (2001). *GEDCOM XML - Draft*. Obtido em 13 de Fevereiro de 2010, de <http://www.familysearch.org/GEDCOM/GedXML60.pdf>

About.com: Genealogy. (s/d). *GEDCOM*. Obtido em 07 de Novembro de 2010, de <http://genealogy.about.com/library/weekly/aa110100a.htm>

About.com: Genealogy. (2009). *History of Genealogy Software*. Obtido em 20 de Fevereiro de 2010, de <http://www.arqnet.pt/portal/estudo/genealog.html>

Almeida, L. F. (1999). *A Metodologia de Disseminação da Informação Geográfica e os Metadados*. Obtido em 10 de Abril de 2010, de Centro de Ciências Matemáticas e da Natureza - Rio de Janeiro: http://www.cprm.gov.br/publique/media/dou_fernandobarbosa.pdf

Amaral, M. (2008). *O Portal da História - Ciências Auxiliares da História*. Obtido em 13 de Março de 2010, de <http://www.arqnet.pt/portal/estudo/genealog.html>

Barbosa, F. (2000). *Genealogia*. Obtido em 13 de Março de 2010, de <http://www.barbosa.acthus.com/genealogia.html>

Battersby, J. (2007). *MacWorld*. Obtido em 10 de Abril de 2010, de http://www.macworld.com/reviews/product/405349/review/reunion_904.html?expand=true

Beecher, J. (2010). *Best Genealogy Software*. Obtido em 29 de Janeiro de 2010, de <http://genealogy-software-review.toptenreviews.com/>

Butler, D. (2006). Virtual Globes: The web-wide world. *Nature volume 439* .

Caldeira, C. (2008). *DATA WAREHOUSING -- Conceitos e modelos.1ª Edição*. Lisboa: Edições Sílabo ISBN 978-972-618-479-9.

Caldeira, C. (2007). Prosopografia, bases de dados e análise de redes: pressupostos teórico-metodológicos. *Seminário de Metodologias*. Instituto de Investigação Científica Tropical, Lisboa.

Chapple, M. (2010). *Microsoft Access 2007 Review*. Obtido em 23 de Março de 2010, de <http://databases.about.com/od/productinfo/gr/accesspr.htm>

Cobra, R. Q. (2001). Técnicas de Pesquisa: Determinação de Datas Aproximadas para Nascimentos. Obtido em 11 de Dezembro de 2010, de <http://www.cobra.pages.nom.br/gen-pesqdatas.html>

ConsumerSearch. (2010). *Compare Genealogy Software*. Obtido em 10 de Abril de 2010, de <http://www.consumersearch.com/genealogy-software/compare>

Correia, F. A. (2006). *A Família Correia (Pinto) da Penajóia*. Obtido em 10 de Março de 2010, de <http://www.familiacorreiapinto.com/Introducao.htm>

Crowder, D. (2007). *Google Earth for Dummies*. Wiley Publishing.

Crume, R. (2010). *Family Tree Magazine*. Obtido em 10 de Abril de 2010, de <http://www.familytreemagazine.com/ResearchToolkit/SoftwareGuide>

Damas, L. M. (2005). *SQL - Structured Query Language - 6ª Edição Actualizada e Aumentada*. FCA - Editora Informática.

Day, M. (1998). *Issues and Approaches to Preservation Metadata*. Obtido em 10 de Abril de 2010, de UKOLN: The UK Office for Library and Information Networking.: <http://www.ukoln.ac.uk/metadata/presentations/rlg-npo/warwick.html>

Dempsey, L. & Heery, R. (1998). Metada: A Current View of Practice and Issues. *Journal of Documentation*, v. 54, n.2 .

Doane, G. (1992). *Searching for Your Ancestors: The How and Why of Genealogy, 6th edition*. Minneapolis: University of Minnesota Press.

DoroTree Technologies Ltd. (2005). *Official Flyer*. Obtido em 23 de Janeiro de 2010, de http://www.dorotree.com/Documents/Flyer_Eng.pdf

Dullius, W. M. (2004). *Comentários aos Sistemas de Numeração em Genealogia*. Obtido em 21 de Abril de 2010, de http://buratto.org/gens/gn_comentarios.html

Dupâquier, J. (2009). *A População Mundial no Século XX*. Instituto Piaget Editora.

Eastman, D. (2001). *Clarification of the Use of GEDCOM/XML*. Obtido em 13 de Fevereiro de 2010, de <http://www.ancestry.com/learn/library/article.aspx?article=3788>

Eastman, D. (2004a). *GEDCOM*. Obtido em 07 de Novembro de 2009, de Encyclopedia of Genealogy: <http://www.eogen.com/GEDCOM>

Eastman, D. (2004b). *GeniML*. Obtido em 15 de Março de 2010, de <http://www.eogn.com/archives/news0404.htm>

Eastman, D. (2010). *Eastman's Online Genealogy Newsletter*. Obtido em 10 de Abril de 2010, de http://blog.eogn.com/eastmans_online_genealogy/2010/01/rootsmagic-essentials.html

EDUCASE Learning Initiative. (2006). *7 things you should know about...Google Earth*. Obtido em 06 de Fevereiro de 2010, de net.educause.edu/ir/library/pdf/ELI7019.pdf

eHow. (s/d). *The Disadvantages of Google Earth & Internet Mapping Service*. Obtido em 13 de Fevereiro de 2010, de http://www.ehow.com/list_6527114_disadvantages-earth-internet-mapping-service.html

Family Chronicle. (1999). *History of Genealogy*. Obtido em 29 de Março de 2010, de Magazine: <http://www.familychronicle.com/HistoryOfGenealogy.html>

Family Historian Official Webpage. (2010). *Reviews & Awards*. Obtido em 23 de Janeiro de 2010, de <http://www.family-historian.co.uk/reviews>

Family History Department - The Church of Jesus Christ of Latter-day Saints. (1996). *The Gedcom Standard 5.5*.

FCT - Fundação para a Ciência e Tecnologia. (2006). *Ministério da Ciência, Tecnologia e Ensino Superior*. Obtido em 01 de Maio de 2010, de Projecto PTDC/HAH/66107/2006: http://www.fct.mctes.pt/projectos/pub/2006/Painel_Result/vglobal_projecto.asp?idProjecto=66107&idElemConcurso=1326

Félix, R. F. (2002). *XML na Demografia Histórica: Anotação de Registos Paroquiais*. Obtido em 10 de Abril de 2010, de http://repositorium.sdum.uminho.pt/bitstream/1822/2335/1/CAPSI2002_XML_RFF_FF_MYS_PRH.pdf

Ferreira, A. (2004). *Sistemas Informáticos para análise de dados*. Obtido em 11 de Abril de 2010, de <http://repositorium.sdum.uminho.pt/bitstream/1822/894/1/Granada.pdf>

Filho, K. & Saraiva, M. (2010). *Sistemas de Coordenadas*. Obtido em 10 de Abril de 2010, de <http://astro.if.ufrgs.br/coord.htm>

Floridi, L. (2005). *Is Semantic Information Meaningful*. Obtido em 13 de Janeiro de 2010, de *Philosophy and Phenomenological Research*: <http://philsci-archive.pitt.edu/archive/00002536/01/iimd.pdf>

Flynn, P. (2010). *Frequently-Asked Questions about the Extensible Markup Language*. Obtido em 10 de Abril de 2010, de <http://xml.silmaril.ie/>

Ged2X. (2006). *Ged2X: Genealogy Data Services*. Obtido em 28 de Março de 2010, de <http://www.ged2x.com/about.html>

GenSoftReviews. (2010a). *Family Historian Reviews*. Obtido em 23 de Janeiro de 2010, de <http://www.gensoftreviews.com/?p=14&sel=&new=&lic=&pla=&type=&sort=>

GenSoftReviews. (2010b). *Family Tree Builder*. Obtido em 10 de Abril de 2010, de <http://www.gensoftreviews.com/?p=223>

GenSoftReviews. (2010c). *Personal Ancestral File Reviews*. Obtido em 23 de Janeiro de 2010, de <http://www.gensoftreviews.com/?p=126>

Gibin, M., Singleton, A., Milton, R., Mateos, P., & Longley, P. (2008). *Applied Spatial Analysis and Policy. An Exploratory Cartographic Visualisation of London thorough the GoogleMaps API*.

Goffart, W. (2003). *Historical Atlases - The First Three Hundred Years 1570-1870*. Chicago: University Of Chicago Press.

Google Earth Official Webpage. (s/d). *Google Earth*. Obtido em 06 de Fevereiro de 2010, de <http://earth.google.com/index.html>

GoogleCode. (2007). *KML Documentation*. Obtido em 6 de Fevereiro de 2010, de <http://code.google.com/intl/pt-PT/apis/kml/documentation/whatiskml.html>

Gouveia, L. M. (1996). *Sistemas de Informação*. Obtido em 26 de Fevereiro de 2010, de http://www2.ufp.pt/~lmbg/textos/si_texto.pdf

Groh, M. R., Stockman, J. C., Powell, G., Prague, C., Irwin, M. & Reardon, J. (2007). *Access 2007 Bible*. Wiley Publishing, Inc.

Hearn, S. (2006). *Why Study Your Genealogy?* Obtido em 03 de Abril de 2010, de <http://ezinearticles.com/?Why-Study-Your-Genealogy?&id=285980>

Howe, D. (1996). *Free on-line Dictionary of Computing (FOLDOC)*. Obtido em 10 de Abril de 2010, de <http://wombat.doc.ic.ac.uk/>

Jacobus, D. L. (1968). *Genealogy as Pastime and Profession, 2nd edition*. Baltimore: Genealogical Publishing.

Jorge, A. M., Maria, A., Vilar, H., Henriques, P. & Lopes, S. (2004). *Construção e Exploração de uma Base de Dados Prosopográfica normalizada do Clero Catedralício Português na*

Idade Média. Obtido em 10 de Janeiro de 2010, de www3.di.uminho.pt/~gepl/SIEP/docs/GranadaFastiv7.doc

Loureiro, H. (2008). *Access 2007 Macros & VBA - Curso Completo*. FCA - Editora de Informática.

Miller, F., Vandome, A. & McBrewster, J. (2009). *GEDCOM: De facto, Specification (technical standard), Genealogy, Data, Genealogy software, The Church of Jesus Christ of Latter-day Saints, Research, Character encoding, ANSEL, ASCII, Metadata*. Alphascript Publishing.

Morán, G. (s/d). *Personal Ancestral File 5.2*. Obtido em 10 de Abril de 2010, de Informer Technologies, Inc.: <http://personal-ancestral-file.software.informer.com/5.2/>

MyHeritage. (2010). *Family Tree Builder*. Obtido em 10 de Abril de 2010, de <http://www.myheritage.com.pt/FP/Company/family-tree-builder-learn-more.php>

No1Reviews.com. (2009). *FamTree Reviews*. Obtido em 04 de Fevereiro de 2010, de <http://genealogy-software.no1reviews.com/famtree.html>

Nunes, P. (2008). *Conceito de Comunicação*. Obtido em 13 de Fevereiro de 2010, de <http://www.knoow.net/cienceconempr/gestao/comunicacao.htm>

Nunes, P. (2009). *Sistemas de Informação (SI)*. Obtido em 03 de Março de 2010, de Ciências Económicas e Empresariais : <http://www.knoow.net/cienceconempr/gestao/sistinform.htm>

Olival, F. (2008). *Inquirir da honra: comissários do Santo Ofício e das Ordens Militares em Portugal*. Obtido em 03 de Junho de 2010, de <http://www.cidehus.uevora.pt/investigacao/progcien/linv/projaprov/santoficio.htm>

Oliveira, M. (2008). *Sistema de Informação Geográfica Distribuído*. Obtido em 13 de Março de 2010, de <http://repositorium.sdum.uminho.pt/bitstream/1822/8664/2/tese-Marco%20oliveira.pdf>

Owe, C. (2010). *GenXML*. Obtido em 29 de Março de 2010, de <http://cosoft.org/genxml/>

Portal Oficial ANTT. (s/d). *Página Oficial Arquivo Nacional da Torre do Tombo*. Obtido em 19 de Junho de 2010, de <http://antt.dgarq.gov.pt/>

Portal Oficial Projecto FCT - PTDC/HAH/64160/2006. (s/d). *Comissários do Santo Ofício e das Ordens Militares*. Obtido em 18 de 11 de 2009, de <http://www.santooficio.uevora.pt/>

Powell, K. (s/d). *Personal Ancestral File*. Obtido em 10 de Abril de 2010, de About.com: Genealogy: <http://genealogy.about.com/cs/softwarereviews/gr/paf.htm>

Ratliff, E. (2007). *Google Maps is Changing the Way We See the World*. Obtido em 03 de Fevereiro de 2010, de Wired Magazine, Issue15.07: http://www.wired.com/techbiz/it/magazine/15-07/ff_maps?currentPage=all

Refsnes, J. E. (1999a). *XML DTD - An Introduction to XML Document Type Definitions*. Obtido em 10 de Maio de 2010, de <http://www.xmlfiles.com/dtd/>

Refsnes, J. E. (1999b). *XML XSL - An Introduction to XML Extensible Stylesheet Language*. Obtido em 10 de Maio de 2010, de <http://www.xmlfiles.com/xsl/>

Shannon, C. (2003). *The lattice theory of information*. Obtido em 13 de Janeiro de 2010, de http://ieeexplore.ieee.org/xpl/freeabs_all.jsp?arnumber=1188572

Skonnard, A. (2004). *XML Data Migration Case Study: GEDCOM*. Obtido em 13 de Fevereiro de 2010, de MSDN Magazine - Microsoft: <http://msdn.microsoft.com/en-us/magazine/cc188730.aspx>

Smith, T. (2010). *Family Tree Builder review*. Obtido em 10 de Abril de 2010, de <http://www.computeractive.co.uk/computeractive/downloads/2256365/family-tree-builder>

Sousa, A. A. (2002). *Bases de Dados, Web e XML*. FCA.

Taylor, C. (1999). *An Introduction to Metadata*. Obtido em 10 de Abril de 2010, de University of Queensland Library.: <http://www.library.uq.edu.au/iad/ctmeta4.html>

TopTenReviews. (2010). *Genealogy Software Review*. Obtido em 29 de Janeiro de 2010, de <http://genealogy-software-review.toptenreviews.com/>

W3C. (2003). *Extensible Markup Language (XML)*. Obtido em 10 de Maio de 2010, de <http://www.w3.org/XML/>

W3C. (1999). *XSL Transformations (XSLT)*. Obtido em 10 de Maio de 2010, de <http://www.w3.org/TR/xslt>

w3schools. (2008). *XML Encoding*. Obtido em 19 de Maio de 2010, de http://www.w3schools.com/XML/xml_encoding.asp

Walsh, N. (1998). *A Technical Introduction to XML*. Obtido em 10 de Maio de 2010, de <http://www.xml.com/pub/a/98/10/guide0.html>

Wilson, T. (2008). *OGC KML 2.2.0 - Document #07-147r2*. Obtido em 06 de Fevereiro de 2010, de Open Geospatial Consortium: <http://www.opengeospatial.org/standards/kml/>

XBI Software Inc. . (2006). *XGenML - a global consortium for reporting genealogical information*. Obtido em 27 de Março de 2010, de <http://xgenml.org/>

ANEXOS

ANEXO I – DOCUMENT TYPE DEFINITION – GEDCOM XML

```
<!-- GEDCOM 6.0 DTD
  < xml version="1.0" encoding="UTF-8"?>
  <!DOCTYPE GEDCOM SYSTEM "/GEDCOM60.dtd">
  <GEDCOM>
  <HeaderRec>
  <FileCreation Date=""></FileCreation>
  <Submitter>
  <Link Target="" Ref=""></Link>
  </Submitter>
  </HeaderRec>
  <IndividualRec Id="testID">
  </IndividualRec>
  </GEDCOM>
<!-- GEDCOM root -->
<!ELEMENT GEDCOM
  (HeaderRec,
  FamilyRec*,
  IndividualRec+,
  EventRec*,
  LDSOrdRec*,
  ContactRec*,
  SourceRec*,
  RepositoryRec*,
  GroupRec*)>
<!-- elements common to multiple records -->
<!ENTITY % RecordCom
  "ExternalID*,
  Submitter?,
  Note*,
  Evidence*,
  Enrichment*,
  Changed*">
<!-- header record -->
<!ELEMENT HeaderRec
  (FileCreation,
  Citation?,
  Submitter,
  Note?)>
<!-- family record -->
<!ELEMENT FamilyRec
  (HusbPath?,
  WifeMoth?,
  Child*,
  BasedOn?,
  %RecordCom;)>
<!ATTLIST FamilyRec
  Id ID #REQUIRED>
<!-- individual record -->
<!ELEMENT IndividualRec
  (IndivName*,
  Gender?,
  DeathStatus?,
  PersInfo*,
  AssocIndiv*,
  DupIndiv*,
  %RecordCom;)>
<!ATTLIST IndividualRec
  Id ID #REQUIRED>
<!-- event record -->
<!ELEMENT EventRec
  (Participant+,
  Date?,
  Place?,
  Religion?,
  %RecordCom;)>
<!ATTLIST EventRec
  Id ID #REQUIRED
  Type CDATA #REQUIRED
  VitalType (birth | marriage | divorce | death ) #IMPLIED>
<!-- LDS ordinance record -->
<!ELEMENT LDSOrdRec
  (Participant+,
```

```

        OrdStat*,
        TempleCode?,
        Date?,
        Place?,
        BasedOn?,
        %RecordCom;)>
<!ATTLIST LDSOrdRec
  Id ID #REQUIRED
  Type (B | C | E | SM | SP | SS | TO | W) #REQUIRED>
<!-- contact record -->
<!ELEMENT ContactRec
  (Name,
   MailAddress*,
   Phone*,
   Email*,
   URI*,
   Public?,
   Note*,
   Changed*,
   SameIndiv*)>
<!ATTLIST ContactRec
  Id ID #REQUIRED
  Type CDATA #IMPLIED>
<!-- source record -->
<!ELEMENT SourceRec
  (Repository*,
   Title,
   Article?,
   Author?,
   URI*,
   Publishing?,
   Note*,
   Changed*)>
<!ATTLIST SourceRec
  Id ID #REQUIRED
  Type CDATA #IMPLIED
  xml:lang NMTOKEN #IMPLIED>
<!-- repository record -->
<!ELEMENT RepositoryRec
  (Name,
   MailAddress*,
   Phone*,
   Email*,
   URI*,
   Note*,
   Changed*)>
<!ATTLIST RepositoryRec
  Id ID #REQUIRED
  Type CDATA #IMPLIED>
<!-- group record -->
<!ELEMENT GroupRec
  (Name,
   Contact*,
   Member*,
   ParentGroup*,
   MailAddress*,
   Phone*,
   Email*,
   URI*,
   %RecordCom;)>
<!ATTLIST GroupRec
  Id ID #REQUIRED
  Type CDATA #IMPLIED>
<!-- subrecord elements -->
<!ELEMENT Addressee
  (#PCDATA)>
<!ELEMENT AddrLine
  (#PCDATA |
   Addressee |
   PlacePart)*>
<!ELEMENT Age
  (#PCDATA)>
  <!-- content = YY[ y MM m DD d], e.g., 26 y 4 m 21 d -->
<!ELEMENT Article
  (#PCDATA)>
<!ELEMENT Association
  (#PCDATA)>
<!ELEMENT AssocIndiv
  (Link,
   Association,
   Note*,
   Citation*)>
<!ELEMENT Author
  (#PCDATA)>
<!ELEMENT b
  (#PCDATA)>

```

```

<!ELEMENT BasedOn
  (Link*,
   Note*)>
<!ELEMENT blockquote
  (#PCDATA)>
<!ELEMENT br
  EMPTY>
<!ELEMENT CallNbr
  (#PCDATA)>
<!ELEMENT Caption
  (#PCDATA)>
<!ELEMENT center
  (#PCDATA)>
<!ELEMENT Changed
  (Contact?,
   Note*)>
<!ATTLIST Changed
  Date CDATA #REQUIRED
  Time CDATA #REQUIRED>
<!-- Date = D MMM YYYY, e.g., 3 MAR 1842 -->
<!-- Time = H:MM[:SS[.F]], e.g., 5:42:03.8 -->
<!ELEMENT Child
  (Link,
   ChildNbr?,
   RelToFath?,
   RelToMoth?)*>
<!ELEMENT ChildNbr
  (#PCDATA)>
<!ELEMENT Citation
  (Link,
   Caption?,
   WhereInSource?,
   WhenRecorded?,
   Extract*,
   Note*)>
<!ELEMENT Contact
  (Link)>
<!ELEMENT Coordinates
  (#PCDATA)>
<!-- example: N35.6756 W67.8967 -->
<!ELEMENT Copyright
  (#PCDATA)>
<!ELEMENT Date
  (#PCDATA)>
<!-- content = [QUAL] [[D ]MMM ]YYYY, e.g., 3 MAR 1842 -->
<!-- QUAL = ABT|CAL|EST|AFT|BEF|BET|FROM|TO -->
<!ATTLIST Date
  Calendar CDATA #IMPLIED>
<!-- Calendar = Gregorian|Julian|Hebrew|French|Roman|unknown -->
<!ELEMENT DeathStatus
  (#PCDATA)>
<!-- content = dead|stillborn|infant|child -->
<!ELEMENT DupIndiv
  (Link,
   Note*,
   Citation*)>
<!ELEMENT Email
  (#PCDATA)>
<!ELEMENT Enrichment
  (Citation*)>
<!ELEMENT Evidence
  (Citation*)>
<!ELEMENT ExternalID
  EMPTY>
<!ATTLIST ExternalID
  Type CDATA #REQUIRED
  Id CDATA #REQUIRED>
<!-- Type = User|AFN -->
<!ELEMENT Extract
  (#PCDATA |
   u | b | i | blockquote |
   br | p | sub | sup | center)*>
<!ATTLIST Extract
  xml:lang NMTOKEN #IMPLIED>
<!ELEMENT FamilyNbr
  (#PCDATA)>
<!ELEMENT FileCreation
  (Product?,
   Copyright?)*>
<!ATTLIST FileCreation
  Date CDATA #REQUIRED
  Time CDATA #IMPLIED>
<!ELEMENT Gender
  (#PCDATA)>
<!-- content = F|M|U -->
<!ELEMENT HusbFath

```

```

        {Link,
        FamilyNbr?}>
<!ELEMENT i
  {#PCDATA}>
<!ELEMENT IndivName
  {#PCDATA |
  NamePart |
  IndNameVariation}*>
<!ATTLIST IndivName
  Type CDATA #IMPLIED
  Method CDATA #IMPLIED
  xml:lang NMTOKEN #IMPLIED>
  <!-- Type = nickname|alias|aka|married|maiden|... -->
<!ELEMENT IndNameVariation
  {#PCDATA |
  NamePart}*>
<!ATTLIST IndNameVariation
  Type CDATA #IMPLIED
  Method CDATA #IMPLIED
  xml:lang NMTOKEN #IMPLIED>
<!ELEMENT Information
  {#PCDATA}>
<!ELEMENT Link
  EMPTY>
<!ATTLIST Link
  Target CDATA #REQUIRED
  Ref IDREF #REQUIRED>
<!ELEMENT Living
  {#PCDATA}>
  <!-- content = Y|N -->
<!ELEMENT MailAddress
  {AddrLine*}>
<!ATTLIST MailAddress
  xml:lang NMTOKEN #IMPLIED>
<!ELEMENT Member
  {Link,
  Role?}>
<!ELEMENT Name
  {#PCDATA}>
<!ATTLIST Name
  Type CDATA #IMPLIED
  xml:lang NMTOKEN #IMPLIED>
  <!-- Type = product|person|business|organization|government agency|... -->
<!ELEMENT NamePart
  {#PCDATA}>
<!ATTLIST NamePart
  Type CDATA #REQUIRED
  Level CDATA #IMPLIED>
  <!-- Level 1=surname, 2=maiden name, 3=given name -->
<!ELEMENT Note
  {#PCDATA}>
<!ATTLIST Note
  xml:lang NMTOKEN #IMPLIED>
<!ELEMENT OrdStat
  {Date?}>
<!ATTLIST OrdStat
  Code {bic|canceled|child|cleared|completed|dns|
  qualified|stillborn|submitted|uncleared} #REQUIRED>
<!ELEMENT p
  {#PCDATA}>
<!ELEMENT ParentGroup
  {Link}>
<!ELEMENT Participant
  {Link,
  Role?,
  Living?,
  Age?}>
<!ELEMENT PersInfo
  {Information?,
  Date?,
  Place?}>
<!ATTLIST PersInfo
  Type CDATA #IMPLIED>
<!ELEMENT Phone
  {#PCDATA}>
<!ATTLIST Phone
  Type CDATA #IMPLIED>
<!ELEMENT Place
  {PlaceName?,
  Coordinates?,
  PlaceNameVar*}>
<!ELEMENT PlaceName
  {#PCDATA |
  PlacePart}*>
<!ATTLIST PlaceName
  Method CDATA #IMPLIED

```

```

        xml:lang NMTOKEN #IMPLIED>
<!ELEMENT PlaceNameVar
  (#PCDATA|
  PlacePart)*>
<!ATTLIST PlaceNameVar
  Method CDATA #IMPLIED
  xml:lang NMTOKEN #IMPLIED>
<!ELEMENT PlacePart
  (#PCDATA)>
<!ATTLIST PlacePart
  Type CDATA #IMPLIED
  Level (1|2|3|4|5|6|7) #IMPLIED>
  <!-- Level 1=country, 2=US state, 3=US county, 4=city/town, 5=postal code -->
<!ELEMENT Product
  (ProductId?,
  Version,
  Name,
  Supplier?)>
<!ELEMENT ProductId
  (#PCDATA)>
<!ELEMENT Public
  (#PCDATA)>
  <!-- content = Y|N -->
<!ELEMENT Publishing
  (#PCDATA)>
<!ELEMENT Religion
  (#PCDATA)>
<!ELEMENT RelToFath
  (#PCDATA)>
  <!-- content = biological|adopted|foster|sealing|... -->
<!ELEMENT RelToMoth
  (#PCDATA)>
  <!-- content = biological|adopted|foster|sealing|... -->
<!ELEMENT Repository
  (Link,
  CallNbr?)>
<!ELEMENT Role
  (#PCDATA)>
<!ELEMENT SameIndiv
  (Link,
  Note*)>
<!ELEMENT sub
  (#PCDATA)>
<!ELEMENT Submitter
  (Link)>
<!ELEMENT sup
  (#PCDATA)>
<!ELEMENT Supplier
  (Link)>
<!ELEMENT TempleCode
  (#PCDATA)>
<!ELEMENT Title
  (#PCDATA)>
<!ELEMENT u
  (#PCDATA)>
<!ELEMENT URI
  (#PCDATA)>
<!ELEMENT Version
  (#PCDATA)>
<!ELEMENT WhenRecorded
  (#PCDATA)>
<!ELEMENT WhereInSource
  (#PCDATA)*>
<!ELEMENT WifeMoth
  (Link,
  FamilyNbr?)>
<!-- ===== -->

```