

população  
cadernos neps  
sociedade

**SIAD`04 - Sistemas Informáticos para Análise de Dados  
Demográficos**

Guimarães, 2004

Título:

*SIAD'04*

*Sistemas Informáticos para a Análise de Dados Demográficos*

**Editores:**

Fernanda Faria  
Pedro Henriques

**Colecção:**

Cadernos NEPS 5

**Edição:**

Núcleo de Estudos de População e Sociedade  
Instituto de Ciências Sociais  
Universidade do Minho  
Guimarães/2001

**Capa**

Fotografia de Francisco Botelho

**Depósito legal:**

219091/04

**ISBN:**

972-95433-6-4

**Composição:**

JDF/NEPS

**Impressão e acabamento:**

Eden Gráfico, S.A.  
Rua dos Casimiros, 21  
Apartado 2047  
3510-061 Viseu

**Direitos reservados:**

NEPS  
Núcleo de Estudos de População e Sociedade  
Universidade do Minho  
Campus de Azurém  
4800-058 GUIMARÃES

**Edição integrada no projecto:**

*Espaços Urbanos e Rurais*  
*Microanálise de Comportamentos Demográficos*  
*Mobilidade Geográfica e Social e*  
*Dinâmicas Culturais (séculos XVI-XX)*

### Nota Introdutória

«Em determinado momento sente-se burburinho num caminho que converge para o largo onde se condensa a multidão, e vê-se investir pelo meio desta um grupo de mascarados farrapões armados de desajeitados paus que manejam no jeito de quem abre cadabulho ou espalha estrume com engajo por entre as pernas da multidão. Estes mascarados, entre os quais aparecem por vezes alguns fardados de soldados e armados de caçadeiras, ao mesmo tempo que isso fazem vão arredando a gente para a berma do caminho, a abrirem espaço para execução da lavoura.

Aparece então, guiada por outro mascarado, uma parrelha de cães a puxar um pequeno arado de pau que outro mascarado agarrado às rabiças orienta conforme pode. Os cães, que nem sempre se entendem um com o outro, fazem entre si uma bulha que resulta ficar em perigo as canelas do candeeiro, o que serve de galhofa e de entretenimento para os espectadores. Os primeiros mascarados e outros que vão aparecendo conservam a multidão afastada servindo-se dos paus ou toscos utensílios que manejam agora em jeito de quem anda a picar seitas, fazendo de vez em quando arremetidas contra os mais avançados ou alguma moça mais guapa.

Aparece então outro mascarado com um avental cheio de baganha que vai semeando pelo chão, e também, às vezes, por cima da cabeça dos espectadores.

Surge finalmente uma junta de cabras atrelada a uma pequena grade que percorre o terreno da lavoura como se andasse a agradá-lo, continuando os mascarados a manter a multidão fora do terreno onde a cena se desenrola.»<sup>[1]</sup>

O diabo regressa personificado para integrar esta manifestação festiva, quase nu, numa negritude assustadora, circulando familiarmente pelo meio da representação agrícola dos mascarados.

Saíndo de um local inesperado irrompe pela multidão, em grande correria, agarra uma rapariga que se esperneia em grande gritaria. A assistência agita-se deliciada perante este espectáculo, como par-

ticipante duma orgia excitante do caos colectivo que todos envolve..

Por entre esta caótica representação uma possante personagem encarna a figura de São Bartolomeu, coberto com uma capa encarnada e, na mão direita exhibe uma majestosa espada prateada. O diabo, quando solto das suas mãos permite-se a todos os devaneios e loucuras, provocando gritarias e a confusão na multidão que veio assistir, para depois se tornar manso e ordeiro quando o santo o acorrenta, num jogo delicioso e nunca ensaiado, tranquilizando por momentos a assistência.

Se o diabo se torna na personagem mais vistosa do conjunto, pela irreverência e pelos movimentos assustadores, outras não menos irreverentes constroem um cenário de verdadeira orgia, tal como o mascarado condutor das cabras, que de quando em vez, arranca jactos de leite jorrados das suas tetas fartas na direcção dos espectadores.

Por entre, o cortejo outras personagens, circulam em gestos que se foram incorporando, ao longo do tempo, na manifestação: um padre, benzendo os espectadores que ladeiam o trajecto com uma vassoura feita issope, mergulada em água suja suspeitando, a multidão, tratar-se de urina; um barbeiro, aqui e além para barbear um participantes do cortejo; casal de noivos que mais tarde terá casamento.

Da multidão saem ofertas de líquidos, servidos em penicos para refrescar as gargantas dos mascarados cansados e sequiosos. Tratando-se de vinho branco, logo se vêem os seus efeitos etílicos nos gestos desmedidos da comitiva dos mascarados.

A noite cai e os habitantes da freguesia acordam já entrados no novo ciclo agrário, festejando 182,5 dias depois, em 24 de Fevereiro, o carnaval, regressando novamente as mesmas máscaras para outros rituais mágicos festivos..

<sup>[1]</sup> Ob. Cit.



## Prefácio

Estando envolvidos há anos em diversas tarefas de concepção e desenvolvimento de Sistemas Informáticos para apoio a estudos em vários campos da História, nomeadamente na Demografia, sentimos fortemente tentados a aproveitar a oportunidade criada pelos Organizadores do VII Congresso da ADEH e abalançamo-nos a dinamizar uma sessão especial onde pudéssemos confrontar o nosso trabalho com o de outros investigadores ibéricos envolvidos em actividades análogas.

Convictos que da troca directa e aberta de ideias, em ambiente próprio - como é o caso deste Congresso luso-espanhol de Demografia Histórica - só pode resultar um aprofundamento dos conhecimentos e um aumento na qualidade dos trabalhos, propusemos uma sessão dedicada aos sistemas suportados em computador para o levantamento, arquivo, pesquisa e análise de dados históricos e divulgação de resultados.

Desse projecto resultou o conjunto coerente de 5 comunicações que se compilam neste volume e que, a nosso ver, pela mensagem que cada uma individualmente trás e pelo diálogo cruzado que propiciam, permitiram alcançar totalmente o fim desejado.

Incluimos ainda a Nota de Abertura que revê a evolução do binómio demografia-informática e de alguma forma sintetisa os contributos dos trabalhos que foram apresentados.

Agradecemos aos Autores o empenho demonstrado.

Aos Organizadores do Congresso, o nosso muito obrigado pela possibilidade criada e todo o apoio prestado.

Resta-nos agora esperar que a sessão seja do agrado de todos os participantes e traga uma nova luz que incentive o progresso nesta área de simbiose entre a História e a Informática.

Março de 2004

Antero Ferreira  
Fernanda Faria  
Pedro Henriques



## ÍNDICE

- 11 Antero Ferreira (Nota de Abertura) - Sistemas Informáticos para análise de dados demográficos: uma abordagem histórica.
- 21 Fernanda Faria e Pedro Henriques - Análise Espacial de BD Paroquiais: Antes e Depois da Fusão.
- 33 Carles Millás Castellví e Carmen Sisón Bernal - Una Aplicación de las Bases de Datos en Demografía Histórica: la Reconstrucción de Familias en MsAccess.
- 49 Ana Jorge, Ana Maria Rodrigues, Hermínia Vilar, Pedro Henriques e Sandra Lopes - Construção e Exploração de uma BD Prosopográfica Normalizada do Clero Catedralício Português.
- 69 Laura Lorenzo Carrascosa - Reconstrucción de Familias y Trayectorias de Vida a partir del Análisis de Datos Locales.
- 83 Albert Esteve, Angels Tonents e Clara Cortina - Proyecto IPUMS, Integrated Public Use of Microdata Series: aplicabilidad a un estudio sobre la Emigración Espanola a Florida entre 1880 y 1920.

## Sistemas Informáticos para Análise de Dados Demográficos: uma abordagem histórica

Antero Ferreira

Ao delinear esta sessão tínhamos como principal objectivo proporcionar um momento de encontro e de reflexão a todos os investigadores que trabalham nesta área de intercepção entre a Demografia e a Informática. Interessava-nos, acima de tudo, reflectir sobre os vários ensaios de tratamento das fontes, sobre os esforços de reconstrução nominativa e sobre a extracção e análise da informação das bases de dados demográficas.

### 1. A Demografia Histórica: fontes e metodologias

O nascimento da demografia histórica, em meados dos anos 50, com os trabalhos pioneiros de Pierre Goubert e Louis Henry, esteve associado à exploração de um tipo particular de fontes: os registos paroquiais, existentes na generalidade dos países europeus a partir do século XVI. O desenvolvimento de uma metodologia específica para a sua exploração, o método de reconstituição de famílias de Fleury-Henry, possibilitou o alargamento da análise demográfica para períodos anteriores aos recenseamentos modernos.

A “descoberta” dos registos paroquiais como fonte para a história da população, associada a uma metodologia rigorosa e que produzia resultados seguros, abria o caminho a um melhor conhecimento do passado, particularmente do homem comum. Esta linha de investigação teve uma enorme repercussão, pois partia de um tipo de fonte que existia, com características semelhantes, em quase todos os países cristãos o que possibilitaria a utilização das mesmas metodologias e a comparação de resultados, abrindo fecundas perspectivas de investigação. A multiplicação deste tipo de estudos por toda a Europa foi apresentando uma realidade diversa das expectativas iniciais, pois as fontes não tinham em todos os países, por variadas razões, a

mesma qualidade. Desta diversidade foram surgindo metodologias alternativas, adaptadas às especificidades das fontes, das quais destacaríamos, entre outras, os trabalhos do Cambridge Group for the History of Population and Social Structure, que, perante as limitações das fontes inglesas, desenvolveu técnicas de análise agregativa, e a Metodologia de Reconstituição de Paróquias, desenvolvida por Norberta Amorim nos anos 70 para as fontes portuguesas, que permite acompanhar em encadeamento genealógico o percurso de vida de cada indivíduo.

Sendo por natureza um campo interdisciplinar, ponto de encontro de historiadores, geógrafos, sociólogos, antropólogos, economistas, médicos, a demografia histórica foi abrindo as portas a outro tipo de fontes nominativas (testamentos, fontes fiscais e administrativas, registo de propriedades, registo dos tribunais civis e eclesiásticos, etc.) que, cruzadas com os registos paroquiais, permitiam iluminar novas dimensões da vida dos homens no passado.

Todas estas investigações, independentemente das opções metodológicas, assentam num princípio comum que poderíamos denominar, de modo genérico: o cruzamento de informação nominativa, isto é, a tentativa de reconstituir as relações existentes entre referências nominativas, respeitantes a indivíduos ou famílias, encontradas em diferentes documentos. Esta perspectiva metodológica coloca grandes dificuldades ao nível do rendimento do trabalho do investigador pois ao tempo dispendido na recolha da informação registada nos registos paroquiais, acresce o manuseamento de uma enorme quantidade de fichas para se proceder a cada um dos cruzamentos nominativos. Louis Henry estimava que eram necessários seis meses de trabalho para um investigador isolado conseguir reconstituir um período de 100 anos numa comunidade de 500



habitantes. Esta situação conduzia a uma dupla redução da investigação: privilegiava-se o estudo das pequenas comunidades e limitava-se o período de investigação.<sup>1</sup> Não é de estranhar que perante tais dificuldades os historiadores demógrafos comesçassem a interessar-se pelos avanços de outra jovem ciência, deveras promissora, a Informática.

A este propósito, no prefácio à obra de Louis Henry sobre as antigas famílias de Genebra, em 1956, Alfred Sauvy afirmava que “A automação não pode ser introduzida na exploração dos registos paroquiais e das genealogias”<sup>2</sup>. Nessa altura não seria possível imaginar, certamente, o desenvolvimento e a importância que a informática assumiu no nosso mundo actual. Ainda no início dos anos 80, trinta anos mais tarde, Jacques Dupâquier afirmava que “O custo da introdução de dados é actualmente tão elevado que tem de se pensar duas vezes antes de tomar a decisão de os tratar através de técnicas informáticas.”<sup>3</sup> Apontava nessa oportunidade como principais limitações à utilização da informática, a necessidade de formalizar os ficheiros, o tratamentos e mesmo a recolha dos dados em função dos sistemas de tratamento da informação, das linguagens e até da estrutura das máquinas. A descrição de alguns dos obstáculos que o historiador enfrentava não podem, hoje, deixar de nos fazer sorrir: “linguagens máquina desenvolvidas para serem utilizadas pelos matemáticos e físicos”, “cartão perfurado”, “banda magnética”, etc.

Esperançado com o advento da micro-informática declarava premonitoriamente: “Com os mini-computadores dotados de um teclado, de um ecrã e de uma impressora, o historiador-demógrafo pode, ele próprio, introduzir os dados e elaborar os seus programas. O material não custa sequer o preço de um automóvel, e é provável que ainda baixe mais.” (Dupâquier, 1984:130)

Se a Informática aparecia como uma solução para as dificuldades da demografia histórica, o seu incipiente desenvolvimento, o grau de especialização que exigia, os enormes investimentos envolvidos, dificultavam, na altura, a aposta nesta solução. Alguns dos obstáculos referidos pelos investigadores na relação com a informática, são bem o testemunho das dificuldades encontradas nesse período:

problemas de introdução de informação; necessidade de desenvolvimento de software específico, o que colocava o historiador na dependência do programador; linguagens de programação que não foram desenvolvidas para o tratamento de informação nominativa e, finalmente, dificuldade de acesso aos próprios equipamentos que eram raros e caríssimos.

Perspectivando a evolução desta relação, abordaremos de seguida as principais soluções de recolha e tratamento da informação demográfica com recurso à tecnologia informática.

## 2. A Introdução de Informação

O primeiro problema a enfrentar pelo investigador ao enveredar pela utilização da informática é o da introdução da informação no computador. A digitalização da documentação com vista ao reconhecimento óptico da informação manuscrita é, ainda hoje, um sonho. Optou-se, portanto, pela transcrição do documento, num processo que em nada se distinguiu do manual. O objectivo final era a constituição de uma ficha padronizada que representava a informação essencial do registo original.

Os investigadores estavam, por analogia com a recolha manual, conscientes da importância desta etapa. Antoinette Fauve Chamoux, em 1972, afirmava que “Para um ficheiro ser rentável, ele tem que antes de tudo ser bem feito. (...) Se a transcrição não é rigorosa e estável, muitas informações serão eliminadas ou mal interpretadas. Os riscos de erro na perfuração juntam-se àqueles cometidos na altura da recolha.”<sup>4</sup> Esta recomendação, ainda hoje muito pertinente, recorda-nos a importância decisiva do acto, na maior parte das vezes único e irrepitível, de transcrição da informação do documento. Recorde-se que, na maior parte dos casos, estamos perante uma dupla transcrição, o que traz ainda mais dificuldades: do documento original para uma ficha em papel; desta para a ficha informática. Para obviar a estas dificuldades A. Chamoux dava-nos notícia das vantagens do sistema FORCOD, desenvolvido por M. Couturier e R. Abehassera, na hora da recolha da informação:

*“Munido de um gravador de som, o investigador dita o conteúdo do seu documento tal qual ele se apresenta, sem se preocupar com a ordem dos elementos. (...) Como em qualquer ficheiro, a unidade de base é a ficha, mas esta não terá nem estrutura nem dimensão fixas e será baptizada como ‘artigo’.”* (CHAMOUX, 1972:1084). Numa abordagem surpreendentemente actual<sup>5</sup> informa-nos que o registo sonoro do documento será pontuado com definidores (por exemplo: «casamento», ou «sepultura») e com descritores (por exemplo: «data do casamento»), podendo o investigador criar livremente os descritores que entender. A informação assim recolhida era posteriormente introduzida no sistema informático por um operador.

Este processo de transcrição da fonte documental para o suporte informático é, sem dúvida, um dos aspectos críticos de qualquer sistema. Para além da dificuldade de representar num modelo informático a riqueza e diversidade de documentos que, em regra, não são normalizados o investigador enfrenta ainda o problema da codificação e da introdução dos dados. As possibilidades de erro durante estas operações são muito elevadas, pois trata-se de processos mecânicos e repetitivos, e dificilmente detectáveis no futuro.

É portanto fundamental que o investigador defina claramente o âmbito da sua pesquisa e efectue uma análise exaustiva da informação disponibilizada pela fonte, de modo a recolher o máximo de informação possível numa única operação de consulta.

Um outro aspecto, muitas vezes descurado mas fundamental para determinar o sucesso de um sistema informático, é o interface com o utilizador. É fundamental que ele se aproxime das lógicas habituais de trabalho do investigador, para garantir a sua aceitação, que seja funcional e que proporcione automaticamente algumas validações de coerência básicas (por exemplo: data de óbito não pode ser inferior à data de nascimento, etc.)

Nos dias de hoje dispomos de outras possibilidades nem sempre exploradas! Podemos transportar o computador para junto dos documentos, introduzindo os dados directamente no sistema, reduzindo as operações de transcrição e

consequentemente as hipóteses de erro; poderíamos recuperar o processo de transcrição sonora do documento, apostando em técnicas de reconhecimento de voz; poderíamos ainda investir na digitalização dos documentos, armazenando na BD, para além da informação normalizada, a imagem do documento. Esta última possibilidade, cada vez mais exequível graças ao desenvolvimento tecnológico e à diminuição de preço do hardware de armazenamento, tem a enorme virtualidade de possibilitar a qualquer momento o regresso ao documento, diminuindo assim as hipóteses de erro.

### 3. O Cruzamento de Informação Nominativa

Nas palavras de Charbonneau e Légarè: “O cruzamento nominativo é o processo de pesquisa e de reagrupamento dos elementos relativos a uma mesma unidade (indivíduo, casal, família). Repousa essencialmente nas relações de parentesco, que associam os nomes entre si e permitem identificar os indivíduos e casais.”<sup>6</sup>

Nos trabalhos pioneiros com aplicações informáticas já se notavam diferentes tipos de abordagem a esta problemática: alguns investigadores acentuavam a perspectiva do cruzamento automático da informação, outros apostavam mais numa reconstrução semi-automática, em que o investigador reserva para si um papel fundamental. Este percurso deixou-nos importantíssimos contributos teóricos que convém recordar pela actualidade que ainda hoje se revestem.

Todos temos presente as dificuldades que a especificidade das fontes utilizadas pela demografia histórica aportam ao cruzamento nominativo: o caso dos homónimos, as variações ortográficas dos nomes, a irregularidade na transmissão e utilização dos apelidos, a identificação incompleta dos indivíduos e dos laços de parentesco nos registos paroquiais, etc.

Perante tais dificuldades a resposta mais óbvia seria a de que o computador jamais conseguirá, de forma automática, reconstruir estas relações – este papel estará sempre reservado ao historiador. Neste sentido, Charbonneau e Légarè afirmavam que “os computadores não podem ser tomados como

*uma panaceia. (...) (as máquinas) não suprimem todo o trabalho manual, especialmente quando se trata de resolver os casos muito ambíguos.*” (Charbonneau, 1979:98). Em sentido inverso, cabe aqui recordar o desafio de Roger Schofield: “se o raciocínio do historiador tem alguma pretensão de respeitabilidade intelectual, os princípios em que se baseia têm de ser especificáveis em forma algorítmica e, portanto, ser executáveis automaticamente pelo computador, sem qualquer intervenção humana.” (Schofield, 1992:75)

Grande parte das tentativas de reconstituição automática de famílias baseia-se, directa ou indirectamente, nos trabalhos pioneiros de H. B. Newcombe que, num famoso artigo de 1959, intitulado “Automatic Linkage of Vital Records”, apresentava uma série de hipóteses muito sugestivas. Para ele: “O registo dos vários factos respeitantes a um indivíduo efectuado rotineiramente nas sociedades contemporâneas constituiria, se reunido, uma história extensivamente documentada da sua vida.”<sup>7</sup> O principal obstáculo para atingir esse objectivo era, na sua opinião, o elevado custo da pesquisa manual de um elevado número de documentos em ficheiros de grandes dimensões. A mecanização destas pesquisas poderia ser efectuada se se demonstrasse que as máquinas poderiam realizar as ligações entre registos de um modo rápido, barato e com suficiente precisão.

Newcombe estava interessado em analisar a fertilidade diferencial das famílias, tendo em consideração a presença ou a ausência de doenças hereditárias. O processo que desenvolveu procurava ligar automaticamente os registos de nascimento ao registo de casamento correspondente. O ficheiro de casamentos era pesquisado em busca da ocorrência dos nomes/apelidos dos progenitores, previamente codificados foneticamente segundo o Russel Soundex Code<sup>8</sup> para evitar os problemas com diferentes grafias. Para possibilitar ao computador decidir qual casamento correspondia a um determinado nascimento eram utilizadas outras variáveis (o nome completo da mãe e do pai, as suas naturalidade, idades, ocupações, etc.). A cada concordância era atribuída uma pontuação positiva, a cada discordância uma pontuação negativa. O nascimento

era relacionado com o registo de casamento que obtivesse a pontuação mais elevada – este é, sem dúvida, o aspecto mais inovador desta proposta.

Os projectos de reconstituição automática de famílias mais conhecidos, o desenvolvido pelo Cambridge Group na Inglaterra e o da Universidade de Montreal, no Canadá, baseiam-se nos princípios enunciados por Newcombe.

O sistema desenvolvido no Cambridge Group, pela equipa de Wrigley e Schofield, surge associado à tentativa de reconstrução dos registos de paróquias urbanas, ou de grupos de paróquias, que, pela grande quantidade de registos envolvidos, desaconselhava a reconstrução de famílias pelos processos tradicionais. A sua grande aposta metodológica centrou-se na clarificação da natureza das decisões que o investigador assume no estabelecimento de relações entre dois registos nominativos, de modo a construir um algoritmo que o computador pudesse utilizar com o mesmo grau de sucesso que o investigador.

Um dos principais problemas que o sistema tinha de enfrentar era o das variações ortográficas, que teria de ser previamente resolvido pelo investigador de modo a que os registos que se referem à mesma pessoa pudessem ser relacionados. Esta operação, mesmo que realizada pelo computador, terá sempre de ser assistida pelo investigador, pois exige um profundo conhecimento das fontes e do meio sobre o qual incide a investigação.

Admitindo que se consegue ultrapassar eficazmente o problema das variações ortográficas, um outro importante problema se coloca, o do relacionamento de registos que se referem ao mesmo indivíduo, dadas as possibilidades de existência de indivíduos homónimos. No estabelecimento destes relacionamentos o Grupo de Cambridge recorre à proposta de Newcombe, tentando resolver as ambiguidades existentes com recurso à comparação de todos os identificadores disponíveis que são pontuados em função da existência ou ausência de informação concordante. O sistema aceitará como verdadeira a relação que obtiver maior pontuação.

Os testes de controlo efectuados permitiram aos investigadores concluir que não havia significativas diferenças entre os resultados da reconstituição

automática e os da reconstituição manual, embora os primeiros sejam obtidos em muito menos tempo. Contudo, Roger Schofield diminui bastante o nosso entusiasmo ao afirmar que “é importante enfatizar o grande dispêndio de tempo nas tarefas de preparação das fontes para introdução no sistema e para o cruzamento da informação.” (Schofield, 1992:78)

O projecto canadiano, “Programme de recherche en démographie historique” (PRDH), desenvolvido pela equipa de Hubert Charbonneau e Jacques Légarè a partir de 1968, foi o primeiro a apresentar resultados significativos aplicando a reconstituição automática. Contudo, reconhecia logo à partida que esta não podia resolver os casos mais ambíguos e que teria de ser completada manualmente numa fase final. Salientava ainda, tal como Schofield, que as operações manuais de preparação da informação eram muito prolongadas. Os procedimentos informáticos, semelhantes aos descritos para Grupo de Cambridge, eram realizados em três etapas: a triagem – subdivisão do ficheiro em subgrupos interligando os elementos mais susceptíveis de pertencer a uma mesma unidade; o acasalamento – aproximação dos elementos reagrupados pela triagem; o emparelhamento – comparação dos elementos acasalados e decisão sobre qual dos pares formados é o mais adequado<sup>9</sup>. Um outro projecto canadiano, o BALSAC, desenvolvido pela Universidade do Québec a partir de 1971, representa também um importante esforço nesta direcção da reconstituição automática, situando-se em termos teóricos dentro dos pressupostos já apresentados. Num artigo em que apresenta as linhas gerais deste projecto, Gérard Bouchard salienta os constrangimentos que o objectivo da pesquisa representa para os procedimentos da reconstituição automática. Um estudo médico ou genealógico, por exemplo, não se satisfaz com cruzamentos baseados em critérios estatísticos. Nestes casos, o cruzamento tem de estar baseado na verificação de um conjunto rigoroso de critérios, reduzindo assim, obviamente, o número de cruzamentos efectuados, mas aumentando o seu rigor<sup>10</sup>.

O grande desenvolvimento da micro-informática, particularmente nos anos 90 do século passado, colocou nas mãos dos historiadores um

hardware acessível, muito poderoso e ferramentas informáticas cada vez mais próximas do utilizador (bases de dados, folhas de cálculo, ferramentas estatísticas). Esta “revolução” libertou o historiador da dependência de grandes sistemas informáticos, forçosamente muito rígidos e dispendiosos, e permitiu-lhe adaptar, com ou sem a colaboração de técnicos informáticos, estas ferramentas às suas investigações<sup>11</sup>.

Neste processo ganham especial relevo os sistemas de gestão de bases de dados (SGBD) (entre outros o dBase e, mais tarde, o Access) que, pelo seu poder e simplicidade, permitem que sejam os próprios historiadores a criar as suas bases de dados, ou a participar activamente no seu desenvolvimento. Ao longo da sua evolução, estes SGBD vão permitindo o armazenamento de um volume cada vez maior de informação, inclusivamente de imagens. O seu verdadeiro poder revela-se na hora de analisar a informação, pois a flexibilidade do seu sistema de consultas não limita o investigador a uma série de questões pré-determinadas. Francisco Ramírez Gámiz, num interessante artigo em que apresenta a adaptação informática que desenvolveu para a reconstrução de famílias nesta região de Granada, indica-nos claramente o novo papel que o investigador atribui à informática ao referir que o computador é um “poderoso aliado” do investigador: “Ainda que numa primeira fase o computador proporcione importante ajuda para armazenar e apresentar a informação, é quando nos dispomos a analisá-la e a manejá-la que a informática revela a sua maior utilidade” [Ramírez Gámiz, 2000:3].

O exemplo português é sintomático dos avanços que a micro-informática trouxe à demografia histórica. Desde 1986 que Norberta Amorim tem desenvolvido trabalhos de reconstituição de paróquias com recurso à informática. A metodologia desenvolvida por esta investigadora, partindo do cruzamento dos registos de baptizados, casamentos e óbitos, apresenta como último objectivo acompanhar o percurso de vida de cada residente da comunidade em encadeamento genealógico. As bases de dados assim constituídas são passíveis de cruzamento com qualquer outro tipo de fonte nominativa que aos mesmos indivíduos se reporte. Os passos

iniciais de Norberta Amorim foram dados com o desenvolvimento de um sistema informático recorrendo ao dBase III Plus, em colaboração com investigadores do Departamento de Informática da Universidade do Minho: Luís Lima, Cecília Moreira e Pedro Henriques. Os resultados obtidos ao nível do rendimento do trabalho foram tão significativos que esta adaptação informática da sua metodologia passou a ser a base de uma vasta série de trabalhos de investigação.

Este sistema informático, na sequência da metodologia manual desenvolvida por Norberta Amorim, não segue os passos clássicos de Henry. O investigador introduz os vários eventos, começando pelos nascimentos, mas cria logo nesse passo todas as relações familiares. Num registo de nascimento típico cria três fichas de indivíduos (o filho, o pai e a mãe) e uma família, no caso em que esta família ou estes indivíduos, pai e mãe, ainda não existem na base de dados. Este procedimento, em que a introdução de dados e o cruzamento são feitos simultaneamente, representa um ganho em termos de tempo e em termos de segurança na reconstituição. Em termos de tempo, porque com a facilidade de pesquisa que o sistema permite a identificação é, em regra, muito rápida, e porque o maior dispêndio de tempo se verifica no primeiro nascimento, nos seguintes só acrescentamos informações relativas ao novo filho e verificamos se houve alguma alteração relativa à família. Em termos de segurança, porque a reconstituição é feita num único passo, no momento em que temos a fonte na mão, o que limita as possibilidades de erro em sucessivas transcrições.

A outra virtualidade deste sistema advém do facto de centrarmos a reconstituição no indivíduo e não na família. Esta abordagem permite que sejam introduzidas na base de dados todos os indivíduos referidos nos registos, independentemente de terem ou não relações familiares na paróquia<sup>12</sup>. Deste modo, recolhemos informações que serão muito úteis no sequente processo de cruzamento das bases de dados de várias paróquias, considerando a grande mobilidade dos indivíduos nas sociedades do passado. Por outro lado, possibilita-nos também o cruzamento nominativo com todo o tipo de fontes em

que esse indivíduo seja referido.

Estes sucessos e os posteriores desenvolvimentos dos Sistemas de Gestão de Bases de Dados Relacionais tornaram o historiador demógrafo progressivamente mais ambicioso: rompeu com as barreiras cronológicas tradicionais da disciplina ao alcançar nas suas investigações o presente das comunidades em análise e abriu estimulantes perspectivas para o cruzamento inter-paroquial, tendendo para a formação de uma base de dados central.

A internacionalização deste sistema de recolha e tratamento de informação demográfica tem vindo a ser ensaiada em Espanha, nomeadamente na Galiza com o trabalho de Delfina Rodríguez Fernández, no Brasil e na Argentina.

Este projecto que aqui, de um modo muito simples, descrevemos, é um exemplo dos novos processos de trabalho que estão ao dispor do investigador nesta área da demografia histórica. Já não estamos a falar de enormes sistemas informáticos, sobre os quais se discutia o custo/hora de utilização dos laboratórios de uma qualquer universidade. Estamos perante programas que podem ser utilizados num sistema pessoal, desenvolvidos em SGBD relacionais, com uma estrutura de dados estandardizada. Trata-se de sistemas abertos, nos quais o investigador pode facilmente introduzir uma nova variável de análise ou uma nova hipótese de trabalho.

Actualmente o investigador social tem também à sua disposição uma série de ferramentas que lhe permite proceder de forma autónoma à exploração da informação organizada através dos SGBD. De facto, se estes permitem realizar a maior parte das contagens e até alguns cálculos estatísticos, não têm, obviamente, o desempenho de ferramentas desenvolvidas especificamente para esse efeito.

Ao longo das duas últimas décadas do século XX, dois tipos de aplicações desempenharam este papel: as folhas de cálculo (como o Lotus e o Excel) e os pacotes estatísticos (como o SAS ou o SPSS). Este tipo de ferramentas, integrando-se perfeitamente com os SGBD, e apresentando grande simplicidade de utilização, coloca na mão do historiador um poder de análise que depende mais da sua ima-

ginação do que das limitações do software.

A Internet veio neste campo, como em tantos outros, abrir novas possibilidades aos vários grupos de investigação. Gostaríamos aqui de abordar um aspecto em particular: o da difusão das bases de dados genealógicas. A importância da informação recolhida pelo historiador-demógrafo não se limita à sua disciplina! As comunidades desejosas de conhecer as suas raízes, nomeadamente os genealogistas amadores, têm grande interesse na consulta destes dados. A Internet fornece-nos o meio de proporcionar-lhes o acesso à informação. Existem já alguns exemplos desta orientação sobre os quais devíamos reflectir, até pela utilidade social que representam: no Canadá o “Programme de recherche en démographie historique” ([www.genealogy.umontreal.ca](http://www.genealogy.umontreal.ca)) e o Projecto BALSAC ([www.uqac.ca/balsac](http://www.uqac.ca/balsac)), na Inglaterra, o Projecto de Macfarlane sobre Earls Colne ([linux02.lib.cam.ac.uk/earlscolne](http://linux02.lib.cam.ac.uk/earlscolne)), em Portugal o projecto do NEPS, ([www.eng.uminho.pt/~neps](http://www.eng.uminho.pt/~neps)).

#### 4. Conclusão

A informática é uma ciência com um papel fundamental no trabalho do investigador social. Não é, contudo, nenhuma panacea milagrosa! Aliás, pela facilidade de tratamento de grandes quantidades de informação, pode até tornar-se uma fonte de problemas quando o investigador não delimita concretamente o seu campo de trabalho. Quando bem utilizada, pela capacidade de análise que proporciona, desafia-nos a alargar os horizontes de investigação, a enriquecer a reflexão metodológica e, conseqüentemente, a colocar novas interrogações para melhor compreender a complexidade do homem.

A evolução da relação da informática com a demografia histórica, que aqui expusemos de forma breve e pouco exaustiva, é um exemplo de sucesso. Parece-nos que esse sucesso radica em parte no facto de a demografia histórica dispor, desde Henry, de uma metodologia segura para a reconstituição nominativa e da importância que a dimensão quantitativa tem nesta área das ciências sociais. Contudo, a evolução mais recente da

informática permitiu-lhe também responder ao desafio do tratamento de dados qualitativos, sedimentando ainda mais o seu importante contributo para a demografia.

A apresentação que fizemos da evolução de alguns dos principais sistemas informáticos de recolha e análise de informação demográfica, pretendeu acima de tudo recordar o percurso que nos conduziu até hoje e a diversidade de soluções avançadas para dificuldades que ainda não ultrapassamos. Para além das diferentes opções teóricas que fomos produzindo, gostaríamos de destacar neste percurso duas grandes fases: a primeira, a dos grandes projectos, alicerçados no desenvolvimento de grandes sistemas informáticos centralizados, envolvendo grandes recursos humanos e financeiros – é o caso dos projectos do Canadá, de Cambridge e, de certo modo, do INED em França; a segunda fase, a dos pequenos grupos de investigação que, graças aos recentes desenvolvimentos da informática, centrados no utilizador, recorre a software de utilização genérica para, em pequenas equipas, prosseguir os seus projectos de investigação – é o caso de alguns projectos referidos em Espanha, Portugal, Dinamarca, Noruega, etc..

O trabalho com estas ferramentas informáticas, acessíveis a qualquer investigador, facilita o intercâmbio de experiências e até de dados entre diversos projectos. O confronto entre diversas soluções para problemáticas semelhantes conduz, necessariamente, a um enriquecimento da investigação. É esta, sem dúvida, a principal motivação desta reunião.

## Bibliografia

- AMORIM, Maria Norberta. “Rebordões e a sua População nos séculos XVII e XVIII. Estudo Demográfico.” Lisboa: Imprensa Nacional Casa da Moeda, 1973.
- AMORIM, Maria Norberta. “Evolução Demográfica de três paróquias do Sul do Pico.” Lisboa: Universidade do Minho - Instituto de Ciências Sociais, 1992.
- BEAUCHAMP, Pierre, CHARBONNEAU, H., DESJARDINS, B., and LÉGARÉ, J. «La reconstitution automatique des familles: un fait acquis.» Population Número Especial (1977): 374-399.
- AMORIM, Norberta et al. “Reconstituição de paróquias e formação de uma base de dados central.” VI Congresso da ADEH, Castelo Branco (2001):
- AMORIM, Maria Norberta; LIMA, Luís. “Demografia Histórica e Micro-Informática uma experiência sobre uma paróquia açoreana.” Boletim do Instituto Histórico da Ilha Terceira XLIV (1986):
- AMORIM, Maria Norberta. “Guimarães de 1580 a 1819. Estudo demográfico.” Lisboa: INIC, 1987.
- AMORIM, Maria Norberta. “Informatização normalizada de arquivos, reconstituição de paróquias e história das populações.” Lisboa: Universidade do Minho - Instituto de Ciências Sociais, 1995.
- BOUCHARD, G., and RAYMOND, R.; OTIS, Y. “Registre de Population et Substitutions Nominatives.” Population 3 (1985): 473-490.
- BOUCHARD, G. “Current issues and new prospects for computerized record linkage in the province of Quebec.” Historical Methods 25.2 (1992): 67-74.
- CHÁCON JIMÉNEZ, F., and GARCIA GONZÁLEZ, F. “Informe sobre el encuentro internacional “Família, Genealogía e Informática”.” Boletín de la Asociación de Demografía Histórica X.3 (1992): 143-148.
- CHAMOUX, A. F. “La reconstitution des familles: espoirs et réalités.” Annales ESC 27.4-5 (1972): 1083-1090.
- CHARBONNEAU, H., and LÉGARÉ, J. “O Uso de Computadores em Demografia Histórica.” Demografia Histórica Ed. Marcílio, Maria Luísa S. Paulo: Vozes, 1979. 97-118.
- DAUBEZE, Y.; PERROT, J.-C. “Un programme d’étude démographique sur ordinateur.” Annales ESC 27.4-5 (1972): 1047-1070.
- DUPÂQUIER, Jacques. “Pour la demographie Historique.” Paris: PUF, 1984.
- FÉLIX, Rafael, FARIA, Fernanda, SANTOS, Maribel, and HENRIQUES, Pedro R. «XML na Demografia Histórica: anotação de registos paroquiais.» Braga: Universidade do Minho, 2002.
- FERREIRA, Antero. «Demografia Histórica e estudos genealógicos.» Boletim Informativo do NEPS.27 (2002): 1-2.
- FERREIRA, Antero. “Sistema de Aquisição de Dados para a Reconstituição de Paróquias.” Braga: Universidade do Minho, 2001.
- FLEURY, M.; HENRY, L. “Nouveau manuel de dépouilement et d’exploitation de l’état civil ancien.” Paris: INED, 1965.
- FURE, Eli. “Interactive Record Linkage: The cumulative construction of Life Courses.” Demographic Research 3 (2000):
- MORRIS, R. J. “Qualitative to Quantitative by way of coding and nominal record linkage. The search for the british middle class.” History of Computing 11.1 e 2 (1999): 9-30.
- NEWCOMBE, H. B. et al. “Automatic Linkage of Vital Records.” Science 130.3381 (1959): 954-959.
- RAMÍREZ GÁMIZ, Francisco. “Ordenadores e microanálisis nominativo.” Scripta Nova - Revista Electrónica de Geografía y Ciencias Sociales.65 (2000).
- REHER, David. “The history of the family in Spain: Past development, present realities, and future challenges.” History of the Family 3.2 (1998): 125-137.
- REHER, D. et al. “Processos de modernización y trayectorias de vida. Propuesta para el análisis sociodemográfico a partir de datos locales en España.” VI Congresso da ADEH, Castelo Branco (2001):
- SCHOFIELD, Roger. “Automatic family reconstitution.” Historical Methods 25.2 (1992): 75-80.
- SCHOFIELD, R. “La reconstitution des familles par ordinateur.” Annales ESC 27.4-5 (1972): 1071-1082.
- SCHOFIELD, Roger. “History, Computing and the emergence of the Population History of England, 1541-1871: a reconstruction.” History and Computing 11.1 e 2 (1999): 79-102.
- SCOTT, Ana Sílvia. “O desafio do cruzamento nominativo a partir de fontes portuguesas.” Boletim Informativo do NEPS.3 (1998): 5-12.
- WRIGLEY, E. A.; SCHOFIELD, Roger. “Identifying people in the past.” Londres: Edward Arnold Publishers, 1973. 1

## Notas de Rodapé

- 1 Sobre estas limitações metodológicas ver Jacques Dupâquier (1984). Pour la demographie Historique. Paris: PUF.
- 2 Citado por H. Charbonneau e J. Légarè. “O Uso de Computadores em Demografia Histórica.” Demografia Histórica Ed. Marcílio, Maria Luísa S. Paulo: Vozes, 97-118.
- 3 Jacques Dupâquier (1984). Pour la demographie Historique. Paris: PUF.
- 4 Antoinette Chamoux (1972). La reconstitution des familles: espoirs et réalités. Annales ESC, 27(4-5), 1083-1090.
- 5 Note-se a semelhança entre esta abordagem e a da anotação de documentos em XML (FÉLIX et al, 2002).
- 6 H. Charbonneau e J. Légarè (1979). O Uso de Computadores em Demografia Histórica. S. Paulo: Vozes, p. 100
- 7 H. B. Newcombe; J. M. Kennedy, S. J. Axford & A. P. James (1959). Automatic Linkage of Vital Records. Science, 130(3381), 954-959.
- 8 O Russel Soundex Code é um algoritmo desenvolvido originalmente no início do século XX por Margaret K. Odell e Robert C. Russel no U.S. Bureau of Archives para simplificar o tratamento dos recenseamentos. Traduz-se o nome por um código fonético de forma a poder ultrapassar o problema de diferentes grafias do mesmo nome. O próprio Henry desenvolveu também um sistema de transcrição fonética de nomes e outros exemplos podem ser encontrados nos trabalhos desenvolvidos por David Reher para Espanha.
- 9 H. Charbonneau e J. Légarè (1979). O Uso de Computadores em Demografia Histórica. S. Paulo: Vozes.
- 10 BOUCHARD, G. (1992). Current issues and new prospects for computerized record linkage in the province of Quebec. Historical Methods, 25(2), 67-74.
- 11 Na bibliografia apresentamos um conjunto de títulos representativo deste processo em países como Portugal, Espanha, Noruega, Dinamarca, Brasil.
- 12 A todos os indivíduos é atribuída uma classificação que indica o conhecimento que possuímos dele no que respeita a actos vitais, por exemplo: nascido na paróquia, falecido na paróquia, etc.





## Análise Espacial de BD Paroquiais: antes e depois da Fusão

Fernanda Faria<sup>1</sup>  
Pedro Rangel Henriques

### Resumo

O objectivo deste artigo é mostrar as potencialidades da exploração dos dados demográficos recolhidos nas paróquias segundo o método de Norberta Amorim, MRP, através de um sistema de Informação Geográfica que permita analisar o comportamento social numa paróquia ou entre paróquias de uma zona ao longo dos séculos.

Tal exploração pressupõe a aquisição dos dados que se extraem dos registos de baptizado, casamento e óbito e o seu arquivo em bases de dados normalizadas, trabalhando individualmente com as BD paroquiais e uma BD central que segue a mesma estrutura.

### 1 Introdução

A Demografia Histórica (DH) é um dos campos de investigação da História e estuda os principais fenómenos demográficos num espaço temporal de algumas centenas de anos até à actualidade. Em Portugal, esse estudo pode ser feito a partir de meados do século XVI, dado que a sua principal fonte de informação encontra-se nos registos paroquiais (RP) — Registos de Baptizado, Registos de Casamento e Registos de Óbito — cujo início da sua sistematização ocorreu por essa altura.

Os RP serviram de base à Metodologia de Reconstituição de Famílias, desenvolvida por Louis Henry e Michel Fleury, e posteriormente à Metodologia de Reconstituição de Paróquias (MRP) concebida por Norberta Amorim, professora e investigadora do Núcleo de Estudos da População e Sociedade (NEPS). Esta última metodologia permite a reconstituição de uma paróquia pela identificação de cada indivíduo e das famílias com que se relaciona através do cruzamento dos dados constantes nos três tipos de RP [Amo91]. A partir da reconstituição da vida dos indivíduos e das famílias é possível analisar indicadores demográficos de Natalidade, Fecundidade, Nupcialidade, Mortalidade e Mobilidade.

É na sequência da cooperação entre o NEPS e

docentes e investigadores do Departamento de Informática (DI) da mesma universidade que surge a idealização do Sistema para o Estudo da Evolução Demográfica (SEED), que pretende dar suporte informático aos processos de Análise Demográfica dos dados recolhidos segundo o MRP. Na figura 1 é apresentada a arquitectura do SEED que compreende todos os processos desde a aquisição até à divulgação dos dados [Rod00].

No esquema da figura 1 identificam-se 5 camadas:

- Camada de Aquisição, inclui a interface para introdução e validação das fichas de indivíduos e famílias (após o levantamento dos RPs) e o sistema de armazenamento (BDPs).
- Camada de Fusão, destinado a fazer a adição dos dados de uma paróquia à BDC identificando indivíduos semelhantes e resolvendo colisões;
- Camada de Armazenamento Normalizado, consiste no próprio repositório de dados normalizados de forma a poderem ser utilizados pela camada de Análise, concentrando as diversas paróquias estudadas;
- Camada de Análise, para a realização de diversos tipos de análise sobre a informação obtida dos RPs eventualmente cruzada com outras fontes documentais;

• Camada de Gestão do Conhecimento, para criação e manutenção de um repositório do co-

nhecimento adquirido, incluindo as funções de armazenamento, verificação-validação e divulgação.

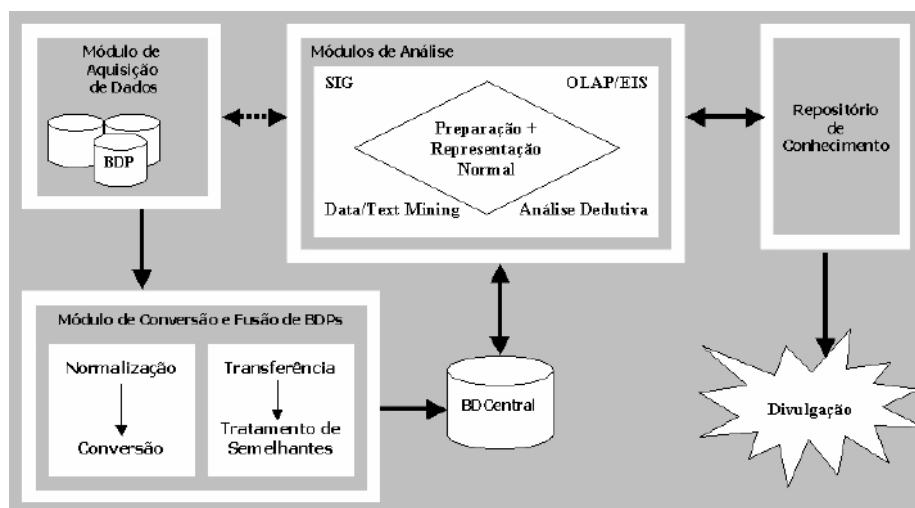


Figure 1: Arquitectura do Sistema SEED

A exploração dos dados, questão central desta comunicação, é realizada recorrendo a ferramentas EIS de Análise Multidimensional (OLAP) como o TM1 ou o SAS, ferramentas de Análise Estatística como o SPSS, ou ferramentas de Data Mining como o Clementine; geralmente os resultados são visualizados em forma de tabela ou gráfico. Mas pela natureza dos RP esta informação contém uma referência de localização pelo que muitas análises podem ser visualizadas de uma outra perspectiva — Perspectiva Georeferenciada — geralmente fornecida por Sistemas de Informação Geográfica (SIG). Os resultados das análises obtidas dos dados georeferenciados são visualizados num mapa através de cores permitindo uma visão mais nítida e directa de todo o conjunto em estudo.

Este artigo tem como principal finalidade a reflexão sobre as potencialidades dos SIGs aplicados à DH e ao MRP em concreto, a uma paróquia ou a um conjunto de paróquias.

A próxima secção faz uma breve descrição dos SIG e apresenta as possibilidades da sua aplicação a um sistema de dados de indivíduos e famílias; enumeram-se estudos que se podem fazer ao nível da paróquia e faz-se referência a estudos interparoquiais. Entretanto, e ainda nessa secção,

referem-se o tipo de fontes e suas relações que tornam possíveis essas análises geo-espaciais.

Na terceira secção, apresentam-se as aplicações do SEED que já estão desenvolvidas e em teste — o SRP que engloba todo o processo de aquisição dos dados para carregamento das Bases de Dados Paroquiais (BDP); o ConFusBDP que é responsável pela conversão (normalização) e fusão das BDPs na Base de Dados Central (BDC) — que vão permitir o estudo mono e multiparoquial georeferenciado. Serão também apresentados alguns resultados obtidos de simulações realizadas com um SIG específico (Geomedia), onde se fizeram, entre outros, análises de mobilidade.

A última secção apresenta as conclusões e algum trabalho futuro que interesse desenvolver.

## 2 Exploração Demográfica Georeferenciada

Por georeferenciação, entende-se a atribuição de uma componente localizacional à nossa informação. Estima-se que 80% da informação manipulada no nosso dia-a-dia possui uma referência espacial. Como informação geograficamente referenciada ou georeferenciada entende-se como informação relativa a qualquer entidade susceptível de ser localizada no espaço através de coordenadas geográficas. As entidades podem ser visíveis ou palpáveis, como por exemplo, edifícios, ruas e árvores. Podem ser também não palpáveis, como sejam limites territoriais (distritos, freguesias,...) e acontecimentos (acidentes, crimes, incêndios, ...). Ou seja, informação georeferenciada é aquela que é relativa a algo que possua localização espacial.

A maior parte da informação tem implícita uma componente espacial que permite localizá-la no espaço: um endereço, um país ou uma cidade e um SIG permite a sua manipulação, visualização e análise de uma forma que não é possível com a informação em formato tabular ou estatístico. O armazenamento de informação num SIG difere do armazenamento de informação numa qualquer base de dados no facto de a informação possuir uma referência espacial. Enquanto uma base de dados armazena informação que simplesmente contenha referência a uma dada localização, isto é, possui um endereço, o SIG utiliza a georeferenciação como principal forma de armazenar e aceder à informação.

Os principais conceitos de um SIG são:

- Visualização — fornece uma nova perspectiva de visualização das entidades espaciais e dos resultados das análises efectuadas;

- Criação de Mapas — permite criar mapas personalizados possibilitando ao utilizador indicar: as entidades espaciais a ver (por exemplo, colocar visível as paróquias e as igrejas e colocar invisível os concelhos, as ruas e os imóveis); a simbologia a adoptar (por exemplo, as igrejas representadas por um ponto laranja, as paróquias representadas com fundo amarelo e limite azul marinho); e a escala a usar na produção do mapa;

- Consultas topológicas — reconhece e analisa relações topológicas entre os vários elementos representados num mapa através da determinação de condições:

- de adjacência (Que freguesias se encontram adjacentes à freguesia S. José de S. Lázaro? );

- de conteúdo (Que ruas se encontram na freguesia de S. Vítor? );

- de proximidade (Que freguesias se encontram num raio de 10 km do rio Este? );

- e de intersecção (Que ruas cruzam com a Av. da Liberdade? )

- Mapas temáticos — permite efectuar todo o tipo de mapas temáticos que consistem em apresentar os elementos espaciais coloridos com uma cor determinada pelos valores que têm em relação a determinado critério ou em determinado atributo. Considerando um exemplo, descreve-se de seguida como se cria um mapa temático.

Uma Câmara Municipal pretende implementar um sistema de incentivos à natalidade nas suas freguesias. Para isso, pretende estudar a natalidade em cada uma delas a partir do número de nascimentos.

Inicialmente identificamos quais os intervalos de valores a usar na nossa análise; supondo que se utilizariam 3 intervalos, estes poderiam ser:

- até 10 nascimentos;
- de 11 a 25 nascimentos;
- a partir de 26 nascimentos.

De seguida, indica-se a cor correspondente a cada intervalo definido, podendo ser um dégradé de tons de uma só cor (várias tonalidades de uma cor, começando pela tonalidade mais clara até a tonalidade mais acentuada) ou um conjunto de várias cores. No nosso exemplo, consideremos para cada um dos intervalos de valores as seguintes cores:

- 1ª intervalo: Vermelho;
- 2º intervalo: Laranja;
- 3º intervalo: Verde.

O mapa temático é então criado colorindo as freguesias do concelho tendo em conta o número de nascimentos e o intervalo de valores em que se encontram. As freguesias com menos de 11 nascimentos estão pintadas a vermelho, as freguesias que tiveram entre 11 e 25 nascimentos encontram-se

pintadas a laranja e as restantes a verde.

Outros exemplos de mapas temáticos podem ser: mapa da natalidade em Portugal no ano de 2002, mapa dos partidos eleitorais eleitos por concelho, mapa dos restaurantes de Lisboa por tipo de cozinha (Tradicional, Francesa, Italiana, Chinesa, Espanhola, ...);

- Padrões de Comportamento — potencia o aparecimento de padrões de comportamento dada a perspectiva visual que fornece [San01]. Tomando o exemplo dos incentivos à natalidade referido no ponto anterior, a visualização do número de nascimentos por freguesias numa perspectiva geográfica permite, por exemplo, verificar que as freguesias situadas a oeste do concelho têm uma taxa de natalidade mais baixa ou que as freguesias em que o rio passa têm uma natalidade mais alta, e a partir dessas observações procurar as suas causas. Adicionando outras fontes de dados ao mapa, podem ser visualizadas características que possam ser a causa ou origem das observações atrás referidas. Por exemplo, se se visualizar o rendimento per capita médio das famílias por freguesia, poderia ser constatado que as famílias residentes nas freguesias situadas a oeste do concelho têm efectivamente um menor rendimento que as famílias residentes na outra ponta do concelho o que pode constituir uma causa para a baixa taxa de natalidade nessa zona.

Também nos RP, encontramos informação que é possível representar no mapa. No tratamento dos registos de baptismo, existe informação sobre o local de baptismo e podemos encontrar informação sobre a naturalidade e residências dos pais e padrinhos. Ao tratarmos os registos de casamento, é possível localizar o casamento e, por vezes, a naturalidade e a residência dos nubentes, das suas famílias e dos padrinhos. No caso dos registos de óbito, além do local do enterro, podemos ter referida a naturalidade e a última residência do falecido.

Os elementos georeferenciados obtidos no levantamento dos RP correspondem a paróquias, localidades, concelhos, países, etc. Estes locais podem ser hierarquizados da seguinte forma: um país contém vários distritos, um distrito contém vários concelhos, um concelho contém várias paróquias, uma paróquia contém vários lugares, e um lugar

contém vários sítios ou ruas — possibilitando o estudo georeferenciado a qualquer nível na hierarquia de locais.

Como é sabido, na DH não são os RPs que nos interessam isoladamente, mas sim a informação sobre o Indivíduo e sobre a Família que a partir deles se reconstrói, para então se calcularem variáveis que são indicadores do comportamento geral da sociedade em diferentes meios e épocas. Assim, a partir dos RPs passa a ser possível georeferenciar o trajecto de vida do indivíduo e das famílias a que este está associado e, portanto, abre-se a possibilidade de associar os indicadores sociais calculados a posições geográficas.

Importa agora identificar o valor acrescentado que esse facto — a capacidade de localizar os valores médios de fecundidade, nupcialidade, mortalidade, etc. sobre um mapa — nos pode aportar, na medida em que as facilidades resultantes de associar a BDParoquial ou Central a um SIG se traduza essencialmente numa forma diferente de ver a informação.

Para representar num mapa uma determinada variável, por exemplo a fecundidade, o que se faz é escolher uma cor numa escala em que se associam cores a intervalos de valores diferentes. Portanto se no período de tempo em estudo se pintar a paróquia P com a cor C e ao lado se apresentar a legenda com o dito código de cores, o ganho que se obtém é pequeno: uma leitura rápida do valor calculado (obtido com igual facilidade pela simples leitura do valor absoluto ou percentagem).

Este efeito visual é particularmente propício ao estabelecimento de comparações. Assim, se se colocarem lado a lado mapas com a paróquia pintada de cores diferentes em diferentes períodos de tempo ou em relação a indicadores diferentes, já o ganho é maior pois este estudo comparativo torna-se mais simples/intuitivo (mas talvez não tão intuitivo como em um comum gráfico de barras).

É óbvio que as vantagens desta forma de análise crescem se se estudarem em simultâneo várias paróquias — nesse caso é possível mostrar sobre o mesmo mapa as diferentes paróquias pintadas a cores diferentes permitindo assim, numa rápida inspecção, a extracção de resultados comparativos

entre o comportamento de determinadas zonas relativamente a um parâmetro em estudo. Torna-se pois evidente que os benefícios de um SIG aplicado à BD do SEED são distintos consoante se trabalhe numa situação mono ou multiparóquia.

Por exemplo, o estudo da taxa de natalidade de uma paróquia, analisado através de um SIG, corresponde a ver a paróquia em estudo com determinada cor. Essa cor isolada não nos diz nada. Mas se se comparar várias paróquias, cada uma com a sua cor dependendo do intervalo de valores em que a natalidade de cada uma se encontra, já se obtém uma percepção do significado das cores de cada paróquia, isoladamente ou em conjunto. Se, por exemplo, pudéssemos criar um mapa temático das várias paróquias de um concelho, no que toca à taxa de fecundidade ilegítima, poderíamos facilmente detectar a que tinha uma taxa menor — representada por uma cor de tonalidade mais suave — e a que tinha a taxa mais elevada — representada por uma cor de tonalidade mais forte.

Para a maioria das variáveis tradicionais, o estudo monoparóquia só faz sentido se for para comparar resultados em diferentes épocas. Nomeadamente, o estudo da Mobilidade só pode ser feito sobre a BD Central onde existam várias paróquias.

Para além desta análise imediatista de representação geográfica dos indicadores demográficos tradicionais, a DH abre outras hipóteses quando se relaciona informação existente nos RP com outras fontes, como registos prediais, róis de confessados, ... Seguem-se alguns exemplos de análises que nesse caso podem ser efectuadas por um SIG:

Genéricas:

- Longevidade versus Valor Patrimonial ou Dimensão da propriedade;
- Idade ao casamento versus Valor Patrimonial;
- Distribuição das casas pelo número de elementos do agregado familiar;
- Distribuição das localidades da paróquia por taxas de mortalidade
- Distribuição dos terrenos pelo tipo de cultura;
- ...

Específicas:

- Distribuição das casas com rendimentos su-

periores a 30 réis anuais;

- Acompanhamento dos membros de determinada família e seus descendentes relativamente às sucessões e aquisições;
- Distribuição das propriedades do Sr. António na paróquia;
- Localização dos terrenos com mais de 4000 metros quadrados e nome dos respectivos proprietários;
- Distribuição por herdeiros, não necessariamente descendentes, dos bens imóveis legados pelo falecido Sr. José;
- Identificação dos indivíduos residentes numa área de 250 metros da igreja paroquial;
- Distribuição dos bens imóveis dos habitantes elegíveis e respectivo valor patrimonial;
- Localização dos terrenos de cultivo da paróquia;
- ...

Sendo possível a ligação das fontes documentais cruzadas às entidades geográficas, podem-se efectuar todo o tipo de análises espaciais não possíveis com ferramentas de análise comuns. Por exemplo, a listagem das propriedades do Sr. António não fornece a mesma perspectiva que a visualização da sua distribuição. Esta última perspectiva fornece à primeira vista a ideia da quantidade, da proximidade, da adjacência e das áreas das propriedades, bem como o enquadramento espacial destas no conjunto de propriedades da paróquia.

## 2.1 Exploração Multiparóquia Georeferenciada

Um SIG quando aplicado a análises multiparóquias permite efectuar todas as que são possíveis e aquelas que se revelaram sem interesse num âmbito monoparóquia.

Assim, a visualização de análises de natalidade, nupcialidade, fecundidade e mortalidade fornecem uma visão geral de determinado conjunto de paróquias ou outro limite administrativo de nível superior facilitando claramente todos os estudos comparativos.

É possível efectuar comparações entre paróquias

do mesmo concelho ou concelhos diferentes, é possível efectuar a comparação entre uma paróquia e o concelho a que pertence, etc.

No entanto, deve-se realçar a possibilidade do estudo de um fenómeno fundamental da DH impossível de visualizar num contexto monoparóquial — a **Mobilidade** — pois, pela sua natureza, envolve sempre no mínimo um local de origem e um local de destino. A análise espacial da Mobilidade não é, de forma nenhuma, tão linear como a análise dos outros fenómenos geralmente estudados pela DH. Foram estudadas técnicas de representação deste fenómeno por um SIG. São elas:

- Por cores: O local de origem encontra-se colorido com uma cor (por exemplo, verde) e os locais de destino encontram-se coloridos com várias tonalidades de outra cor (por exemplo, vermelho). As várias tonalidades do vermelho indicariam o fluxo migratório entre os dois locais — origem e destino — correspondendo a tonalidade mais clara aos menores fluxos e a tonalidade mais escura aos maiores fluxos de migração.

- Por setas: A base da seta indica a origem do indivíduo a outra extremidade indica o destino de migração. O fluxo migratório pode ser visualizado como uma etiqueta (label) junto à seta correspondente ou na própria grossura da seta — isto é, uma seta mais fina indica um menor fluxo e uma seta mais grossa indica um maior fluxo de migração.

### 3 Aquisição, Armazenamento e Fusão de RPs

Após se ter discutido as diferentes formas em que a ligação de um SIG a um sistema de dados de indivíduos e famílias (baseado, à partida na informação recolhida dos RP, mas prevendo já o seu cruzamento com outras fontes que nos informam sobre bens imóveis e outras propriedades e pertences do indivíduo) pode, em geral, ajudar o processo de análise, pretende-se nesta secção concretizar a proposta tendo por suporte o sistema de informação construído segundo o MRP com base no SEED. Para tanto e antes de ilustrar as ideias com os resultados de estudos georeferenciados (como a mobilidade) vai-se referir o mecanismo de aquisição e construção das BDPs (SRP) e o módulo de con-

versão e fusão com resolução dos conflitos resultantes da ocorrência de indivíduos semelhantes (ConFusBDP).

#### 3.1 O SRP

O SRP supõe um processo prévio manual de levantamento dos dados dos RP e de reconstrução do indivíduo e família, fornecendo então toda a funcionalidade necessária à criação de uma BDP, isto é, suporta as tarefas de inserção, manutenção e cruzamento dos dados de um indivíduo ou de uma família.

Esta aplicação foi desenvolvida tendo sempre em vista os seguintes aspectos:

1. A aplicação deve manter o workflow usado pelos historiadores demógrafos (HD) no processo manual ou nas aplicações desenvolvidas anteriormente, como por exemplo a criação de uma família implica a associação de pelo menos um indivíduo a essa família.

2. A aplicação deve funcionar tendo como principal objectivo a correcção dos dados em detrimento da rapidez da sua introdução na BDP.

Relativamente ao primeiro ponto, os HDs podem utilizar a aplicação usando as suas funcionalidades de acordo com a sua forma habitual de trabalhar. Assim: podem variar a ordem de levantamento dos RPs — começar pelos registos de baptismo ou pelos de casamento; podem variar a forma de inserção dos dados — por exemplo, (1) inserir todos os indivíduos e criar depois as famílias com os indivíduos já existentes ou (2) criar uma família e ao mesmo tempo criar os indivíduos a ela associados.

Quanto ao segundo ponto, a aplicação obriga a que qualquer inserção, alteração ou eliminação seja explicitamente confirmada antes de ter efectivamente lugar na BDP de forma a evitar incongruências nos dados. Obviamente, a necessidade de confirmar / cancelar todas as operações traduz-se em tempo acrescido na inserção de cada RP, mas garante a correcção dos dados.

Foram construídos dois formulários principais que tratam da inserção e manutenção das principais entidades do MRP — Indivíduo e Família — e di-

versos outros formulários para tratamento de informações relacionadas com as duas entidades anteriormente referidas (informações sobre as residências, profissões, herdeiros, sufrágios, ...) ou para tratamento de informação auxiliar (listas de locais, listas de profissões, graus de parentesco, tipos de misas, ...). Far-se-á de seguida uma breve descrição

dos principais formulários do SRP, visto que manipulam os dados fundamentais e básicos de qualquer RP que constituem a base das análises.

Por ser o principal objecto de estudo da DH no contexto paroquial começa-se pela apresentação do formulário de Família ilustrado na figura 2.

**Família**

N.º Família: 8      Situação Família: Legítima

Tipo Família: 2      Famílias das quais conhecemos registo de baptismo de todos os filhos, data de casamento, data de óbito do primeiro cônjuge falecido e desconhecemos a data de nascimento da mulher.

Início Obs.: 13-06-1748  
Fim Obs.: 29-07-1756

Pai: 16959      Alberto Lino Melo  
Mãe: 19368      Maria Ana Ressurreição

C  
A  
S  
A  
M  
E  
N  
T  
O

Data: 13-02-1741      Data Fixa?       Local: Lajes do Pico - F- 460102      interVeniientes

Pai: Idade:      Estado civil:      N.º Casamentos:        
Mãe: Idade:      Estado civil:      N.º Casamentos:     

4 filhos

Nind	Nome	D. Nascimento	D. Óbito
21	José	11-11-1741	
22	Francisco	02-12-1744	
23	Luís Inácio Melo	13-06-1748	
24	Indeterminado	25-03-1751	

Residências      Obs.:

Localizar      Inserir      Alterar      Eliminar      Fechar

Figure 2: SRP – Formulário de Família

Este formulário manipula os dados específicos da entidade Família indicando o(s) indivíduo(s) que a constituíram, a sua classificação segundo o tipo e situação de família, as datas de início e fim de observação, e observações que não se enquadrem nos campos disponíveis. Caso exista o registo de casamento desta família, os dados sobre a data e local de casamento, a idade e número de casamentos anteriores de cada nubente são também registados. É também possível obter uma listagem dos filhos originados pela família em questão apresentando os seus nomes, datas de nascimento, de casamento, de fim de observação e de óbito. Sempre que a operação pretendida seja a associação ou consulta dos indivíduos de uma família, a aplicação

redirecciona-nos para o formulário de Indivíduo ilustrado na figura 3.

O formulário de Indivíduo apresenta a informação específica do indivíduo, como datas de início e fim de observação, a sua classificação e observações. Optou-se pela visualização dos seus dados organizados por três separadores, cada um relativo a cada tipo de RP — Baptismo, Casamento e Óbito — de forma a dar uma perspectiva mais organizada sobre os seus actos vitais.

O primeiro separador apresenta os dados relativos ao nascimento, como pode ser visto na figura 3, que são data e local de nascimento, filiação, idade e estado civil dos seus pais à época do nascimento.



Figure 3: SRP – Formulário de Indivíduo

O segundo separador é ilustrado na figura 4 e apresenta a listagem das relações familiares em que esse indivíduo se pode encontrar, isto é, as famílias legítimas ou ilegítimas que esse indivíduo constituiu.

Nº. Família	Situação	Data de Casamento
8	Legítima	13-02-1741

Figure 4: SRP – Separador das Relações Familiares do Indivíduo

Na figura 5 encontra-se ilustrado o terceiro separador que apresenta a informação sobre o óbito do indivíduo. Podem ser consultados ou editados dados como data e local de óbito, causa de morte, classificação paroquial, entre outros.

Todos os outros formulários da aplicação podem ser acedidos a partir dos dois principais atrás referidos.

Figure 5: SRP – Separador da informação sobre o Óbito

### 3.2 O ConFusBDP

Para se proceder a estudos multiparóquiais, devem ser considerados os seguintes processos:

1. Conversão Já existem diversas BDPs com formato distinto das BDPs obtidas através do SRP. Para que exista coerência e uniformização nos dados usados nas análises de DH, estas têm de ser convertidas para o formato de BDP usado actualmente — SRP.

2. Fusão Dado que o levantamento dos RPs é feito paróquia a paróquia, existem tantas BDPs quantas as paróquias tratadas. As BDPs devem ser aglutinadas numa só, pois o correcto tratamento analítico impõe que não haja duplicação de indivíduos e de famílias. As situações de duplicação dos indivíduos e famílias, quando não devidamente tratadas, ocorrem inevitavelmente dada a possibilidade de mobilidade entre paróquias.

De forma a auxiliar o utilizador nos dois processos referidos, foi desenvolvida uma aplicação — o ConFusBDP — que tomando uma BDP efectua a sua conversão, quando necessário, e a sua fusão para a BDC.

O processo de Conversão de BDPs é feito de forma automática depois de ter identificado manualmente a codificação a usar nos campos relativos a locais e profissões para estar de acordo com o formato usado no SRP.

O processo de Fusão de BDPs é efectuado entre BDPs com o formato SRP. Este processo pretende colocar numa só BD — na BDC — os indi-

víduos e famílias de várias paróquias, uma paróquia de cada vez, o que leva à possibilidade de se encontrarem indivíduos e famílias que já existem na BDC. Este processo efectua-se em quatro passos, são eles:

1. Transferência dos Indivíduos
2. Tratamento dos Indivíduos Semelhantes
3. Transferência das Famílias
4. Tratamento das Famílias Semelhantes

No primeiro passo, a aplicação procura na BDC indivíduos ou famílias com características muito semelhantes aos que se querem inserir; no caso de não encontrar semelhantes faz a sua inserção directa; no caso de encontrar semelhantes regista esse facto e coloca em stand-by a sua inserção.

Terminado o passo de inserção directa dos indivíduos, procede-se ao tratamento dos Indivíduos Semelhantes. O utilizador percorre todos os indivíduos semelhantes de um indivíduo a inserir para aferir se algum é efectivamente o mesmo. Se decidir que o indivíduo que pretende transferir corresponde a um indivíduo semelhante, faz a actualização dos dados do último com os dados do primeiro. Se não encontrar um indivíduo semelhante que o possa considerar o mesmo que o indivíduo a inserir, procede então à sua inserção na BDC. Na figura 6 é apresentado o formulário de Tratamento de Indivíduos responsável por apoiar o utilizador neste passo.

Tanto no primeiro passo como no segundo, sempre que é inserido um indivíduo na BDC, são também transferidos os seus dados relativos a profissões, residências e assinaturas.

**Indivíduo encontrado** Tipo Ind. 3 Início Obs. Fim Obs.

3 Isabel

Datas

Nind	Nascimento	Casamento	Obito
3	19-07-1733		12-02-1785

Locais

Nind	Nascimento	Casamento	Obito
3	460102		460102

**Foram encontrados 2 Indivíduos Semelhantes**

Nind	Nome	Tipo Ind.	Início Obs.	Fim Obs.
619	isabel st. andre	0	07-05-1745	09-10-1769
710	isabel francisca	0	14-09-1780	14-09-1780

Datas

Nind	Nascimento	Casamento	Obito
619			

Locais

Nind	Nascimento	Casamento	Obito
619			

<< < > >> Inserir Atualizar Fechar

Figure 6: ConFusBDP — Tratamento de Indivíduos Semelhantes

O terceiro passo é efectuado da mesma maneira que o primeiro mas relativamente às famílias; ou seja, procura na BDC famílias semelhantes à família que se pretende inserir. Posteriormente no quarto passo é feito o tratamento das famílias semelhantes de forma idêntica ao tratamento dos indivíduos semelhantes.

Visto que em alguns dos passos do processo da Fusão, nomeadamente no segundo e quarto passo, é necessária a intervenção do utilizador este processo não é automático mas semi-automático.

### 3.3 Alguns Resultados obtidos com um SIG

Utilizando o SIG Geomedia da Intergraph, foram já feitas algumas simulações com BDPs de algumas paróquias da ilha do Pico, Açores.

Um primeiro exemplo de resultado obtido com um SIG é a criação de um mapa temático por determinado critério. Neste caso, tomando a taxa de

natalidade, as paróquias são pintadas na cor correspondente ao intervalo de valores em que se encontram. Como se pode ver na figura 7, as paróquias usadas na análise encontram-se pintadas a determinada cor; essa cor indica qual o intervalo de valores em que as paróquias se enquadram. A legenda apresenta a escala de cores e valores usada na representação visual do resultado da análise, começando numa cor mais clara para o intervalo de valores mais baixos até a cor mais escura para o intervalo de valores mais elevados.

A percepção do resultado analítico obtido quando se utiliza um SIG é mais rico que a apresentação desses resultados em forma tabular (folha de cálculo), pois fornece uma visão conjugada do valor da taxa de natalidade com as áreas das paróquias, com as suas situações geográficas na ilha e com os factores proximidade e adjacência entre elas.

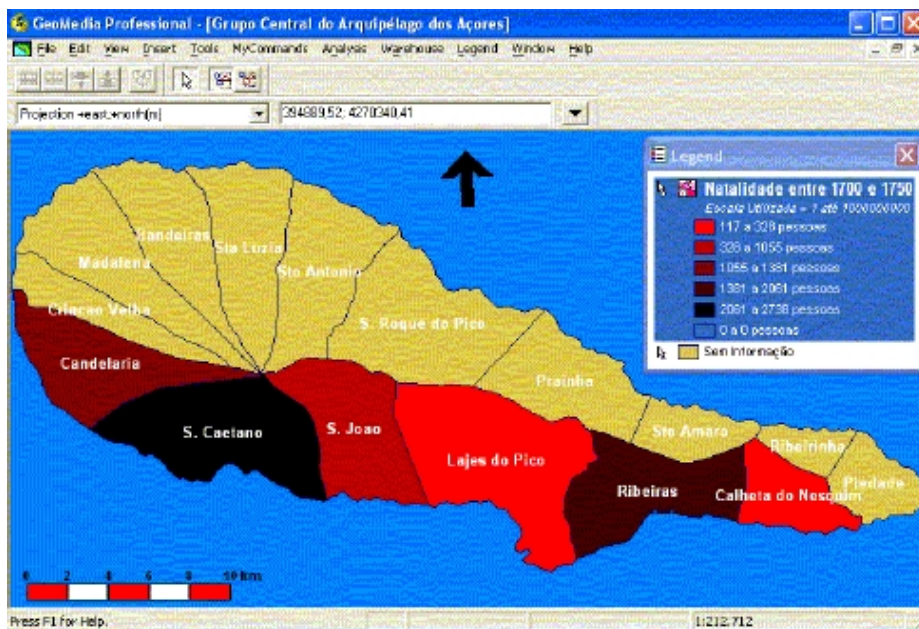


Figure 7: Mapa temático sobre a natalidade em algumas paróquias do Pico

Como já foi referido na secção 2, um outro exemplo do que se pode obter com um SIG é um estudo de mobilidade que está ilustrado na figura 8. O que se pode ver é um fluxo migratório no ano de 1720, representado pela linha, entre duas paró-

quias — Origem: Sta Luzia; Destino: Criação Velha — e um valor associado a esse fluxo — neste caso, 3 pessoas saíram de Sta Luzia para Criação Velha.

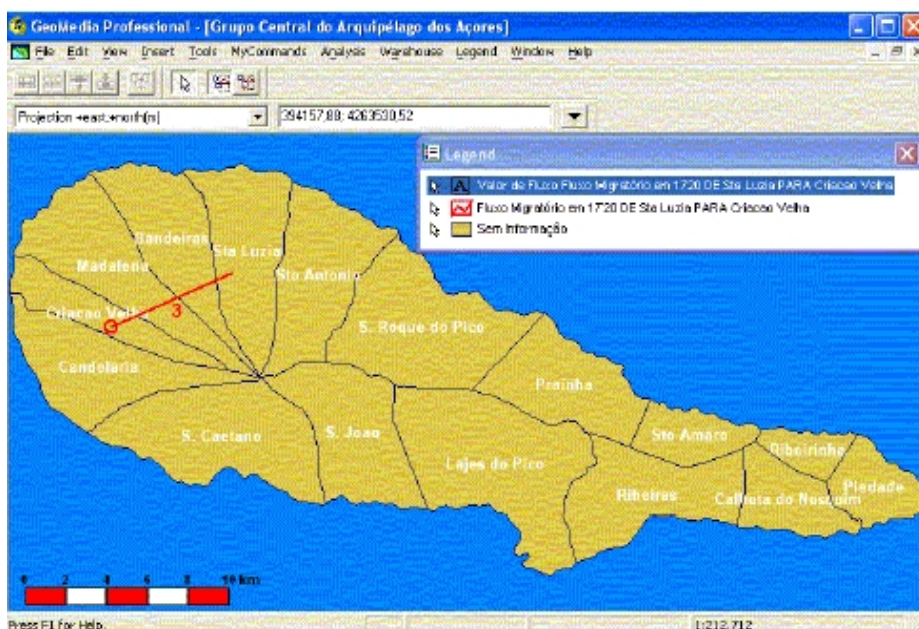


Figure 8: Exemplo de Representação da Mobilidade

#### 4 Conclusão

Neste artigo pretendemos mostrar algumas das potencialidades de um SIG aplicado à DH, referindo alguns exemplos em que este traz mais-valias às análises fundamentais desta área de estudo e outros exemplos em que não existem vantagens na sua utilização. Em conclusão, este tipo de análise deve ser feito quando os dados manipulados por um SIG podem fornecer uma perspectiva espacial intuitiva e elucidativa, não fornecida por outra ferramenta de apresentação de resultados, caso dos estudos comparativos de variáveis espaciais (áreas, posições, etc) e da mobilidade.

Foram também apresentadas as aplicações de-

sempovidas para o suporte à aquisição e armazenamento electrónico dos RPs — o SRP, Sistema de Reconstituição de Paróquias, e o ConFusBDP, Sistema de Conversão e Fusão de BDPs — constituindo as camadas iniciais da arquitectura SEED e a base para qualquer tipo de exploração e divulgação dos dados.

Uma das linhas de trabalho futuro está relacionado com o suporte à aquisição; pretende-se que o sistema permita introduzir (talvez usando técnicas de anotação estrutural de documentos) os registos na íntegra e ajude (de forma semi-automática) a fazer o seu cruzamento com vista à reconstituição paroquial.

#### Bibliografia

- [Amo91] Maria Norberta Amorim. *Evolução Demográfica de Três Paróquias do Sul do Pico*. Instituto de Ciências Sociais, Universidade do Minho, 1991.
- [Rod00] Fátima Rodrigues. *Arquitectura Heterogénea para Extração de Conhecimento a partir de Dados*. PhD thesis, Universidade do Minho, Dezembro 2000.
- [San01] Maribel Santos. *PADRÃO — Um sistema de Descoberta de Conhecimento em Bases de Dados Georeferenciadas*. PhD thesis, Universidade do Minho, 2001.

#### Notas de Rodapé

<sup>1</sup> Bolseira da FCT ao abrigo do Programa SAPIENS 99

## Una Aplicación de las Bases de Datos en Demografía Histórica: la Reconstrucción de Familias en MsAccess

Carles Millàs Castellví  
Carmen Sisón Bernal

### Introducción

Los proyectos de investigación, y especialmente los correspondientes a las ciencias sociales, requieren largas y tediosas tareas consistentes tanto en la recogida de datos, como en la clasificación y tabulación de los mismos para poder proceder a su análisis. Esta fase rutinaria, pero imprescindible, del proceso de investigación conlleva el empleo de grandes recursos en términos de tiempo y esfuerzo, que pueden verse sensiblemente reducidos mediante la utilización de herramientas informáticas adecuadas como, por ejemplo, las bases de datos.

De hecho, el uso de ordenadores en los estudios de población es, hoy en día, prácticamente imprescindible. Desde sus orígenes, la relación entre la Informática y la Demografía Histórica ha ido evolucionando y mejorando considerablemente. No obstante, la conexión entre el microanálisis nominativo y los programas informáticos no ha sido siempre fructífera. De aquí que pretendamos presentar, en forma de proyecto, una adaptación informática aplicada al método de reconstrucción de familias en MSAccess.

Creemos que al trabajar con esta base de datos podremos aligerar sustancialmente algunas de las tareas más rutinarias y costosas, tanto en el ámbito estadístico como en el ámbito temporal. Sin embargo, no pretendemos ofrecer resultados sobre una investigación en curso, sino que queremos exponer en esta comunicación las opciones y los pasos seguidos en la adaptación del método de Fleury y Henry<sup>1</sup>, a la vez que desarrollamos las rutinas y algoritmos empleados y buscamos soluciones a los problemas planteados.

### El Método

La reconstrucción de familias fue diseñada para recoger en una única ficha la información pertinente acerca del matrimonio, el nacimiento y la defunción de los miembros de una pareja concreta y de su descendencia. Una vez reunida, dicha información permite reconstruir las tasas de fecundidad por edades del matrimonio y por duración de éste; por otro lado, permite realizar algunas estimaciones de la nupcialidad —tal como la edad al contraer las nupcias—, de la mortalidad infantil y estimar los intervalos protogenésicos e intergenésicos; también permite abordar las estructuras de la fecundidad, la nupcialidad y, en menor grado, de la mortalidad... Es por ello un método muy válido para los análisis demográficos, pero, la ausencia de datos sociales y económicos suponía una evidente limitación. De aquí que en nuestro proyecto introduzcamos informaciones referentes a estas variables —mediante el oficio de los individuos masculinos, mayoritariamente, o informaciones referentes a la religiosidad popular coetánea— para mejorar así sus resultados cualitativamente.

No es extraño, pues, que el método de reconstrucción de familias sea cuestionado, principalmente, por su escasa representatividad —ya que sólo se tienen en cuenta las parejas sedentarias si se aplica exhaustivamente este método en una única parroquia— y por su laboriosidad —muchas horas de trabajo y, a menudo, pocos resultados<sup>2</sup>. No obstante, se ha afirmado que “es la única manera de saber algo acerca del comportamiento reproductivo de las sociedades antiguas”<sup>3</sup> y, aún así, se ha cuestionado la necesidad de realizar más estudios empleando el método de Fleury y Henry, sugiriendo, a pesar de ello, que “el futuro de este tipo de estudio tanto en España como en el resto de Europa tiene necesariamente que pasar por una renovación de

su enfoque principal, así como del período histórico que suele abarcar<sup>4</sup>. De esta manera, es lógico que hayan surgido variantes, modificaciones e, incluso, nuevos métodos alternativos<sup>5</sup> que permiten enriquecer la reconstrucción de familias al cotejar diferentes fuentes para conseguir, con dicha unión, nuevas propuestas de análisis más osadas y atrevidas<sup>6</sup>.

### Las Fuentes

Como en la mayoría de estudios realizados, he-

mos recurrido a los registros parroquiales y civiles —para recoger los hechos vitales y todas aquellas otras informaciones que nos suministran—, principalmente, sin olvidar las fuentes notariales, municipales, generales y otros documentos eclesiásticos que nos puedan complementar la información de los personajes de la base de datos<sup>7</sup>.

Los cuadros siguientes ilustran toda la información recopilada de los libros de bautismo, de matrimonio y de defunción de los registros parroquiales, documentación básica del período preestadístico<sup>8</sup>.

### LIBROS DE MATRIMONIO

Información general	Información del novio	Información particular
Fecha del evento (día de la semana, día, mes y año)	Nombre, apellidos y apodo del contrayente	Nombre y apellidos de la contrayente
Oficiante (nombre, apellido, cargo —rector, sacerdote...—)	Estado civil	Estado civil
Parroquia o iglesia donde se contrae matrimonio	Edad	Edad
Obispado y provincia al cual pertenece	Oficio, profesión...	
Licencias por la procedencia de otra parroquia y obispado	Naturalidad o residencia anterior (parroquia, obispado...)	Naturalidad o residencia anterior (parroquia, obispado...)
Fecha de las amonestaciones	Nombre y apellido del padre	Nombre y apellido del padre
Testigos de la unión matrimonial (nombres, apellidos, procedencia, estado civil y oficio)	Estado vital del progenitor	Estado vital del progenitor
Dispensas por consanguineidad	Oficio del padre	Oficio del padre
Misas de bendición	Nombre y apellido de la madre	Nombre y apellido de la madre
Realización de capitulaciones matrimoniales (notario, fecha y lugar)	Estado vital de la progenitora	Estado vital de la progenitora

### LIBROS DE BAUTISMO

Información general	Información particular
Fecha del evento (día de la semana, día, mes y año)	Nombre
Oficiante (nombre, apellido, cargo —rector, sacerdote...; comadrona...— )	Progenitor (nombre, apellidos, oficio y nacionalidad)
Parroquia, iglesia o lugar donde se realiza el bautismo	Progenitora (nombre, apellido y nacionalidad)
Obispado y provincia al cual pertenece	Abuelos paternos (nombres, apellidos, procedencia y estado vital)
Fecha de nacimiento (día, mes y año)	Abuelos maternos (nombres, apellidos, procedencia y estado vital)
Padrino (nombre, apellido, estado civil, relación de parentesco...)	Legitimidad (hijo legítimo y natural o ilegítimo)
Madrina (nombre, apellido, estado civil, relación de parentesco...)	Otras informaciones (bautismo de urgencia, parto múltiple...).

### LIBROS DE DEFUNCIONES

Información general	Información particular
Fecha del evento (día de la semana, día, mes y año)	Nombre y apellidos del difunto/a
Hora, fecha y lugar de la defunción	Naturaleza del difunto/a
Oficiante (nombre, apellido, cargo —rector, sacerdote...— )	Estado civil
Parroquia, iglesia o lugar donde se realiza el sepelio	Edad y/o nivel de madurez
Obispado y provincia al cual pertenece	Nombre/s, apellido/s, oficio y estado vital de los progenitores y/o cónyuges
	Causa mortis
	Recepción de sacramentos (penitencia, unción y comunión)
	Realización de testamento (notario, lugar y fecha)

Fuente: Archivo Parroquial de Abrera, Archivo Parroquial d'Olesa de Montserrat y Archivo Parroquial de Pallejà. Sacramentarios. Elaboración propia.



## El Enfoque Informático

Nuestro objetivo es, precisamente, comentar el desarrollo de una aplicación de MSAccess en el proceso de reconstrucción de familias, según el siguiente esquema de relaciones:

- Justificación del diseño establecido.
- Las relaciones entre las tablas.
- Algunas peculiaridades de las tablas establecidas.
- Las consultas.
- Los formularios.
- Los informes.

## Justificación del diseño establecido

Dada la amplitud del lapso de tiempo entre un acontecimiento y otro en la vida de cada individuo, así como las complejas interrelaciones entre los mismos, recomponer los datos biográficos hasta poder hacer el correspondiente análisis demográfico resultaría una tarea muy lenta e intrincada para llevarla a cabo manualmente, a causa de la búsqueda de cada dato a lo largo de los registros eclesiásticos de muchos años. Sin embargo, si basamos toda la tarea en relacionar entre si los individuos, que previamente se van dando de alta en la tabla principal de nuestra base de datos, según van apareciendo en los mencionados registros (ordenados alfabéticamente por la propia aplicación informática a medida que vamos trabajando), podemos seguir la tarea en el mismo orden temporal en que dichos eventos acaecieron. Véase el núcleo de la estructura en la figura 1<sup>9</sup>.

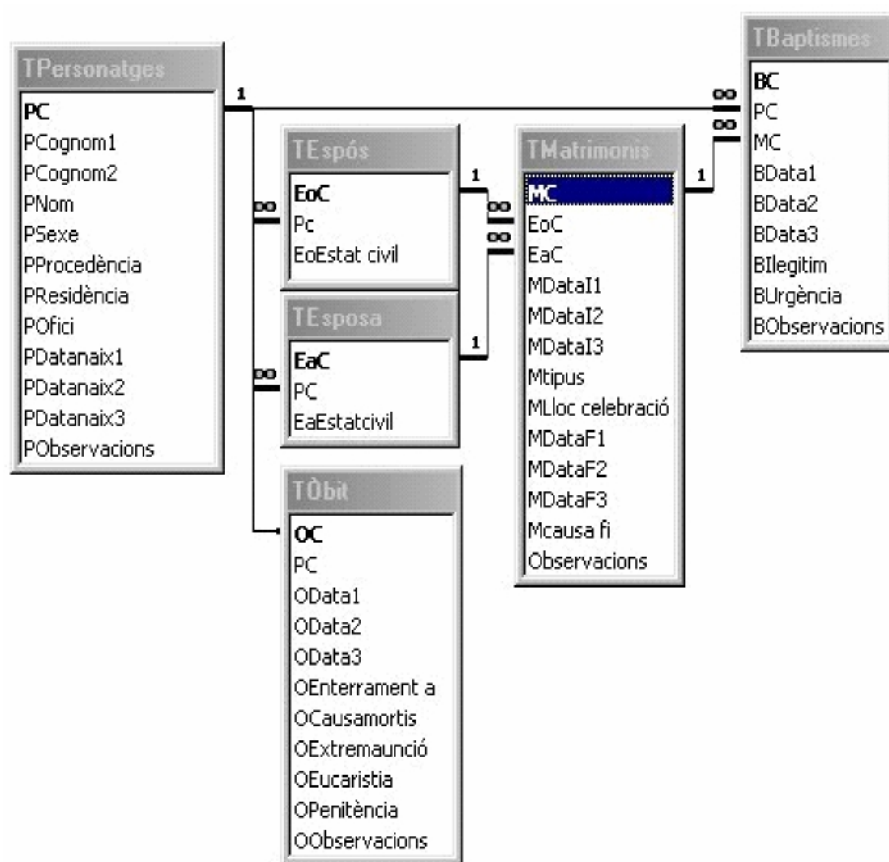


Fig. 1 Tablas principales y relaciones entre las mismas

Así, la utilización de una tabla de partida relativamente neutra, como es la tabla TPersonatges, nos permite introducir los datos de cada persona, independientemente de los acontecimientos que irán jalonando su vida, a fin de:

a.- someter los datos a los criterios de ordenación que más nos convengan para su utilización en cada momento, y que el programa se ocupará de llevar a cabo.

b.- establecer las condiciones para evitar posibles repeticiones y, también, los errores que con toda probabilidad se producirían como consecuencia de las mismas.

c.- tener una referencia única (el individuo) que aglutine toda la información.

En términos generales, cada tabla está formada por una serie de campos (tipos de datos) a cumplimentar para cada registro (elemento caracterizado por dichos datos) de manera que exista una coherencia lógica con el concepto alrededor del cual se estructura la tabla. Es decir, que para cada campo y registro exista una y sólo una información posible.

De esta manera, nuestra tabla principal, TPersonatges, recoge todas las características relevantes del individuo que resulten invariables, pero no podemos incluir datos que pueden variar a lo largo de su vida, como son sus uniones o alumbramientos, por ejemplo, porque requeriría hacer repeticiones de registros para un mismo individuo y la aplicación no podría relacionarlos con un solo individuo; tampoco podemos incluir datos que, a pesar de existir una sola vez para cada individuo (nacimiento u óbito), se producen a tanta distancia temporal uno del otro que requeriría cumplimentar los datos de una misma tabla en veces sucesivas (hacer modificaciones de los registros correspondientes), haciendo el trabajo más lento y favoreciendo la posibilidad de cometer errores. Asimismo, el carácter circular de las relaciones en-

tre las tablas (cada individuo es hijo de alguien y será, probablemente, padre o madre de alguien, que se ha recogido previamente en la tabla TPersonatges) establece una complejidad de las mismas que solamente se pueden recoger adecuadamente estos datos separando las tablas. Estos son los motivos de la creación de dos tablas aparte para recoger, respectivamente, los datos de bautismo (TBaptismes) y los datos de la defunción (TÒbit)

Las tablas, pues, nos permiten reflejar una gama de características individuales (datos de cada sujeto, como son: nombre, apellidos, día, mes y año de nacimiento), o de circunstancias (uniones o matrimonios, alumbramientos, óbitos), añadiendo en cada caso toda la información hallada al respecto en distintas fases, mediante tablas complementarias, siempre ligadas a los registros correspondientes de la tabla principal, y a los particulares de cada una de dichas tablas complementarias (como por ejemplo, en el caso de TÒbit, las observaciones de causa mortis o la administración sacramental).

El mismo criterio de unicidad rige para las relaciones que se establecen entre registros: un registro que pueda tener múltiples relaciones con otro, que también presenta la misma característica, requiere el establecimiento de campos clave tales que permitan volver a establecer las relaciones sin repetición de la clave de relación: la solución es establecer tablas intermedias que nos permiten generar relaciones diferentes bajo una clave distinta cada vez. La tabla TMatrimonis genera una clave para cada unión, permitiendo que exista más de una unión para cada individuo y recogiendo toda la información de cada uno de dichos eventos.

También es necesario que las relaciones se realicen entre registros de tablas diferentes. Por esta razón se han creado tablas en las que se declara a cada individuo esposa (TEsposa) o esposo (TEspòs), según sea el caso, a fin de poder establecer posteriormente las correspondientes uniones entre ellos. La alternativa de mantener los

sexos separados desde el inicio planteaba el inconveniente de crear dos tablas principales, en lugar de una sola, haciendo más lenta la operación de introducción de los datos, y complicando también clasificaciones posteriores de los mismos por conceptos comunes a ambos sexos.

### Las relaciones entre las tablas

Creadas y definidas adecuadamente las tablas, con sus correspondientes campos clave, y campos de datos de información requerida para el estudio que se está realizando, se establecen las relaciones en el sentido de uno a varios, es decir, de la tabla donde un registro genera relaciones con varios registros, a la tabla donde varios registros reciben una relación de un solo registro de la otra tabla, que en la pantalla correspondiente (véase Fig. 1) se representan con los símbolos 1 a ¥ sobre la línea que va de clave a clave de cada tabla relacionada, excepto casos como el de la tabla *TÒbit* y la tabla *TBaptismes*, ya que sólo se puede nacer o fallecer una vez y, por lo tanto, carece de sentido la posibilidad de repetir el evento, y se ha establecido una relación de 1 a 1.

Se ha exigido integridad referencial en todas las relaciones a fin de que, en caso de eliminar un individuo de la tabla principal (supóngase un error de repetición de un individuo), se eliminen en cadena todos los registros de cualquier tabla posterior que se hubiesen relacionado con el registro eliminado en la tabla de origen, garantizando así la fiabilidad de la corrección. De esta manera se pueden recoger los datos de los individuos de más de una fuente, reduciendo el volumen de pérdida de información y manteniendo la facilidad de depuración de los datos mediante una simple consulta que compare los campos de cada registro para detectar coincidencias entre sus respectivas características. La misma consulta ofrece datos sobre los registros en cascada que pueden quedar eliminados, para proceder a su comprobación y, en su caso, a su correcta reconstrucción con nuevos datos<sup>10</sup>.

### Algunas peculiaridades de las tablas establecidas

Cabe destacar, por lo que respecta a las tablas, que las fechas se han separado en día, mes y año, a fin de poder extraer más tarde conclusiones sobre la posibilidad de rechazar o no rechazar la hipótesis de la existencia de una relación entre determinados tipos de evento y un determinado mes (o temporada) con aspectos sociales, económicos, etc.

En las tablas *TEspós* y *TEsposa* se incluye el estado civil para poder recoger si se trata de primeras nupcias o de uniones posteriores y la causa de la disolución del vínculo anterior (que podría no ser exactamente un matrimonio o una defunción), así como anotaciones, en observaciones, del oficio del esposo cuando este dato es conocido. Se declara esposo/esposa cada vez que se establece una unión a fin de poder consignar el estado civil previo de cada contrayente.

En la tabla *TMatrimonis* se incluye un campo (*Mtipus*) para establecer si se trata de una unión espúrea o de un matrimonio formalizado de acuerdo con las normas legales y religiosas del momento. Las denominaciones de esposo/esposa, por lo tanto, deben ser entendidas en sentido muy amplio. Haciéndolo así, se evita que cuando dichos vínculos den lugar a un nacimiento, se pierdan las relaciones de parentesco, al menos, cuando éstas son conocidas. También los datos de los alumbramientos de la madre quedarán debidamente registrados.

La tabla *TBaptismes* es la que permite establecer el nacimiento de cada individuo (recordemos que los datos se extraen, mayoritariamente, de registros parroquiales) y la mayoría de los personajes que se pueden obtener; en dicha tabla se podrán observar los casos de ilegitimidad y los bautismos de necesidad. A través de la información que se recoge en la tabla *TMatrimoni* conoceremos si proceden de un matrimonio o de otro tipo de unión (dado que en la tabla *TPersonatges* se ha incluido

un registro de padre desconocido y otro de madre desconocida, donde se puede recoger cualquier procedencia, desde los expósitos hasta los hijos de madre soltera y padre desconocido).

En la tabla TÒbits se recogen dos aspectos remarcables: por un lado, la posible causa de la muerte (violenta, accidental, enfermedad ...) u otras observaciones al respecto, como pueden ser las referencias a conflictos bélicos, incidencias meteorológicas...; y, por el otro, la administración sacramental (penitencia, eucaristía y extremaunción).

### Las consultas

La ordenación alfabética de los datos de la tabla principal para facilitar su posterior utilización en la introducción de los datos siguientes, se obtiene mediante una consulta sencilla como la que se detalla

en la Fig. 2.

Análogamente, pueden crearse consultas para ordenar cualesquiera campos de una tabla (o de varias tablas relacionadas) a fin de mostrar diversas características de cada individuo.

En nuestro caso, la tabla que se recoge en la Fig. 2 (recuadro superior izquierda) nos permitirá declarar esposo o esposa a cada individuo que establezca una unión mediante un sencillo formulario en el que bastará con seleccionar, mediante el ratón, el citado individuo entre una lista que la consulta nos ofrecerá ordenada, como se verá al comentar los formularios. También podremos establecer sus progenitores, su fecha de bautismo y su óbito a medida que vayamos incorporando la información de las páginas de los registros eclesiásticos que nos sirven de fuente de información.

*Estableciendo orden ascendente en cada uno de los campos, la aplicación ordenará sucesivamente por cada uno de ellos en la misma secuencia en que aparecen en la tabla.*

Campo:	PCognom1	PCognom2	PNom	POfici
Tabla:	TPersonatges	TPersonatges	TPersonatges	TPersonatges
Total:	Agrupar por	Agrupar por	Agrupar por	Agrupar por
Orden:	Ascendente	Ascendente	Ascendente	
Mostrar:	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Criterios:				
o:				

Fig. 2 Consulta para ordenar los registros por apellidos y nombre sucesivamente

Más tarde, las consultas también nos permitirán unir toda la información que deseemos listar, con los datos ordenados, agrupados y distribuidos según convenga al propósito de nuestro trabajo, proceso que se ilustrará en su momento.

Se nos permitirá recordar en este punto que existen diversos tipos de consultas para diversos fines, como por ejemplo las de selección, de creación de tablas, de eliminación de registros, de referencias cruzadas, o de actualización de registros, cada una de las cuales, mediante el establecimiento de los criterios oportunos, nos ofrecen datos discriminados, ordenados y, si conviene, agrupados y sumados, con lo que ello conlleva de ahorro de tiempo y energía en dichas tareas, como ya se había comentado al iniciar este escrito.

Asimismo, las consultas nos permitirán establecer un rápido sistema de depuración de identificaciones buscando coincidencias entre los diversos campos que definen la identidad exacta de un sujeto. Al realizar esta depuración atendiendo a más de un campo, el investigador podrá discriminar adecuadamente

los casos de variación ortográfica de los nombres o apellidos de un mismo sujeto en distintos registros y la coincidencia de nombres de dos o más sujetos distintos, mejorando la fiabilidad del estudio.

Las consultas también permiten concatenar campos de una misma tabla de manera que aparezcan como un solo campo. En la misma consulta efectuada sobre la tabla TPersonatges que nos ha de servir para seleccionar los esposos, esposas, bautismos y óbitos, podemos crear dicha concatenación para que, en el formulario correspondiente, nos muestre los dos apellidos (cuando están disponibles) y el nombre.

La concatenación, aparecería, en el diseño de la tabla, como se muestra en la Fig. 3: Nombre que damos al campo, separado por dos puntos de los campos concatenados, utilizando el símbolo & para cada concatenación, ya sea con los espacios y caracteres intermedios que queramos intercalar, o con otro campo. Véase que los campos, para MSAccess se escriben entre corchetes y los espacios o caracteres que insertamos, entre dobles comillas.

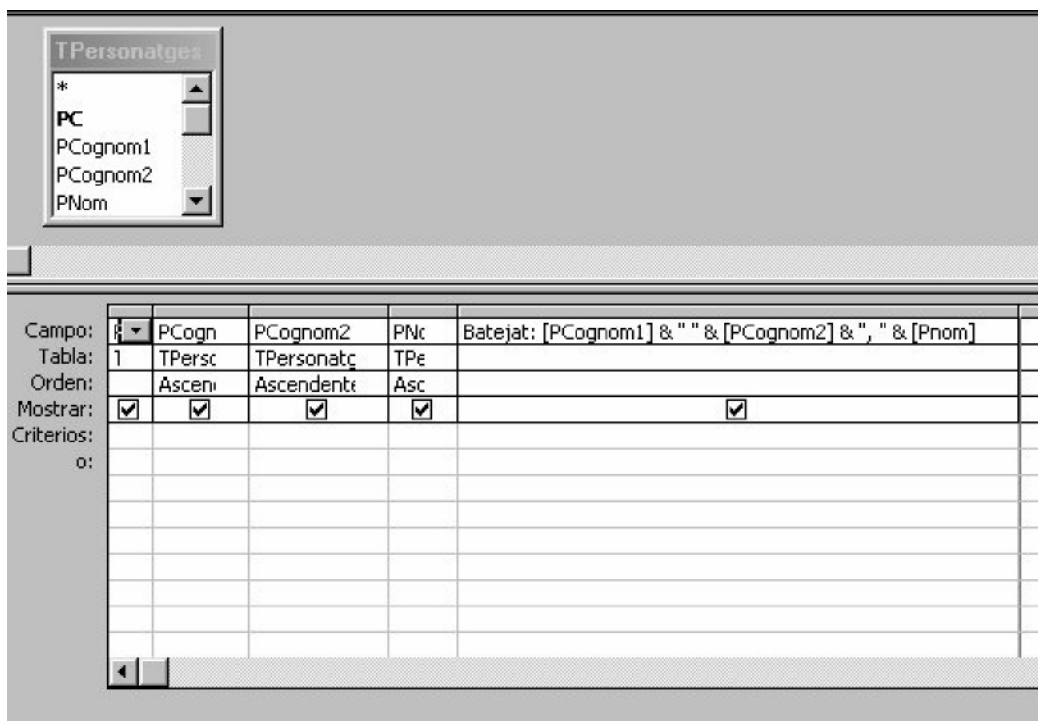


Fig. 3 Ejemplo de concatenación de campos

El resultado buscado es obtener un cuadro combinado en el formulario correspondiente tal como se puede observar en la Fig. 4, donde se ilustra su apariencia sobre el formulario correspondiente a la ficha de bautismos (FBaptismes).

### Los formularios

Los formularios de MSAccess nos sirven para preparar pantallas que facilitan la entrada de los datos, permitiéndonos trabajar con datos de una o más tablas, como es el caso que estamos utilizando como ejemplo en la Fig. 4 y la Fig. 5. Para trabajar

con datos de más de una tabla, utilizaremos los cuadros combinados, o bien los subformularios.

Por lo que se refiere a los cuadros combinados, como en nuestro ejemplo, bastará pulsar el botón del ratón sobre el nombre elegido para que quede registrada su clave de personaje en el campo correspondiente de la tabla de bautismos. El nombre de padre y madre será seleccionado de la misma manera en su correspondiente cuadro combinado, basado en una consulta a partir de las tablas TPersonatges y TMatrimonis (véase Fig. 5). El resto de los datos del formulario se rellenará adecuadamente.

Fig.4 Ejemplo de cuadro combinado en el Formulario de Bautismos

En este punto, y con la perspectiva ganada hasta ahora, conviene puntualizar que la tabla TÒbit, que recogerá los fallecimientos, no se ha enlazado a la de TBaptismes porque ello nos impediría registrar un fallecimiento de un posible personaje que no hubiese recibido el sacramento del bautismo o no tuviésemos datos al respecto, como en el caso de personajes inmigrados de otras zonas, suceso bastante frecuente.

Los números que siguen a los nombres del padre y de la madre son las respectivas fechas de nacimiento, que facilitarán la identificación adecuada de los progenitores a partir de las mismas, cuando exista repetición de nombres, lo cual es bastante frecuente puesto que en muchos casos sólo se dispone de un apellido y el nombre de pila pasa habitualmente de padres a hijos.

**FITXA DE BAPTISME**

Fitxa Personatges

Codi: (mérico)

Batejat: [dropdown]

Data Baptisme: 0/0/0

Il·legítim:

Pares: [dropdown]

Urgència: [dropdown]

Observacions: [text area]

Dropdown list for Pares:  
 Ereter , Ramon; 0-0-0 + Alavedra , Elisabet; 0-0-0  
 Moragues , Salvador; 0-0-0 + Llobateres , Beneta; 0-0-0  
 Sanahuja , Gabriel; 0-0-0 + Puigventós , Caterina; 0-0-0  
 Torralcina , Antoni; 0-0-0 + Sanahuja , Elisabet Beneta; 8-4-1579

Fig. 5 Ejemplo de cuadro combinado para la selección de los padres

Si no se requieren cuadros combinados ni subformularios (si trabajamos con una sola tabla), el formulario se limitará a un encabezado con su nombre, un pie con los botones para navegar entre los registros, y en el cuerpo, los campos de datos a rellenar. También podemos añadir botones para pasar directa y cómodamente de un formulario a otro en caso de tener que cambiar de tarea. Véase la Fig. 6.

Como se puede observar, en el formulario de personajes podemos añadir todos los datos inmutables de cada individuo. Se ha añadido la procedencia para poder distinguir después entre los autóctonos y los llegados de otras zonas (sobre los cuales faltarán seguramente algunos datos por no figurar sino en los registros de otra parroquia).

**FITXA DE PERSONATGE**

Menú d'Inici

Fitxa d'Òbits

Fitxa Matrimoni

Esposa

Espós

Codi: (mérico)

Sexe: [dropdown]

Primer cognom: [text field]

Segon cognom: [text field]

Nom: [text field]

Procedència: [text field]

Lloc de residència: [text field]

Ofici: [text field]

Data naixement: 0/0/0

Observacions: [text area]

Navigation bar: [back] [left] [right] [forward] [stop]

Fig. 6 Ejemplo de formulario simple con botones de desplazamientos

Una vez añadidos los campos en la vista de diseño del formulario y alineados convenientemente (véase Fig. 7), se procederá a establecer el orden de tabulación, a fin de ir avanzando de campo en campo en el mismo orden en el que se han distribuido en el formulario (véase Fig. 8)



Fig. 7 Ejemplo de la vista de diseño del formulario

Bastará seleccionar, desde la zona gris, cada campo y hacerlo deslizarse hasta la posición deseada para establecer el orden en que queremos que avance el cursor para cumplimentarlo.



Fig. 8 Ejemplo de orden de tabulación



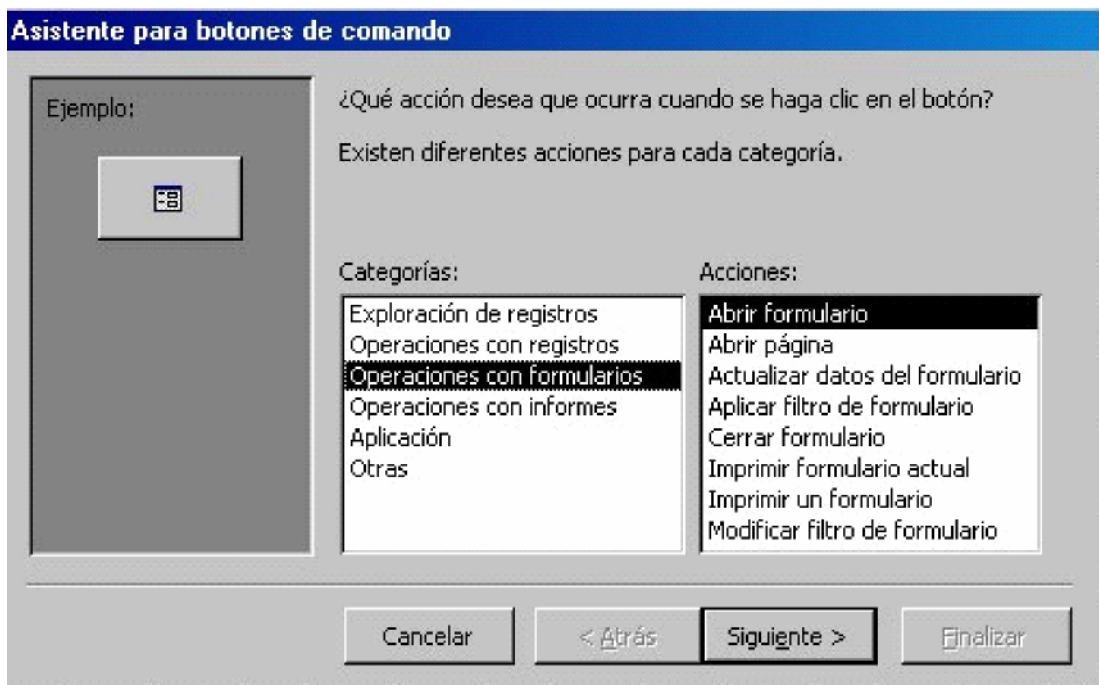


Fig. 9 Inicio del Asistente para creación de botones de comando

Otro factor importante son los botones que automatizan acciones, como por ejemplo, pasar de un formulario a otro. El asistente de la barra de herramientas nos guiará en su creación, como se ve en la Fig. 9. Pero conviene completar su programación desde la edición de código de VisualBasic para que:

a) el formulario que abandonamos quede cerrado (Fig. 10), evitando que se nos vayan acumulando una serie de ventanas abiertas que ocupan memoria y reducen la velocidad de trabajo

b) el formulario que abrimos se maximice para que ocupe toda la pantalla automáticamente (Fig. 11).

#### Los informes

Con independencia de la exportación de los datos, ya depurados, a hojas de cálculo, y concretamente, en nuestro caso, a MS Excel, con el auxilio de las consultas para seleccionar y ordenar los

datos que interesen en cada caso, se pueden obtener listados para recoger los diferentes aspectos de la investigación, inclusive la ficha de familia, según anexo.

#### Criterios de Depuración de Observaciones

Si bien es imprescindible depurar los datos para evitar sesgos en los resultados del proceso estadístico, no lo es menos que la pérdida de información reduce la validez científica de la investigación a la vez que constituye, por si misma, una importante fuente de sesgo, razón por la cual, a diferencia de algunos de los proyectos de investigadores que nos han precedido, y que, ciertamente, nos han proporcionado una valioso punto de referencia, hemos optado por:

- No eliminar los hijos póstumos, por parte de padre, de los registros, siempre y cuando no superasen los nueve meses tras la fecha de finalización de la unión de la pareja, estudiando por

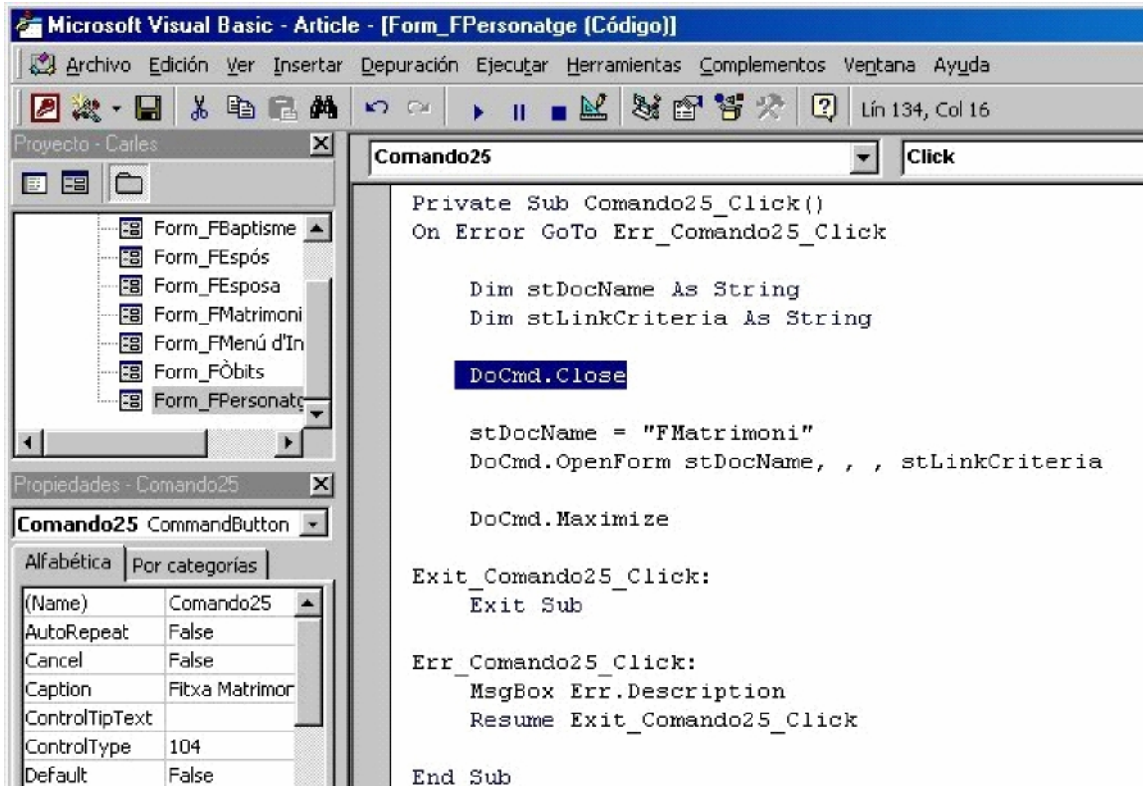


Fig. 10 Inclusión de un comando de cierre del formulario de partida

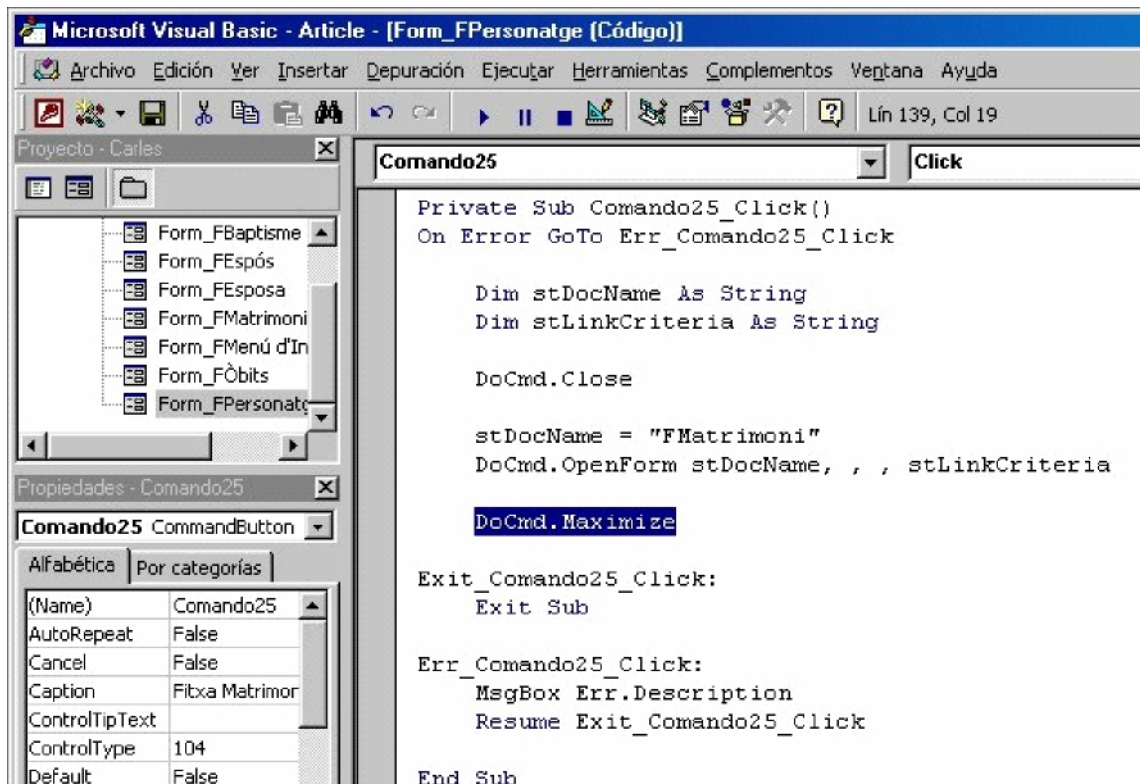


Fig. 11 Inclusión de un comando de maximización del formulario que se abre

separado los casos muy próximos a esa frontera para evitar errores de bulto.

- También hemos dejado abierta la edad de la madre ya que, aunque muy raramente, se pueden dar nacimientos de madres mayores de 50 años, y, puesto que la eliminación del dato del alumbramiento, si no hay razón para considerarlo un error de identificación, debería comportar la eliminación de toda la observación para no causar sesgos al estudio, es decir, la eliminación de todos los hijos de esa familia, y, por tanto, de la ficha de familia. Sí que se ha tenido en cuenta que hay que validar cuidadosamente todos los datos correspondientes a las observaciones extremas para asegurar su fiabilidad. Pero, precisamente por la escasez de dichos casos, se justifica la no automatización de su tratamiento.

- La elaboración de las correspondientes campanas de Gauss, en todo caso, nos pueden dar una primera indicación de la normalidad de los datos recogidos.

- También se ha tenido en cuenta la necesidad de realizar validaciones de datos para evitar las imposibilidades lógicas, para lo cual se han utilizado consultas para la comparación de fechas.

- Como ya se ha mencionado anteriormente,

ante la necesidad de eliminar un dato de nacimiento por ser muy dudoso, o un período de datos de la misma madre, hemos optado por eliminar por completo los datos de esa madre, ya que de otro modo se da lugar a un sesgo importante en los estudios de fecundidad, objetivo principal del método de Henry y Fleury.

### Conclusiones

Dado el avance creciente de los recursos y prestaciones que ofrecen las bases de datos, aun cuando hay que poner especial cuidado en la depuración de los datos y en no sistematizar el tratamiento de datos extremos, la utilización de bases de datos (MS Access, Oracle, etc.), permite la ampliación de las zonas geográficas de estudio y la colaboración en un mismo proyecto de diversos investigadores, mediante la construcción de páginas web que integren los formularios de entrada de dichos datos (usando, por ejemplo, programación en JAVA y ASP) para permitir que cada miembro del equipo introduzca la información de su trabajo de campo en la misma base de datos. El tratamiento de la información (informes, análisis estadísticos, etc.) podría llevarse a cabo tanto para la totalidad del área estudiada, como para determinados sectores de dicha zona, permitiendo la obtención de conclusiones a nivel global, parcial y comparativo.

## Notas de Rodapé

1 FLEURY, M.; HENRY, L. (1956). Des registres paroissiaux à l'histoire de la population. Manuel de dépouillement et d'exploitation de l'état civil ancien. (París).

El mismo año 1956, Henry publicó *Anciennes familles genevoises. Étude démographique* (París), donde puso en práctica su método con un estudio genealógico sobre la burguesía ginebrina y, junto a Gautier, dos años después consiguió el éxito con *La population de Crulai, paroisse normande. Étude historique* (París).

2 PÉREZ GARCÍA, J. M. (2000). “¿Es necesario privilegiar todavía la reconstrucción de familias en España? Una respuesta a David S. Reher”, *Boletín A.D.E.H.*, XVIII, II, p. 106 y siguientes. Concretamente, afirma que: “(...) el ocaso de la reconstrucción de familias queda sentenciado en España ya sea por la laboriosidad del método, su imposibilidad de fijar los cambios y su reclusión a épocas pasadas”. En las páginas siguientes, Pérez García profundiza en las críticas al método realizadas tanto en Francia como en el resto del continente.

3 REHER, D.S. (2000) “La investigación en demografía histórica: pasado, presente y futuro”, *Boletín A.D.E.H.*, XVIII, II, p. 33.

4 REHER, D.S. (2000). “La investigación...”, p. 36.

5 Un buen estado de la cuestión al respecto puede consultarse en PÉREZ Y GARCÍA, J.M. (2000), “¿Es necesario...”, p. 105-122 y REHER, D.S. (2000). “La investigación...”, p. 15-78.

Sin embargo, creemos significativo resaltar la metodología lusitana, que tan buenos resultados ha tenido, de la reconstrucción de parroquias empleando una reconstrucción de familias directamente en fichas de familia; creando ficheros de individuos y, posteriormente, agregando la información proveniente de otras fuentes [AMORIM, M. N (1995). “Informatização normalizada de arquivos. Reconstituição de paròquias e história das populaões”, *Boletín A.D.E.H.*, XIII, II]; la utilización del método henryano en el éxito de la historia total de la escuela de los Annales donde los aspectos demográficos se relacionan con las dimensiones económicas, sociales y culturales [REY, O. (1981). *Aproximación a la historia rural de la comarca de la Ulla (siglos XVII y XVIII)*. Santiago de Compostela, TORRENTS, A. (1993). *Transformacions demogràfiques en un municipi català: Sant Pere de Riudebitlles, 1608-1935*. Barcelona...], la utilización de diferentes métodos en la tesis de Valentí Gual [GUAL, V. (1993). *La familia moderna a la Conca de Barberà. Tarragona*] o la aportación metodológica e informatizada de Francisco Ramírez [RAMÍREZ, F. (1998). *Transformaciones y comportamientos demográficos diferenciales en dos municipios del interior andaluz. Aplicación del método de reconstrucción de familias en las poblaciones de Iznájar y Loja (siglos XVII-XX. Granada)*] que ha demostrado que la rentabilidad y la representatividad de una investigación no están reñidas con la visión microscópica del objeto de estudio utilizando el programa informático dBase III Plus.

6 PÉREZ GARCÍA, J.M. (2000). “¿Es necesario...”, p. 117. Este autor destaca, en este sentido, los trabajos de URRICOECHA, J. (1992). *En una mesa y compañía. Caserío y familia campesina en la crisis de la “sociedad tradicional”. Irún, 1766-1845*. donostia; y GARRIDO, E. (1993). “Familia, parentesco y alianza en la Huerta de Valencia, siglo XVIII. La estrategia familiar de la consanguinidad”, *Estudis*, 18.

7 Específicamente se están vaciando capítulos matrimoniales, inventarios post-mortem, padrones y censos de población, manuales y actas municipales, matrículas industriales, listados de alumnados, mensuales notariales, etc.

8 Omitimos las informaciones complementarias que nos suministran las otras fuentes vaciadas.

9 En la figura se han omitido parte de las tablas en función de una visión más clara de la estructura tipo de la base de datos

10 Este método garantiza la depuración de datos repetidos o erróneos, permitiendo al investigador la validación de las modificaciones y reduciendo a un mínimo el volumen de información perdida.

## Construção e Exploração de uma Base de Dados Prosopográfica Normalizada do Clero Catedralício Português na Idade Média

Ana Maria Jorge

Ana Maria S. A. Rodrigues

Hermínia V. Vilar

Pedro Henriques e Sandra Lopes

### Resumo

Nesta comunicação pretende-se apresentar os fundamentos teóricos e metodológicos e os aspectos técnicos do projecto *Fasti Ecclesiae Portugaliae*, que tem como objectivo a realização e disponibilização à comunidade científica de uma Base de Dados Prosopográfica Normalizada que permita o estudo do clero catedralício português na Idade Média (no imediato, apenas entre 1071 e 1325).

Procurar-se-á justificar, quer a adequação do método prosopográfico à abordagem do referido grupo social, quer a relevância do questionamento efectuado às fontes e a própria selecção destas, assim como se evocarão os principais problemas suscitados pela necessidade de arquivo e manipulação de um tão grande número e variedade de dados referentes a indivíduos conhecidos pelos seus nomes (em particular, a homonímia).

Daqui se retira a necessidade de proceder à recolha da informação, extraída das fontes, em fichas normalizadas que permitam: a) o seu armazenamento em bases de dados; b) o seu tratamento sistematizado.

Identificam-se, então, as várias operações de análise que se pretendem à partida e que se tornam possíveis devido ao suporte em computador.

Por fim, discute-se a solução informática preconizada para apoiar todo o processo, da recolha e normalização ao armazenamento, análise e divulgação.

### 1. Introdução: o Projecto

A construção de uma base de dados prosopográfica normalizada, acessível à comunidade científica e que permita o estudo do clero catedralício português na Idade Média constitui o objectivo básico do projecto *Fasti Ecclesiae Portugaliae - Prosopografia do clero catedralício português (1071-1325)*. Financiado pela Fundação para a Ciência e Tecnologia, iniciou-se em Junho de 2002, prevendo-se o seu prolongamento até 2005.

Este projecto está sediado na Universidade Católica Portuguesa e tem como investigadora responsável a Doutora Ana Maria Jorge sendo a restante equipa composta por treze investigadores oriundos de diferentes Universidades do País<sup>1</sup>, além de quatro bolsiros recrutados por concurso<sup>2</sup>.

Sendo o principal objectivo deste projecto, como já acima referimos, a elaboração de uma base de dados prosopográfica do clero catedralício e tendo em conta as dificuldades que se colocam ao nível da inventariação documental de muitos dos arquivos portugueses, optámos, nesta fase, por restringir a cronologia de forma a podermos abranger todo o espaço eclesiástico nacional. Decidimos, assim, estudar, exclusivamente, o período compreendido entre 1071, data da restauração da primeira diocese portuguesa após a ocupação muçulmana, a de Braga, e 1325, final do reinado de D. Dinis, durante o qual se deu a “pacificação” das relações entre a monarquia e o clero através do estabelecimento dos primeiros acordos gerais, normalmente conhecidos como *Concordatas*. Incluímos, porém,

neste levantamento as nove dioceses existentes neste período – Braga, Porto, Coimbra, Lamego, Guarda, Viseu, Lisboa, Évora e Silves.

## 2. O grupo social em estudo: o Clero Catedralício

O universo social em análise inclui bispos – ou arcebispos, no caso de Braga –, bem como dignidades, cónegos e porcionários constitutivos dos cabidos das catedrais, mas exclui o restante clero auxiliar, como clérigos do coro, capelães e vigários, entre outros. Pretende-se, desta forma, ser exaustivo nesta identificação e não proceder por amostragem, dado por um lado a relativa pequenez do grupo, que não deve exceder os dois milhares no total para o período considerado<sup>3</sup>, e por outro a inexistência de séries documentais completas referentes a esses indivíduos que permitam tal tipo de abordagem.

De facto, existem dificuldades tanto no estabelecimento da amplitude deste grupo social como na identificação da sua composição, face à inexistência de listas de detentores de dignidades eclesiásticas compiladas por eruditos do século XIX ou mesmo de períodos anteriores, que em vários países europeus existem para diversos cabidos<sup>4</sup> e que, em Portugal, apenas se podem encontrar para os bispos e mesmo assim com bastantes erros e imprecisões<sup>5</sup>.

Nas dioceses estudadas até agora<sup>6</sup>, também não foi detectada, para o período considerado, a existência de quaisquer livros de registo de entradas nos cabidos, livros de actas das reuniões capitulares ou livros de distribuições que permitam estabelecer essas listas com alguma precisão como forma de iniciar a pesquisa, enunciando os indivíduos a analisar<sup>7</sup>. Assim, é a partir da própria consulta das fontes que se vai estabelecendo o catálogo dos clérigos ao serviço das diferentes sés nacionais<sup>8</sup>. Tal facto obriga-nos a ser, também, o mais exaustivo possível na escolha das fontes a utilizar, procurando identificar e fichar toda a documentação disponível para o conjunto de anos escolhido.

## 3. As fontes a usar: variedade e lacunas

Contudo, o estado actual de muitos dos nossos arquivos e a existência de deficientes elementos de busca ou mesmo a sua total ausência comprometem, sobremaneira, essa exaustividade que, nem mesmo, a opção feita de percorrer documento a documento alguns dos fundos mais importantes poderá colmatar em absoluto. De qualquer forma, caberá realçar, desde já, a enorme variedade de fundos arquivísticos compulsados ou a compulsar no âmbito deste projecto, desde os mais directamente ligados às respectivas catedrais, como sejam os documentos conservados nos cartórios capitulares e episcopais, aos documentos oriundos de alguns dos mais importantes cenóbios e colegiadas existentes na área das dioceses consideradas.

O recurso a esses fundos justifica-se, sobretudo, porque sabendo nós que o clero catedralício circulava de instituição para instituição ao longo da sua carreira, é nossa intenção acumular e cruzar as informações colhidas não só nos cartórios de todas as catedrais mas também das colegiadas por onde esses clérigos, com frequência, passavam. Igualmente serão postos a contribuição os restantes fundos eclesiásticos – bispos e dignidades eram muitas vezes chamados a arbitrar questões entre igrejas, mosteiros, conventos e ordens militares –, os referentes aos monarcas – em cuja burocracia o alto clero tinha um importante papel –, os do Vaticano e ainda os das dioceses espanholas que até 1393 governaram circunscrições eclesiásticas portuguesas ou, inversamente, estiveram na dependência de metropolitas portugueses<sup>9</sup>.

Mas o recurso a outros fundos que não os oriundos das catedrais, justifica-se adicionalmente pela variedade de documentação disponível para as diferentes dioceses consideradas. Se, para muitas delas, temos fontes relativamente abundantes, se bem que diferenciadas na tipologia e na distribuição cronológica da sua produção, para outras, como é o caso de Lisboa, Guarda e Silves, a documentação produzida directamente pela instituição capitular ou

episcopal é esparsa e mesmo, em alguns casos, como acontece com a diocese egitaniense, inexistente. Neste contexto, a única forma de reestruturar, se bem que de forma parcelar, a composição capitular implica o recurso a outras fontes indirectas que possam informação sobre os seus capitulares e bispos.

Lidamos, pois, e apesar de todas as contingências, com um universo documental vastíssimo, que só a limitação temporal imposta e o grande número de investigadores e bolseiros envolvidos tornam controlável.

#### 4. O Interrogatório colocado às Fontes

Que interrogações são colocadas a essas fontes? O questionário elaborado baseia-se em anteriores experiências semelhantes, quer estrangeiras, referentes ao clero capitular<sup>10</sup>, quer portuguesas, referentes aos burocratas régios<sup>11</sup>. Pretende-se, assim, para cada indivíduo, determinar a sua origem geográfica (local de nascimento, onde tomou as ordens menores, onde se encontra sepultado) e social, reconstruir a rede familiar (identidade e estatuto dos ascendentes, colaterais e descendentes) e clientélica (identidade e estatuto dos patronos, dependentes e servidores) a que pertencia, acompanhar as diferentes etapas da sua carreira eclesiástica (incluindo as ordens sacras alcançadas, os estudos feitos, eventuais serviços prestados ao rei ou ao Papa) e avaliar o seu nível económico (posse de propriedades de avoenga, compradas ou doadas; de bens móveis; de negócios).

Se bem que estas sejam as principais interrogações, é óbvio que nem todas conhecerão uma resposta para um grande número de eclesiásticos. A disparidade dos dados recolhidos impede, em muitos casos, a formulação de uma resposta cabal para muitas destas questões. Aliás, a própria identificação dos diferentes eclesiásticos constitui um problema central.

As armadilhas decorrentes da onomástica medieval, com nomes repetidos e que dificilmente iden-

tificam famílias constitui um dos primeiros problemas, já que a repetição, mesmo se frequente, do mesmo nome não poderá, por si só, sustentar uma identificação que terá de ser deduzida a partir de outros dados adicionais. Contudo, mais uma vez a disparidade de dados, tanto no aspecto quantitativo como qualitativo, impede, frequentemente, a formulação de respostas cabais a muitas das perguntas colocadas na identificação dos diferentes eclesiásticos. Desta forma, esta operação surgirá, sobretudo, como o resultado de um processo de cruzamento e de interpretação dos, por vezes, esparsos ou controversos dados disponíveis.

#### 5. O Método adoptado: justificação

Uma palavra ainda para a justificação do método adoptado. Tal como é referido no próprio título desta comunicação propomo-nos construir e tornar acessível à comunidade científica uma base de dados prosopográfica do clero catedralício português na Idade Média, cientes, porém, das interrogações e dos debates que a aplicação deste método tem suscitado<sup>12</sup>. Com efeito, e sem querer entrar agora numa discussão que dificilmente caberia no âmbito de uma exposição deste género, caberá realçar que esta nossa opção se radica no facto de lidarmos com um grupo alargado, com características comuns e percursos semelhantes, que permitem o desenhar de tendências comuns, sem que, no entanto, o contributo individual ou a originalidade de um percurso ou de uma vida sejam perdidos na definição da tendência geral. É, pois, no cruzamento do individual e do geral que, a nosso ver, o método prosopográfico se insere e aplica.

#### 6. SIEP: a solução informática preconizada

A metodologia seguida para realizar este estudo prosopográfico é formado pelas seguintes tarefas:

1. delimitar o grupo a estudar (objecto de estudo);
2. identificar fontes;
3. elaborar um questionário ao qual as fontes são supostas poder responder de forma directa e

adequada;

4. criar um conjunto de critérios de classificação;

5. recolher a informação, consultando as fontes nas dioceses escolhidas e preenchendo os questionários;

6. distribuir os questionários recolhidos pelos vários Coordenadores de diocese, conforme a catedral do indivíduo a que os dados dizem respeito e proceder ao seu armazenamento normalizado, de acordo com os critérios de classificação acima referidos;

7. proceder à identificação e fusão de indivíduos à medida que se vão juntando questionários obtidos a partir das diferentes fontes;

8. explorar estatisticamente os dados tendo em vista a análise de fenómenos sociais;

9. divulgar os dados recolhidos e os resultados extraídos.

Destas, as 4 primeiras fases realizaram-se uma vez, aquando da concepção do projecto, como explicado com detalhe nas secções precedentes; as restantes 5 fases constituem o dia a dia da realização do projecto. O SIEP nasceu, precisamente, com vista a informatizar estas 5 fases.

### 6.1 Arquitectura do SIEP

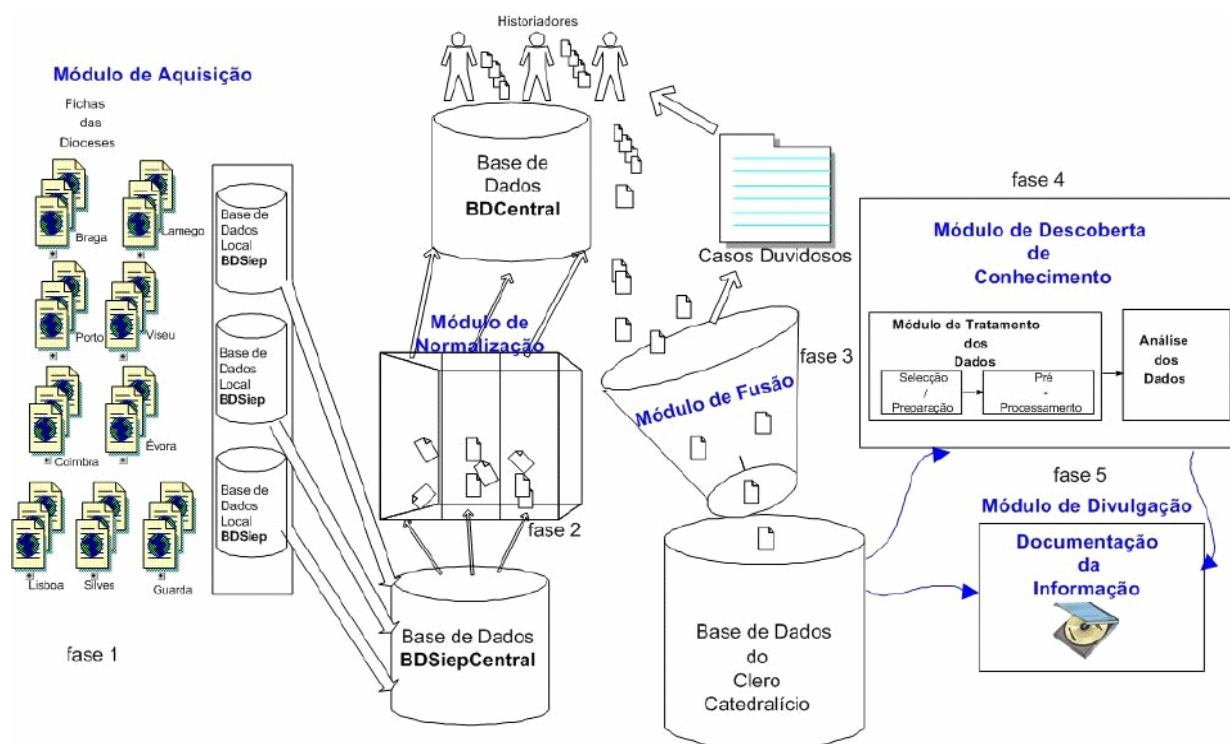


Figura 1 – Arquitectura do SIEP – sistema de suporte ao FASTI.

O SIEP – Sistema de Informação para Estudo Prosopográfico – terá cinco módulos conforme ilustra a Figura 1:

1. Módulo de Aquisição – toda a informação das fichas de documentos e das fichas prosopográficas é transferida do papel (ou directamente da fonte) para o formato electrónico, sendo armazenada em cada local (diocese) de recolha para posterior exportação e envio para o arquivo central.



2. Módulo de Normalização – todo o trabalho dos vários investigadores é importado (a partir das tais bases de dados locais) e concentrado numa base de dados central única para, de seguida, as fichas de documento e prosopográficas serem alvo de normalização, por parte dos coordenadores.

3. Módulo de Fusão – depois do processo de normalização, neste módulo é feita uma transformação da informação, de modo a que a unidade de informação seja a relacionada com os clérigos. Até este ponto a informação era tratada em termos de documentos e fichas prosopográficas. À medida que as fichas prosopográficas associadas aos vários documentos vão sendo processadas, procede-se à junção dos dados dispersos relativos ao mesmo clérigo numa só ficha.

4. Módulo de Descoberta de Conhecimento – neste módulo pretende-se que sejam permitidas diversas formas evoluídas de análise dos dados reunidos. É feita então uma selecção e pré-processamento dos dados de forma a serem tratados por ferramentas, quer de análise multi-dimensional, quer de mineração de dados para ex-

tracção de conhecimento.

5. Módulo de Divulgação – nesta módulo final será feita a documentação da informação de forma a ter-se o produto pretendido, ou seja, uma base de dados com a informação do clero catedralício português para divulgação em cd-rom, via www, ou outra forma que se julgue eficaz e conveniente.

## 6.2 Concepção e Desenho dos Módulos

Depois de uma breve apresentação da arquitectura do SIEP, o que define os grandes blocos funcionais em que o sistema está estruturado e estabelece as ligações entre eles, passamos, agora, a descrever com mais detalhe cada módulo.

### Módulo de Aquisição

Para recolha da informação (fichas de documentos e prosopográficas) junto das fontes, será criada uma interface que permite a introdução destas fichas num formato análogo às actuais fichas em papel.

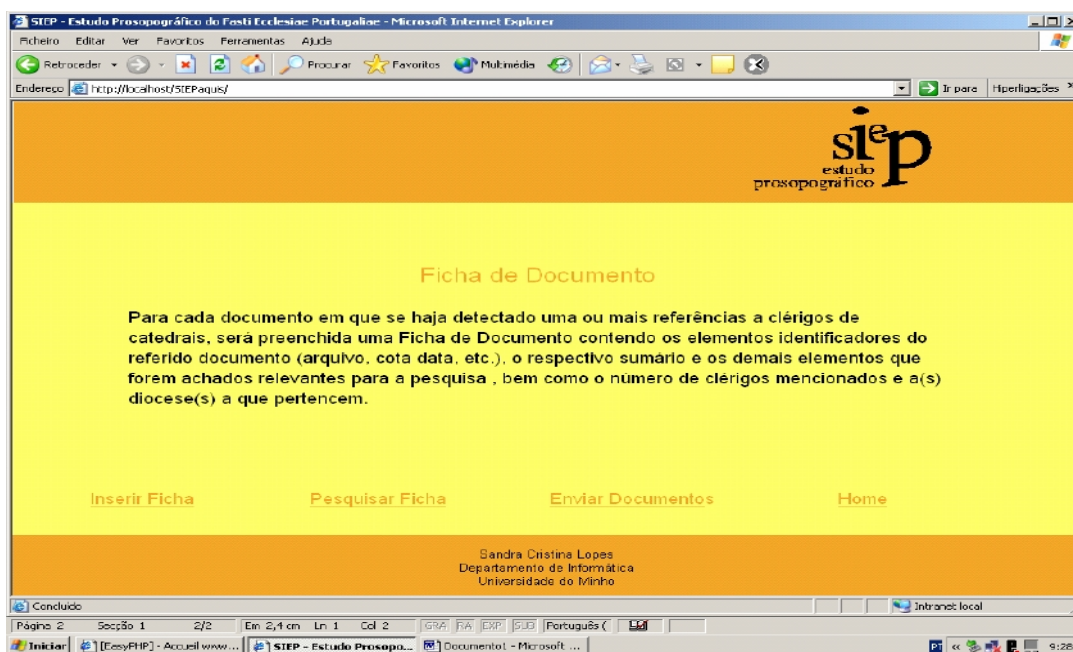


Figura 2: Ecrã que mostra o menu de opções para documentos

Essa interface estará instalada em cada computador pessoal (de secretária ou portátil) dos investigadores envolvidos no projecto, nas tarefas de recolha de material pelos arquivos diocesanos do País, e deverá funcionar em modo local e independente,

não dependendo de nenhuma base de dados central nem requerendo nenhuma ligação à internet. As Figuras 2 e 3 abaixo mostram as opções disponíveis para criar e trabalhar com fichas de documentos e dão uma ideia da ficha em causa.

The screenshot shows a web browser window titled "SIEP - Estudo Prosopográfico do Fasti Ecclesiae Portugaliae". The address bar shows "http://localhost/SIEP/quis/". The page has a yellow background and a logo in the top right corner that says "siep estudo prosopográfico". The main content is a form with the following fields:

- Arquivo:** ADB - Arquivo Distrital de Braga
- Fundo/Título:** Gaveta 1ª das Igrejas
- Maço/Livro:** nº99
- Nº Fólio:** [empty]
- Diocese(s):** Braga
- Tipo de Documento:** Sentença
- Entidade Emissora:** [empty]
- Data Cronológica:** Dous dias do mes de A
- Data Tópica:** [empty]
- Local de Redacção:** Braga
- Publicação:** [empty]
- Sinais de Validação:** Tabelião
- Nº de Clérigos da Catedral a fichar:** 3

At the bottom of the form, it says "Sandra Cristina Lopes, Departamento de Informática, Universidade do Minho". The browser's taskbar shows several open applications, including "GrandaFastiv5.doc - M...", "ecrãs.doc - Microsoft Word", and "SIEP - Estudo Prosopo...".

Figura 3: Ficha de Documento (vista parcial)

The screenshot shows the same web browser window, but now displaying the "Ficha Prosopográfica" page. The address bar is the same. The page content includes:

- A header with the "siep estudo prosopográfico" logo.
- A line of text: "Cota do Documento: Gaveta 1ª das Igrejas - nº99 - de Dous dias do mes de Abril Era de Mil e Trezentos e Sessenta e um anno . A fichar o 0º de 3 clérigos."
- A title "Ficha Prosopográfica" in orange.
- A paragraph of text: "Para cada bispo, dignidade, cónego ou porcionário de catedral referido em cada documento que deu origem a uma Ficha de Documento, será preenchida uma Ficha Prosopográfica contendo a respectiva identificação (nome próprio, patronímico, terceiro elemento do nome quando existente, categoria eclesíastica, diocese) assim como todos os dados que esse documento fornecer sobre as suas origens geográficas e familiares, a educação que recebeu, a sua carreira eclesíastica, os patronos que o impulsionaram e os dependentes que protegeu, os seus bens e rendimentos, etc."
- A navigation menu with four buttons: "Inserir Ficha", "Pesquisar Ficha", "Ficha de Documento" (highlighted in orange), and "Home".
- At the bottom, it says "Sandra Cristina Lopes, Departamento de Informática, Universidade do Minho".

The browser's taskbar shows the same applications as in Figure 3, plus "Fasti" and "Documento1 - Microsoft...".

Figura 4: Ecrã que mostra o menu de opções para fichas prosopográficas

De forma idêntica, as Figuras 4 e 5 ilustram as operações disponíveis para manusear as fichas prosopográficas e a própria ficha principal; informações adicionais (dados que a maioria das vezes

não existem, mas que, se encontrados, fornecem informação complementar importante) são introduzidas após criar a ficha base, como se constata na Figura 5a.

SIEP - Estudo Prosopográfico do Fasti Ecclesiae Portugaliae - Microsoft Internet Explorer

Endereço: http://localhost/SIEP/quiv/

**stiep**  
estudo  
prosopográfico

Cota do Documento: Gaveta 1ª das Igrejas - nº99 - de Dois dias do mes de Abril Era de Mil e Trezentos e Sessenta e um anno . A fichar o 0º de 3 clérigos.  
Hist:'sandra' cota: '21'

**Ficha Prosopográfica**

**Nome Próprio**

**Patronímico**

**Terceiro Elemento do Nome**

**Tipo do Terceiro Elemento**

**Posição no Documento**

**Categoria Eclesiástica**

**Descrição**

**Igreja**

**Diocese**

**Data Início**

Sandra Cristina Lopes  
Departamento de Informática  
Universidade do Minho

Concluído

Página 2 Seção 1 2/3 Em 1,1,7 cm Ln 1 Col 2 GRA PA EXP SUB Português ( )

Iniciar [EasyPHP] - Accuei www... SIEP - Estudo Prosopo... Documental - Microsoft ...

Figura 5: Ficha Prosopográfica base (vista parcial)

SIEP - Estudo Prosopográfico do Fasti Ecclesiae Portugaliae - Microsoft Internet Explorer

Endereço: http://localhost/SIEP/quiv/

**stiep**  
estudo  
prosopográfico

Cota do Documento: Gaveta 1ª das Igrejas - nº99 - de Dois dias do mes de Abril Era de Mil e Trezentos e Sessenta e um anno . A fichar o 1º de 3 clérigos.  
Código do clérigo em edição: 33

**Informação Adicional da Ficha Prosopográfica**

**Categorias Eclesiásticas Complementares**

- Ordens Sacras
- Dispensas
- Gracias em Expectativa/Concessão de Benefícios
- Colações
- Estatutos/Funções
- Formações
- Obras
- Parentes
- Dependentes
- Oficiais
- Bens

Sandra Cristina Lopes  
Departamento de Informática  
Universidade do Minho

Intranet local

Página 3 Seção 1 3/3 Em 2,4 cm Ln 1 Col 2 GRA PA EXP SUB Português ( )

Iniciar [EasyPHP] - Accuei www... SIEP - Estudo Prosopo... Documental - Microsoft ...

Figura 5a: Ecrã que dá acesso às Fichas de Informação Adicional

Tal como no processo manual, os dados serão introduzidos nestas duas fichas (figuras 3 e 5) textualmente, não se procedendo nesta fase a nenhuma operação de validação, ou normalização. À medida que os dados são inseridos, são armazenados numa base de dados local, à qual chamamos BDSiep.

Posteriormente, quando a inserção estiver completa, a informação relativa a um documento e respectivas fichas prosopográficas é exportada e guardada num ficheiro, para poder ser enviada para a sede, onde se juntará com as demais (ver Figuras 6 e 6a).

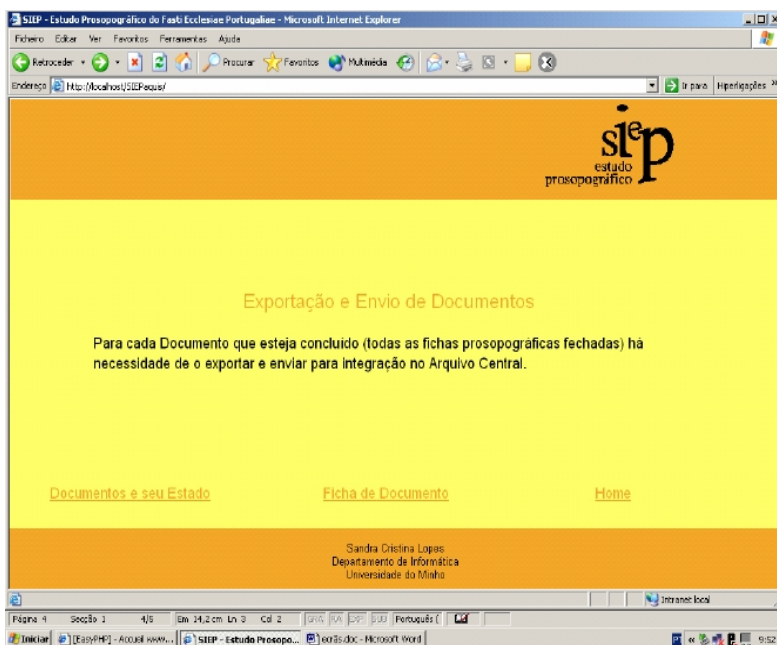


Figura 6: Ecrã de entrada no processo de Exportação de Fichas

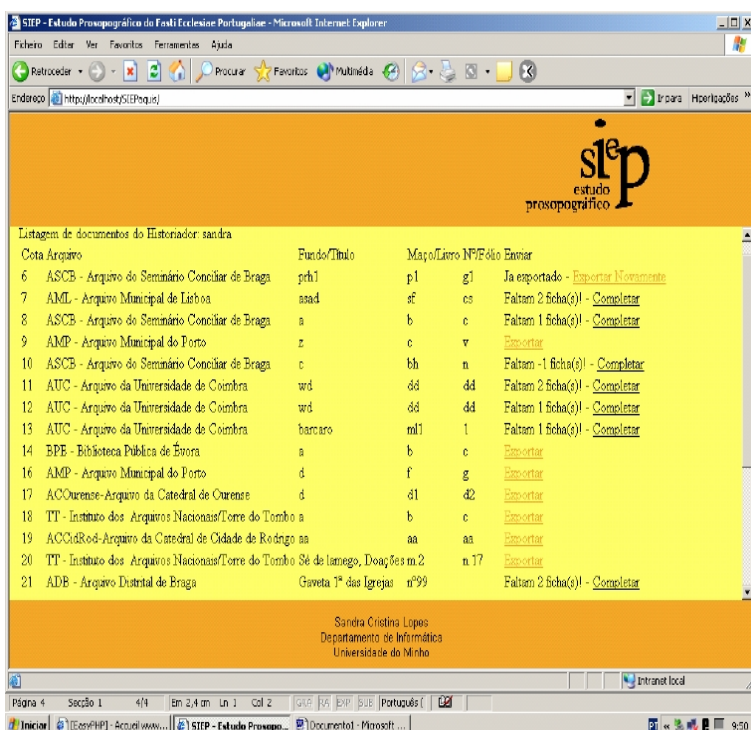


Figura 6a: Ecrã que mostra o estado das Fichas para selecção e Exportação

Embora se executam em dois passos separados, estas duas operações de exportação e envio, estão totalmente automatizadas, bastando que o Historiador indique qual o ficheiro que quer exportar, ou re-exportar, e confirme que o sistema pode proceder à gravação no ficheiro que será então enviado.

Através de um programa de importação, criado para o efeito, a informação contida neste ficheiro será inserida, no sistema central em Lisboa (sede do Projecto Fasti) após a recepção (via e-mail ou disquete/cd-rom), numa base de dados designada por BDSiepCentral. A Figura 7, abaixo, ilustra este processo, que também é executado com grande simplicidade visto estar também automatizado.

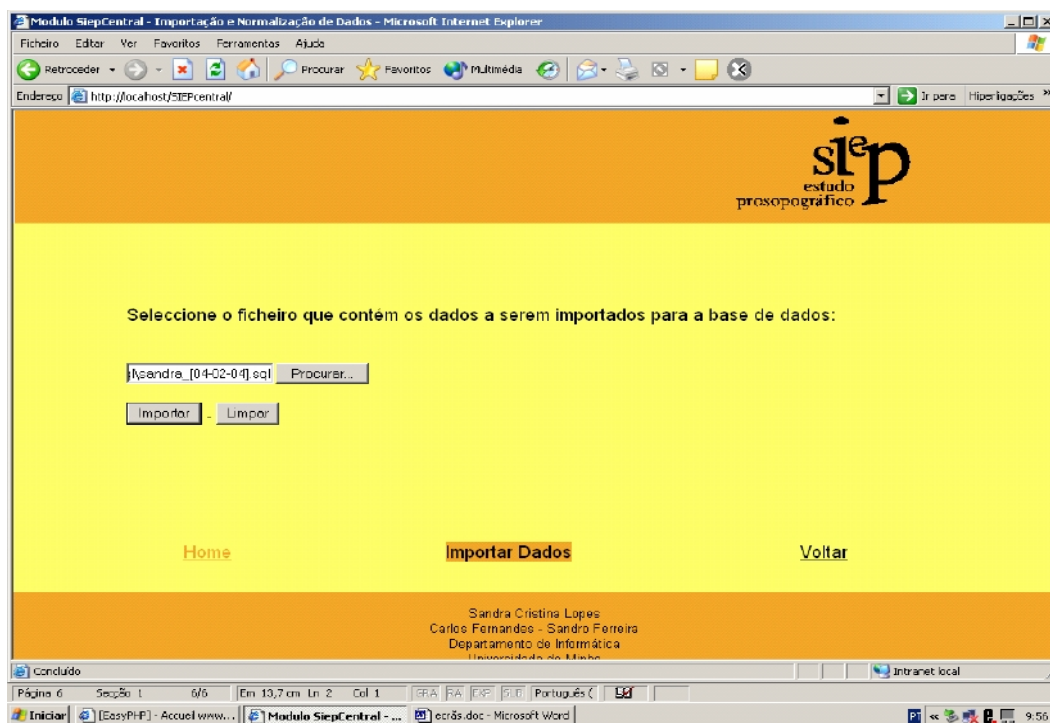


Figura 7: Ecrã que mostra o processo de Importação de Fichas

O envio (de um ou mais documentos) será realizado sempre que o responsável pela recolha o entenda (numa base diária, semanal, etc.).

### Módulo de Normalização

Relativamente a este processo, criou-se uma outra interface, mais uma vez baseada num browser mas agora com ligação às bases de dados centrais via internet.

Esta interface permite ao Coordenador especificar a sua diocese e proceder automaticamente à recolha (a partir da BDSiepCentral) de todas as fichas prosopográficas que digam respeito à diocese em causa. É importante referir que essas fichas prosopográficas vêm sempre associadas à respectiva ficha de documento.

Uma vez encontradas as fichas da diocese especificada, a interface permite ao Coordenador proceder à verificação dos dados recolhidos e à sua normalização. Essa normalização pode consistir em adoptar critérios de representação sistemáticos (do género: escrever os nomes das pessoas em maiúsculas e em latim; escrever o nome das localidades conforme original, mas em maiúsculas; etc.), ou em substituir o texto livre original por um valor padrão escolhido dentro de uma lista de valores possíveis. Esses valores padrão, são apresentados em certos campos das fichas de acordo com os critérios definidos pela Coordenação do projecto, critérios esses que estão armazenados em tabelas (também incluídas na base de dados) para poderem ser facilmente actualizados ao longo do decurso do projecto.

As Figuras 8 e 8a mostram, precisamente, a forma como se procede a esta operação de normalização, que é sempre um processo iterativo, isto é, a 1ª vez os novos campos (na coluna da direita) apa-

recem preenchidos com os valores iniciais, mas nas vezes seguintes (em que o historiador decida rever a ficha) verá os campos com os novos valores fixados até esse momento (na coluna da esquerda terá sempre indicação do valor textual original).

Original	Normalizado
Arquivo: Braga	Braga
Fundo: nao tem	Fundo
Mapa: nao tem	Mapa
Numero/Folio:123	1111
Diocese: Braga	Braga
Tipo Documento: Livro	Mandado régio
Codigo Entidade Emisora:Braga	4
Data Cronologica: 12-12-12	12-12-12
Data Topica: 12-12-21	12-12-21
Local Braga	Braga
Publicado: Sim	Não
Sinais/Val. Selo	Tabelião
Numero de Clerigos Fichados: 2	2
Sumarios: fdfdsfsdfd	ssssssssss

Figura 8: Ecrã que mostra o processo de Normalização da Ficha de Documento

Original	Normalizado
Codigo: sandra502	sandra502
Posicao do Clerigo no Documento: testemunha	Testemunha
Nome Proprio: aaaa	Testemunha
Patronimico: aaaa	Testemunha
Terceiro Elemento: aaa	Outorgante Activo
Tipo Terceiro Elemento: Ausente	Ausente
Data Nascimento: aaa	12-12-1310
Data de Obito: aaa	10-07-1388
Data Com. M: aa	10-07-1388
Local de Nascimento: aaa	Braga
Diocese de Nascimento: aaa	Braga
Local Sepultura: aaa	10-07-1388
Diocese de Sepultura: aaa	Braga
Prebendado	Sim
Observacoes: sasasa	Nada mais a registar!

Qual o estado da Normalizacao: Completa

Figura 8a: Ecrã que mostra o processo de Normalização da Ficha Prosopográfica

A Figura 9 e 9a ilustra esta característica iterativa da tarefa, pois permite ver os ecrãs que mostram o estado de normalização de um ou dos vários documentos.

ção, o Coordenador tiver alguma duvida, pode marcar o documento para que este seja posteriormente analisado por um grupo de coordenadores. Cabe a estes decidir se mantêm a informação original, ou qual a correcção que deve ser efectuada.

Se, após ou durante o processo de normaliza-



Figura 9: Ecrã correspondente à apresentação do status dos Documentos

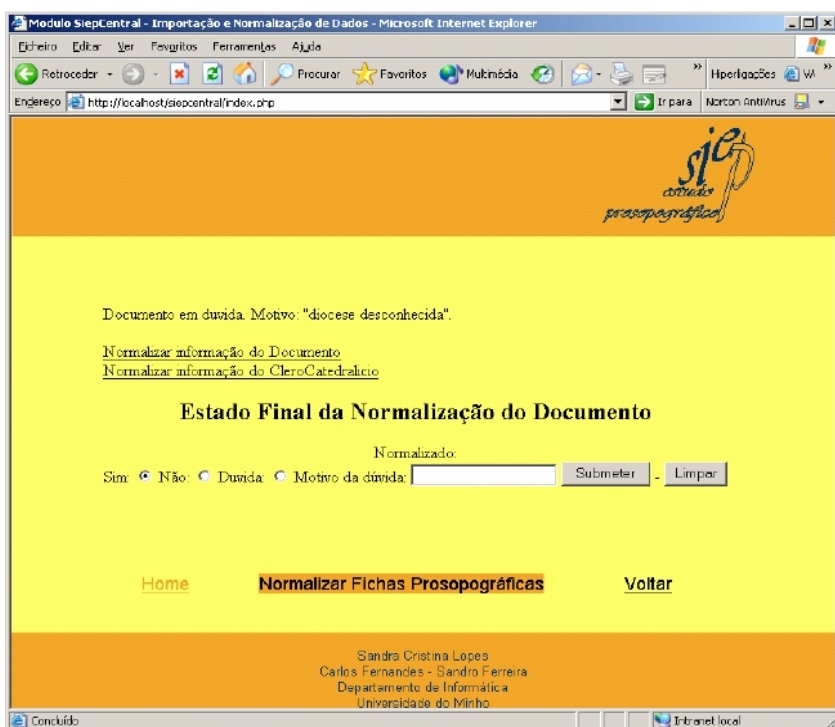


Figura 9a: Ecrã correspondente à apresentação do status de um Documento

É importante compreender que recorreremos a duas bases de dados para suporte a este módulo: a chamada BDSiepCentral, que possui a informação original de cada um dos documentos e que é semelhante à existente no Módulo de Aquisição com a única diferença de a tabela documentos possuir mais um campo que identifica o historiador que o elaborou e uma nova tabela controlo que irá conter o estado actual de cada um dos documentos (normalizado, não normalizado ou em duvida e a respectiva causa); e a BDCentral, uma base de dados que possui as mesmas tabelas da da anterior (com diferentes tipos para os campos categorizáveis<sup>13</sup>) mais um conjunto de tabelas auxiliares e de controlo usadas durante o processo de normalização.

#### Módulo de Fusão

Depois de cada Coordenador proceder à Normalização das suas fichas, pretende-se carregar a Base de Dados do Clero Catedralício (BDCC). Na arquitectura do SIEP esta tarefa é apoiada pelo módulo denominado de fusão que ainda não foi concretizado, mas que seguirá exactamente a mesma abordagem usada para implementar os módulos anteriores.

Tomando uma a uma as fichas individuais normalizadas, geradas no módulo anterior, este módulo irá pesquisar na BDCC todos os indivíduos homónimos que satisfaçam os demais critérios de semelhança. Feita esta pesquisa, mostra ao Coordenador o novo indivíduo em análise e a lista de todos os semelhantes, para que este tome uma de três decisões:

- Juntar a informação da nova ficha à ficha existente (fusão), por considerar que se trata do mesmo indivíduo;
- Inserir a nova ficha, por considerar que se trata de um indivíduo distinto de todos os apresentados;
- Transferir a nova ficha para um ficheiro de casos duvidosos (para serem tratados posteriormente em reunião do grupo), por considerar que não tem elementos suficientes para decidir se trata

de um indivíduo distinto ou não.

Como se disse atrás, a cada clérigo passa a corresponder uma e uma só entrada (ficha) na BDCC, a qual manterá então referencia para os vários documentos (registados na respectiva tabela dessa base de dados) de onde provieram os vários dados agora congregados nessa ficha prosopográfica única.

#### Módulo de Descoberta de Conhecimento

A partir da Base de Dados do Clero Catedralício, através de ferramentas de Análise de Dados Multidimensional (OLAP/EIS) e de Descoberta de Conhecimento (Data Mining), será possível analisar os dados de forma a obter respostas às várias questões levantadas:

- quais as origens de cada membro do clero;
- influência das origens na progressão na carreira dentro do clero;
- influência da proximidade do Rei na carreira dentro do clero;
- em que medida a fortuna prévia é importante;
- de que forma os estudos feitos (por ex: graus universitários obtidos) podem ajudar na progressão da carreira;
- até que ponto a rede de relações (protecção do papa, de um cardeal, através de laços familiares ou dependência) é um factor a considerar, etc;

#### Módulo de Divulgação

Este módulo, também ainda por implementar mas peça chave do sistema por permitir concretizar os objectivos do projecto, destina-se a consultar a BDCC das mais variadas formas (isto é, cruzando a informação de diferentes maneiras), tendo em vista a possibilidade de apresentar a informação, ou conhecimento, que ela contém em diferentes formatos: ecrán do computador; papel (relatórios); CD-ROM; ou páginas WWW.

A sua concretização terá, claramente, duas fases: numa primeira etape, a implementar logo após



o Módulo de Fusão, permitirá interrogar directamente a BDCC, usando as tecnologias clássicas; num segundo passo, a ser dado após o desenvolvimento do Módulo de Descoberta de Conhecimento, permitirá mostrar os resultados de explorar a BDCC com as ferramentas não-convencionais referidas atrás.

### 6.3 Requisitos de Robustez e Controlo do Fluxo

Como acontece com qualquer outra aplicação informática, uma preocupação que tem de estar sempre presente na concepção do sistema (para que se possa, mais tarde, implementar com sucesso) é a sua *robustez*, isto é, a segurança que ele oferece ao utilizador. Por um lado, para garantir que este não perde inesperadamente o seu trabalho; por outro lado (talvez até o mais importante neste contexto), para lhe transmitir a confiança necessária para que o use realmente.

Este conceito é demasiado abrangente e é usado, com frequência, em sentidos diversos.

Essencialmente, pode designar: a resistência e tolerância à falha<sup>14</sup>; a capacidade de protecção e defesa contra intrusos<sup>15</sup>; a satisfação dos requisitos lógicos do processo que asseguram a qualidade dos

dados e a coerência da informação<sup>16</sup>. Enquanto que os dois primeiros casos são tratados de maneira sistemática, o terceiro já é totalmente dependente da aplicação em causa; será, precisamente, aquele que se vai abordar nesta secção.

O acesso a qualquer um dos módulos do SIEP, descritos na secção anterior, será sempre feito através de um ecrã de login (ver Figura 10) no qual se pede a identificação do utilizador via username/password (Un/Pw). Com este mecanismo, controla-se o acesso ao sistema e regista-se o utilizador para que o seu username fique associado à informação que vai inserir e às operações que vai fazer (registo de autoria e mecanismo de responsabilização). No que respeita ao controlo de acesso e a título de exemplo, no ecrã da Figura 10, verifica-se se o utilizador tem autorização para usar o sistema (se o par Un/Pw é válido); no caso de ter acesso, definem-se as suas permissões. Por um lado, só haverá um utilizador (o administrador do SIEP) com permissão para fazer a importação de dados e a gestão das tabelas auxiliares de normalização (ver próxima secção); por outro lado, aproveita-se para identificar a que diocese o utilizador terá acesso de modo a só lhe mostrar as fichas dessa diocese quando ele entrar no processo de normalização.

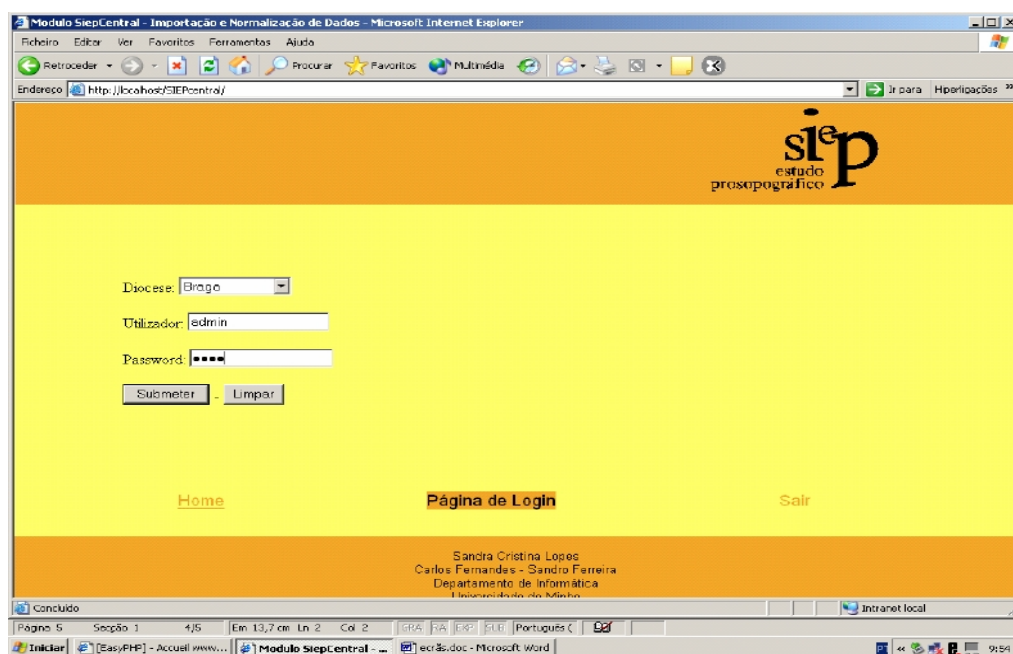


Figura 10: Login do Utilizador no sistema central para Importação e Normalização

Uma outra questão que se afigura de capital importância para manter os princípios exarados no método de trabalho descrito, é o controlo das fichas prosopográficas associadas a um documento.

Atendendo ao que foi dito, a unidade de trabalho, na fase de aquisição, é o documento o qual integra um determinado conjunto de fichas prosopográficas. No processo manual, essa integração é fisicamente assegurada pois as folhas de papel correspondentes são agrafadas à do documento; no sistema informático tínhamos que encontrar uma solução para manter a coesão sem tornar o sistema demasiado restrito e, como tal, pouco prático e cómodo de usar.

A decisão foi registar o número de fichas prosopográficas esperadas, aquando da introdução da ficha do documento, e depois ir controlando à

medida que as fichas prosopográficas são inseridas. O ecrã reproduzido na Figura 11 mostra a forma que o controlo assume: após a gravação de uma ficha prosopográfica verifica-se o total de fichas já inseridos e compara-se com o número esperado; se esse número ainda não foi atingido, surge uma mensagem a dizer quantas faltam e a operação disponível é a de voltar ao ecrã de fichas prosopográficas; caso se tenha chegado ao fim, então a mensagem dá conta desse facto e já surge como operação possível voltar à ficha de documento.

Os mesmos contadores são ainda usados para controlar a exportação. No ecrã patente na Figura 6a, acima, verifica-se que um documento só pode ser exportado se estiver completo, isto é, se o contador de fichas introduzidas e o número de fichas esperadas forem iguais.

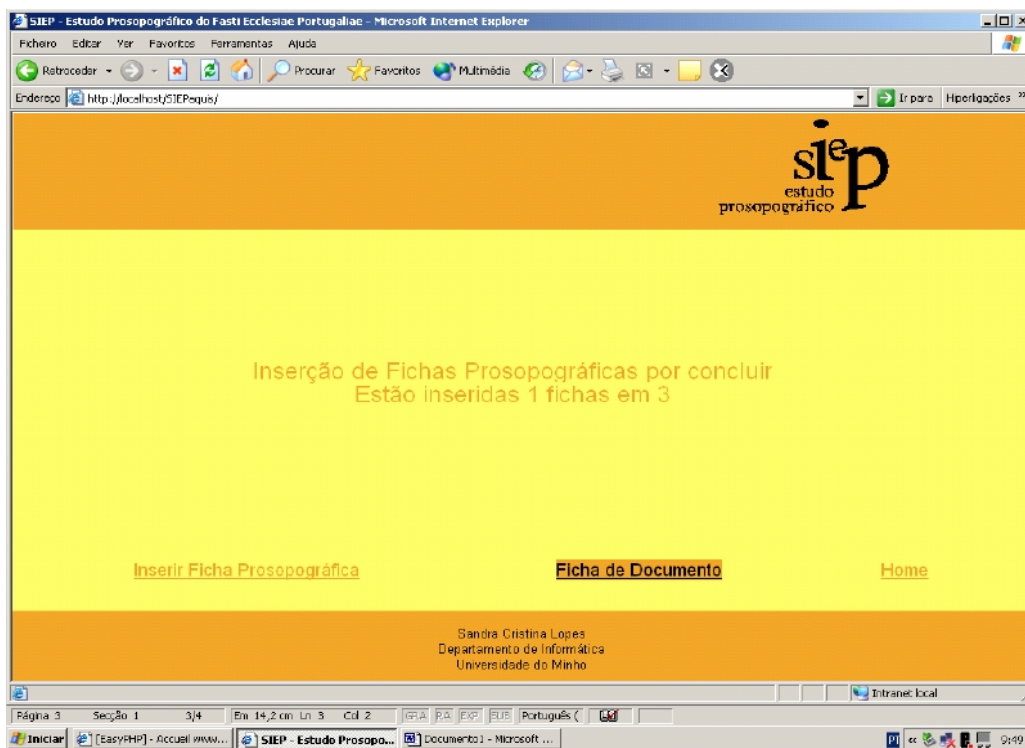


Figura 11: Controlo do preenchimento das fichas de um documento

### 6.4 Modelo de dados do SIEP

Para que se tenha noção do tipo de informação envolvida no SIEP, apresentamos o Modelo de Dados que resultou de uma análise profunda do problema com vista à identificação exaustiva e detalhada de todas as unidades de informação que terão de ser manuseadas. Este modelo ilustra a complexidade do sistema e especifica com rigor as bases de dados em que se vai suportar todo o funcionamento do sistema.

A Entidade Central é *Clero Catedralício*, em torno do qual giram todas as outras Entidades, das quais se listam a seguir as mais relevantes:

- \* Clero Catedralício;
- \* Ordens Sacras;

- \* Graças Expectativas;
- \* Colações;
- \* Formações;
- \* Obras;
- \* Bens;
- \* Últimas Vontades;
- \* Fundações;
- \* Documentos;
- \* Sinais de Validação;
- \* Indivíduos Relacionados.

Estas Entidades informáticas (que modelam todas as bases de dados atrás referidas) encontram-se relacionadas entre si de acordo com o Diagrama Entidades-Relações (DER) apresentado na Figura 12, a seguir.

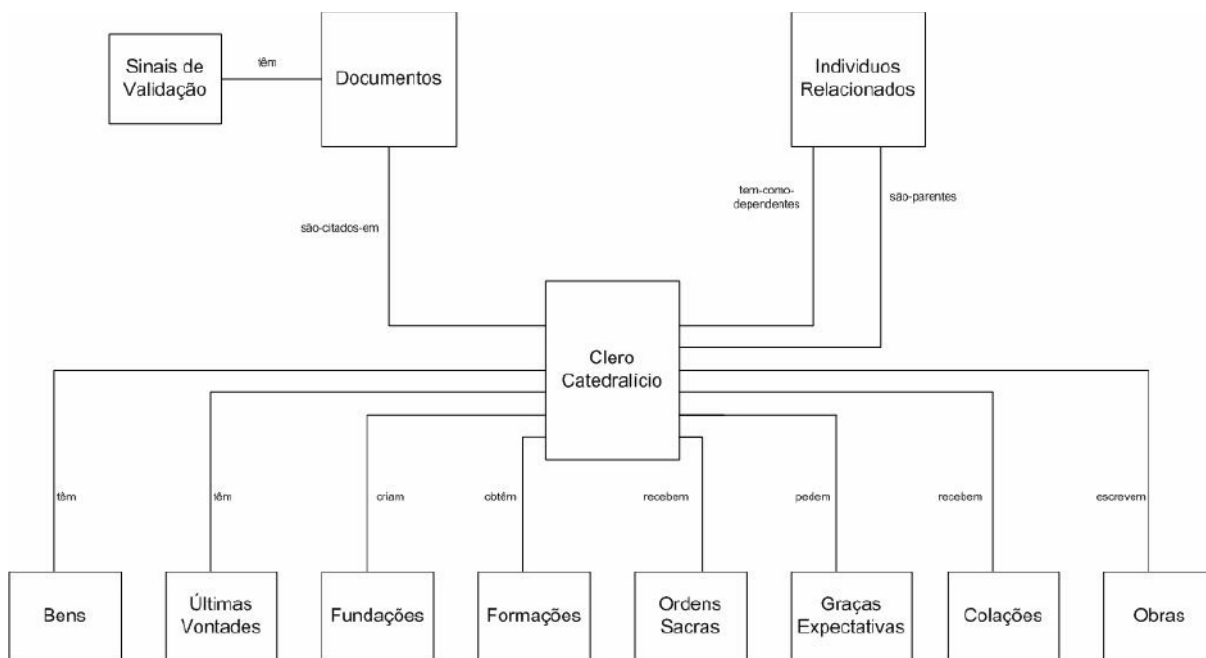


Figura 12: Modelo Conceptual de dados (DER) do SIEP

Porque representa uma mais valia do projecto, é importante enumerar as tabelas classificativas que foram criadas para suportar a normalização; fá-lo-emos através da Figura 13 que mostra o ecrã do

SIEP que dá acesso às operações de manutenção, ou de gestão<sup>17</sup> dessas tabelas (designadas no dito ecrã por Tabelas Auxiliares).



Figura 13: Gestão das Tabelas de Classificação para suporte à Normalização

## 6.5 Implementação e Instalação

Definida a arquitectura e o modelo de dado do sistema de informação, é necessário proceder à sua implementação, devendo para isso fazer-se a escolha do ambiente de trabalho e das ferramentas de desenvolvimento:

Para criar e manter a base de dados escolheu-se o motor de bases de dados MySQL — software livre, multi-plataforma e eficiente, que suporta grandes quantidades de dados. Para desenvolver os programas (quer a camada de interacção com o utilizador, quer as camadas com a lógica do sistema e o acesso aos dados), a escolha recaiu na linguagem de scripting PHP. Como é sabido esta linguagem permite uma programação estruturada e rápida e uma ligação simples às bases de dados MySQL<sup>18</sup>. Por outro lado, também permite programar com facilidade interfaces modernas e ágeis, que podem ser abertas por qualquer browser<sup>19</sup>, servindo assim de suporte, quer ao funcionamento local (off-line), quer ao acesso à distância via Internet (on-line).

Todos os pacotes de software podem ser descarregados via internet sem pagamento de licenças de utilização e existem quer para Linux quer para Windows. O único contra desta opção tecnológica, é que, em funcionamento local, é necessário que a máquina onde se vão correr os programas se comporte como um servidor HTTP.

Para resolver este problema e se proceder à instalação do software desenvolvido (mais concretamente, o Módulo de Aquisição) usou-se um pacote de software, easyPHP, para máquinas que possuam o sistema operativo MS Windows que permite de uma forma fácil, prática e rápida transformar essa máquina num servidor HTTP local, ou para a Internet (servidor Apache), com a capacidade de executar scripts em PHP; pacote esse que inclui já o motor de bases de dados MySQL pretendido. O easyPHP mostrou-se bastante útil e compatível com as aplicações deste projecto até então já desenvolvidas e, para além deste facto, comportou-se bastante bem nas várias máquinas onde foi instalado com o intuito de testar a sua compatibilidade (máquinas com MS Windows 95, 98, ME, 2000 e XP).

Precisamente a mesma configuração foi instalada no computador central que vai fazer de servidor para o projecto, sendo então aí instalado, depois, o Módulo de Normalização.

## 7. Conclusão

A comunicação que aqui se termina teve dois objectivos principais. Por um lado, pretendeu-se apresentar o Projecto Fasti Ecclesiae Portugaliae — Prosopografia do clero catedralício português, em curso, envolvendo investigadores de departamentos de História de diversas universidades portuguesas. A preocupação foi caracterizar bem o objecto de estudo, as fontes a consultar e as interrogações a colocar para, então, se definir o método a seguir com vista a atingir o objectivo do projecto: a construção de uma base de dados normalizada do Clero Catedralício que permita, posteriormente, estudar este importante grupo da sociedade portuguesa da Idade Média, entre a restauração da primeira diocese (1071) e a “pacificação” das suas relações com a monarquia. Por outro lado, quis-se introduzir o SIEP, o sistema de informação concebido e desenvolvido para dar suporte ao projecto, tendo-se descrito a sua arquitectura, funcionalidade dos módulos que o constituem e o modelo de dados de suporte e discutido as decisões que lideraram a sua implementação.

Contudo, a motivação essencial foi sistematizar a abordagem seguida na informatização e consolidar o método de trabalho para desenvolver software de apoio a projectos na área das ciências sociais (método este que já fora testado no âmbito da demografia histórica, etc.). Neste contexto a qualidade das interfaces é fundamental: tem de ser claras e fáceis de usar e mais possível imunes ao ruído para evitar os erros; além disso, todo o processamento deve ser feito de tal forma que inspire segurança/confiança. Para isso é crucial que se siga de perto o método manual com que os historiadores estão familiarizados.

Os dados a recolher são muitos (fichas muito extensas) e a informação é muito complexa (relações várias, multi-entidades e por vezes reflexivas); além disso, esta informação é tipicamente incom-

pleta (em muitos casos as fontes não fornecem dados para preencher todos os campos das fichas), ou incerta/difusa (mesmo encontrando dados, muitas vezes surgem dúvidas sobre os valores lidos), mas é precisamente com esta que temos de lidar e sobreviver. Por tudo isso, com o SIEP pretende-se seguir de perto as directrizes do método prosopográfico concebido para o projecto FASTI, pelo que se desenhou a sua arquitectura e concebeu os respectivos módulos a partir das várias fases desse método; também o modelo de dados e o desenho das interfaces seguiu a estrutura e o conteúdo das fichas pensadas para usar em papel. É importante ainda referir que a palavra Normalização foi uma constante neste projecto. Após a recolha de informação nas fontes, totalmente desprovida de normalização, quer sintáctica quer semântica, é feita a categorização dos dados com base em tabelas classificativas. Mais tarde, no processo de junção dos dados dispersos pelas várias fichas, será fulcral pensarmos em evitar a redundância da informação e em aplicar o algoritmo, tão útil no mundo da História, de fusão de informação de um indivíduo que terá de ser identificado entre um conjunto de hipóteses (indivíduos semelhantes).

O processo vai a meio do caminho, mas das componentes do sistema, já desenvolvidas e provisoriamente instaladas para teste ([www.fasti.ucp.pt](http://www.fasti.ucp.pt)), fica a clara indicação de que, quer a nível das ferramentas usadas, quer a nível de metodologia seguida para a sua implementação, estamos no caminho certo. Fica assim para trabalho futuro reflectir sobre estas questões e o desenvolvimento das três fases restantes do projecto SIEP (Módulo de Fusão, Módulo de Divulgação e o Módulo de Descoberta do Conhecimento) bem como alguns acertos e correcções que sejam necessários aos módulos desenvolvidos até ao presente momento.

Pensamos que o sucesso, ou não, deste modelo pode contribuir para avanços significativos em outros estudos semelhantes no domínio da História.

## Notas de Rodapé

<sup>1</sup> A equipa é constituída por Ana Maria Jorge (Universidade Católica Portuguesa); Maria Helena da Cruz Coelho, Saul António Gomes e Maria do Rosário Morujão (Universidade de Coimbra); Armando Luís de Carvalho Homem, Maria Cristina Cunha e Luís Amaral (Universidade do Porto); Ana Maria Rodrigues e Bernardo Sá Nogueira (Universidade de Lisboa), Hermínia Vasconcelos Vilar (Universidade de Évora); Maria João Branco (Universidade Aberta); Anísio Saraiva (bolseiro da FCT, doutorando da Universidade de Coimbra); Maria Justiniana Maciel Lima e Maria Antonieta Moreira da Costa (professoras do Ensino Básico/Secundário, mestres pela Universidade de Minho)

<sup>2</sup> Os bolseiros que têm trabalhado no âmbito deste projecto têm sido: António Castro Henriques que cessou a sua colaboração em Junho de 2003; Marta Castelo Branco; Mário Farelo; André Evangelista Oliveira Marques.

<sup>3</sup> Este número é deduzido a partir dos estudos já elaborados para dioceses como Braga, Évora e Lamego. Assim, no caso de Braga, entre 1245 e 1374, encontram-se identificados 63 dignidades e 182 cónegos; enquanto para Évora para o período compreendido entre o final do século XII e 1423, se encontra um total de 209 indivíduos e para Lamego, entre 1296 e 1349, um total de 61.

<sup>4</sup> A primeira edição dos *Fasti Ecclesiae Anglicanae*, de Le Neve, data de 1716, e foi ampliada em 1847 por Cotton e 1854 por Hardy, antes de ser de novo actualizada a partir dos anos 60 do século XX. Em França, quase todas as dioceses dispõem de catálogos antigos de capitulares, muitas vezes feitos por eclesiásticos, como se pode comprovar pelos volumes já editados de *Fasti Ecclesiae Gallicanae*, vol. 1 – Diocèse d’Amiens, par Pierre Desportes et Hélène Millet, Turnhout, Brepols, 1996, vol.2 – Diocèse de Rouen, par Vincent Tabbagh, Turnhout, Brepols, 1998, vol. 3 – Diocèse de Reims, par Pierre Desportes Turnhout, Brepols, 1998, vol. 4 – Diocèse de Besançon, par Henri Hours, Turnhout, Brepols, 1999; vol. 5 – Diocèse d’Agen, par Fabrice Ryckebusch, Turnhout, Brepols, 2001; vol. 6 – Diocèse de Rodez, par Mathieu Desachy, Turnhout, Brepols, 2002.

<sup>5</sup> Fortunato de Almeida (*História da Igreja em Portugal*, 1ª ed., 1910-1928) baseou-se nas obras anteriores de P. B. Gams (*Series Episcoporum Ecclesiae Catholicae*, 1ª ed., 1873) e C. Eubel (*Hierarchia Catholica Medii Aevi*, 1ª ed., 2 tomos, 1913-1914); por sua vez, Miguel de Oliveira (*História Eclesiástica de Portugal*, 1ª ed., 1940) apoiou-se nele. As listas deste último foram recentemente actualizadas por Ana Maria C. M. Jorge (“*Episcopologio*”, *Dicionário de História Religiosa de Portugal*, vol. C-I, Lisboa, Círculo de Leitores, 2000) a partir dos muitos estudos de pormenor desde então publicados sobre a questão, mas continua a faltar uma pesquisa de base que o projecto agora lançado permitirá, finalmente, levar a cabo.

<sup>6</sup> Braga, Évora e Lamego são, como já referimos, algumas das diocese trabalhadas, se bem que nem sempre na mesma cronologia definida pelo actual estudo. A primeira foi objecto de um projecto, também financiado pela FCT e que decorreu entre 1997 e 1999, subordinado ao tema “O poder económico, social e político do Cabido da Sé de Braga em finais da Idade Média (sécs. XIII-XVI)”, tendo como Investigadora Responsável Ana Maria S. A. Rodrigues. Entre os trabalhos dele resultantes realcem-se, como mais directamente ligados ao âmbito do projecto agora apresentado, as dissertações de Maria Antonieta Moreira da Costa, *O Cabido de Braga na segunda metade da centúria de Duzentos (1245-1278)*, Braga, Universidade do Minho, 2000 e de Maria Justiniana Pinheiro Maciel Lima, *O Cabido de Braga no tempo de D. Dinis (1278-1325)*, Lisboa, Patrimonia, 2003. As dioceses de Évora e de Lamego foram, respectivamente, estudadas por Hermínia Vasconcelos Vilar, *As Dimensões de um poder. A diocese de Évora na Idade Média*, Lisboa, Estampa, 1999 e por Anísio Miguel de Sousa BemHaja Saraiva, *A Sé de Lamego na primeira metade do século XIV (1296-1349)*, 2 vols, *Dissertação de Mestrado*, Faculdade de Letras, Universidade de Coimbra, 1999.

<sup>7</sup> Situação que não se repete para os séculos seguintes, onde já se consegue ter acesso a listas, se bem que não totalmente exaustivas, dos cónegos catedralícios.

<sup>8</sup> As histórias das dioceses (como, por exemplo, as de D. Rodrigo da Cunha, *História Eclesiástica dos Arcebispos de Braga...*, 2 vols., Braga, 1634-1635, *História Eclesiástica da Igreja de Lisboa...*, Lisboa, 1642, e *Catálogo dos Bispos do Porto*, Porto, 1742) fornecem alguns dados importantes, mas não são suficientes para este propósito.

<sup>9</sup> Várias dioceses galegas e castelhanas tinham jurisdição sobre territórios portugueses: Tui no espaço entre os rios Minho e Lima, Cidade Rodrigo sobre a zona de Riba-Coa, Badajoz sobre os concelhos de Olivença, Campo Maior e Ouguela. Quanto às dioceses de Lamego e Guarda, Évora e Silves, estavam submetidas, respectivamente, a Compostela e Sevilha. Em contrapartida, o arcebispo de Braga exercia autoridade metropolitana sobre as Sés de Astorga, Tui, Lugo, Orense e Mondonhedo. Cf. A. H. de Oliveira Marques, *Portugal na Crise dos Séculos XIV e XV*, Lisboa, Presença, 1986, pp. 365-366.

<sup>10</sup> Hélène Millet, *Les Chanoines du Chapitre Cathédral de Laon 1272-1412*, Rome, Ecole Française de Rome, 1982; Jacques Pycke, *Le Chapitre Cathédral de Notre-Dame de Tournai de la fin du XI<sup>e</sup> à la fin du XIII<sup>e</sup> siècle. Son organisation, sa vie, ses membres*, Louvain-la-Neuve/Bruxelles, 1986 e Répertoire Biographique des Chanoines de Notre-Dame de Tournai 1080-1300, Louvain-la-Neuve/Bruxelles, 1988; Elisabeth Mornet, “Préliminaires à une prosopographie du haut clergé scandinave: le chapitre cathédral de Roskilde 1367-1493”, in *Medieval Lives and the Historian. Studies in Medieval Prosopography*, ed. N. Bulst, J.-Ph. Genet, Medieval Institute Studies, Western Michigan University, 1986, pp. 139-162; David Lepine, *Brotherhood of Canons Serving God. English Secular Cathedrals in Later Middle Ages*, Woodbridge, The Boydell Press, 1995.

<sup>11</sup> Armando Luis de Carvalho Homem, *O Desembargo régio (1320-1433)*, Porto, INIC-Centro de História da Universidade do Porto, 1990; Eugénia Maria Pereira da Mota, *Do Africano ao Príncipe Perfeito. Percursos da Burocracia Régia*, 2 vols., dissertação de Mestrado, Faculdade de Letras, Universidade do Porto, 1989; Judite Gonçalves de Freitas, *A Burocracia do “Eloquente” (1433-1438). Os Textos, as normas, as gentes*, Cascais, Patrimonia, 1996; Idem, “Teemos por bem e mandamos”. *A Burocracia régia e os seus oficiais em meados de Quatrocentos (1439-1438)*, Cascais, Patrimonia, 2001; Vasco Rodrigo dos Santos Machado Vaz, *A Boa Memória do Monarca: Os escrivães da chancelaria de D. João I (1385-1433)*, 2 vols., dissertação de Mestrado, Faculdade de Letras, Universidade do Porto, 1995; Armando Paulo Carvalho Borlido, *A Chancelaria régia e os seus Oficiais em 1463*, dissertação de Mestrado, Faculdade de Letras, Universidade do Porto, 1996; Ana Paula Pereira Godinho Almeida, *A Chancelaria Régia e os seus Oficiais em 1462*, dissertação de Mestrado, Faculdade de Letras, Universidade do Porto, 1996.

<sup>12</sup> Sobre esta questão vejam-se, entre outros, os estudos coligidos em duas colectâneas, relativamente distantes no tempo mas que representam dois pontos de situação sobre o debate suscitado pela aplicação deste método. Falamos, respectivamente, das colectâneas *Prosopographie et genèse de l’État Moderne*, ed. Françoise Autrand, Paris, 1986 e *Resources Sources*, ed. K. S. B. Keats-Rohan, Oxford, 2002.

<sup>13</sup> Note-se que na BDSiepCentral todos os campos são do tipo texto para permitir a introdução livre da informação retirada da fonte; na BDCentral já alguns desses campos passam a ser de tipos enumerados para se restringir o seu valor à lista de hipóteses previstas.

<sup>14</sup> Mecanismos que asseguram que o sistema não avaria (não interrompe abruptamente o seu funcionamento), ou que em caso de falha recuperam o estado e os dados sem transtornos para o utilizador (a política de back-up é um dos truques usados mais simples).

<sup>15</sup> Mecanismos que vão desde o controlo de acesso à aplicação, estabelecendo níveis de permissão (de impedido, até pode fazer todas as operações sobre todos os dados, passando por permissões intermédias de só poder consultar, ou só poder alterar uma parte dos dados), até aos processos criptográficos para encriptar a informação que transita nos canais de comunicação e assim assegurar a sua confidencialidade e invulnerabilidade quanto a escutas ou outras formas de pirataria.

<sup>16</sup> Mecanismos que dão a garantia de que o utilizador da aplicação cumpre os preceitos de boas práticas e faz as operações nos momentos correctos sobre os dados correctos.

<sup>17</sup> Operações de inserção de valores em cada tabela prevista e previamente criada, de consulta, alteração e remoção de valores; a criação e eliminação de tabelas não são operações permitidas.

<sup>18</sup> Além de que existe um bom sistema de gestão para estas bases de dados (SGBD) em PHP, o PhpMyAdmin, que é usado diariamente no desenvolvimento do projecto e nos testes de instalação.

<sup>19</sup> Leitor de páginas HTML que permite aceder a hiper-documentos na World-Wide Web (WWW).

## Reconstrucción de Familias y Trayectorias de Vida a Partir del Análisis de Datos Locales

Laura Lorenzo Carrascosa\*

### 1. Introducción

El análisis microanalítico ha sido uno de los enfoques con mayor tradición en la demografía histórica. Este análisis permite realizar descripciones detalladas de las estructuras demográfica y familiar de un lugar determinado, así como del contexto social y económico en el que se inscriben (Reher, 2000: 29).

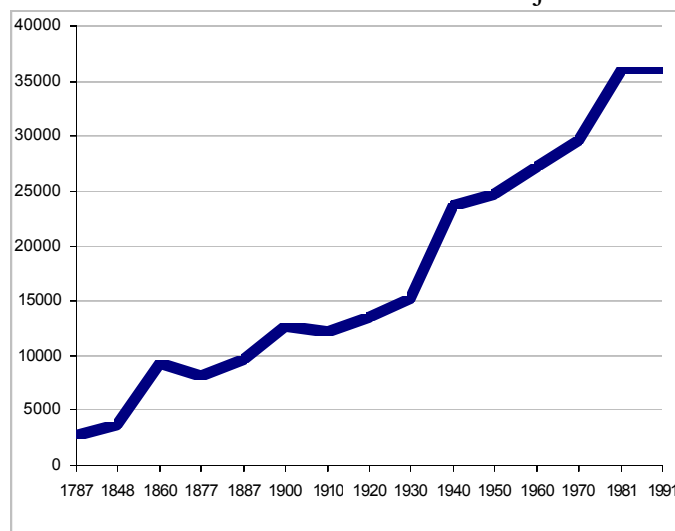
Si bien esta perspectiva analítica cuenta con muchos detractores, debido a la escasa representatividad de los resultados y a la dificultad para estudiar los procesos de cambio, son indiscutibles las posibilidades de estudio que presenta en lo que a los comportamientos demográficos se refiere, en donde prima la experiencia individual y familiar. Asimismo, estos inconvenientes aducidos a los estudios microanalíticos y a la reconstrucción de familias, son solventables cruzando información de diferentes fuentes no sólo demográficas, sino también eco-

nómicas, culturales y sociales, que profundicen en las causas de los comportamientos individuales y familiares (Reher, 2000: 63).

La reconstrucción de familias se fundamenta en la explotación de fuentes estadísticas oficiales como son el Censo de Población y Vivienda y el Registro Civil, entre otras. Si bien muchas de las fuentes demográficas tienen el problema de que los datos se presentan de forma agregada, debido al compromiso de preservar la confidencialidad, la disponibilidad de datos nominales que nos permita identificar a los sujetos facilita la reconstrucción de las trayectorias de vida de los individuos y las familias. Esta es la metodología empleada en el proyecto de investigación de Aranjuez.

La investigación Procesos de modernización social y demografía en la España interior. Dinámicas individuales y familiares, 1870-1990, se centra en el análisis de la población de la localidad madrileña de Aranjuez durante el período comprendido entre 1870 y 1990.

**Gráfico 1:** Evolución de la Población de Aranjuez. 1877-1981



Fuente: Base de datos del proyecto de investigación.

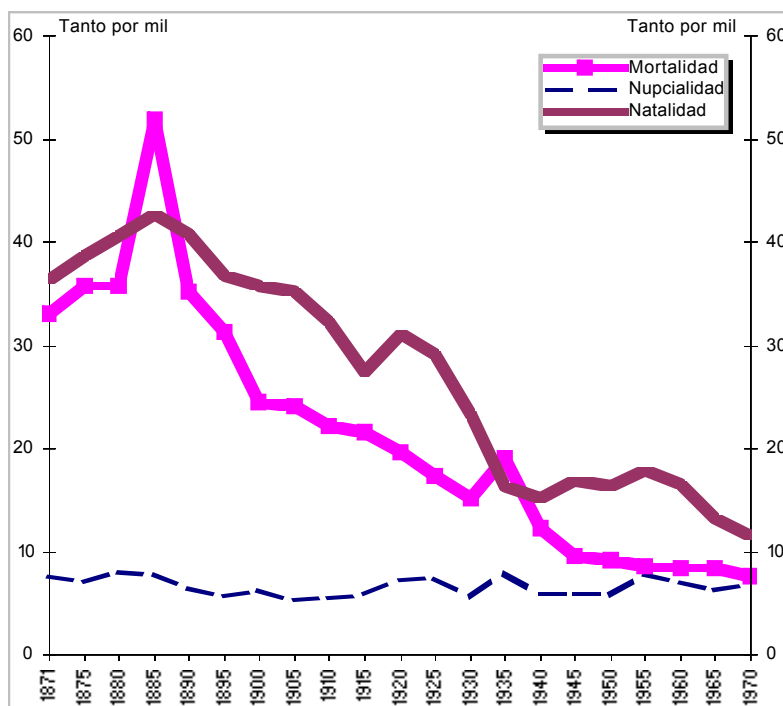


En estos años la población de Aranjuez experimenta un importante crecimiento como se puede observar en el Gráfico 1, pasando de los 2.653 habitantes recogidos en el Censo de Floridablanca en 1787, a 15.245 en 1930 y a una población de 35.997 en el año 1991<sup>1</sup>.

Si bien durante el siglo XIX, momento en el que comienza esta investigación, Aranjuez es fundamentalmente una localidad agrícola con un sector

servicios bastante importante, tras la Guerra Civil se convierte en un centro de atracción migratoria y de implantación de una actividad fabril nada despreciable. Estas características de la modernización social y económica de Aranjuez, junto con un proceso de cambios demográficos similar al del resto de España y Europa que queda reflejado en el Gráfico 2, convierten a Aranjuez en un buen exponente de las tendencias de la población en el período denominado de la transición demográfica.

**Gráfico 2:** Evolución de las Tasas Brutas de Mortalidad, Nupcialidad y Natalidad en Aranjuez. 1871-1970



Fuente: Base de datos del proyecto de investigación.

El proyecto de investigación tiene como objetivo estudiar el proceso de modernización de la sociedad madrileña y española. Para ello se ha procedido a la identificación, recopilación e informatización de la información de distintas fuentes de carácter demográfico y socioeconómico de Aranjuez en el período comprendido entre 1871 y 1975. La base de datos elaborada, integra la información de diversas fuentes tales como el Registro Civil, el Padrón

Municipal de Habitantes, las medidas antropométricas de los quintos y otras relativas a renta y patrimonio de fuentes paralelas.

La información aportada por estas fuentes permite analizar de forma longitudinal diferentes aspectos del proceso de modernización demográfica desde una perspectiva individual.

## 2. Diseño y creación de la base de datos.

Como se ha dicho anteriormente, el objeto de esta investigación es estudiar diferentes aspectos del proceso de modernización social y demográfica de Madrid y España.

La base de datos creada reúne la información facilitada por distintas fuentes tales como las partidas de nacimientos, matrimonios y defunciones del Registro Civil, los Padrones de población, las medidas antropométricas de los quintos y otras referentes a rentas y patrimonio, durante el período 1871-1975.

Pasaré ahora a analizar de forma más detallada la información facilitada por el Registro Civil y el Padrón Municipal de Habitantes, principales fuentes utilizadas en la creación de la base de datos de Aranjuez<sup>2</sup>.

### 2.1. Los datos del Registro Civil.

El Registro Civil es el origen de la creación de la base de datos de Aranjuez, debido a que la información que dicha fuente aporta ha permitido elaborar las trayectorias de vida de los individuos de la localidad, así como identificar y registrar sus familias y vínculos de parentesco<sup>3</sup>.

La informatización del Registro Civil ha posibilitado la integración de 40.279 partidas de nacimiento, 10.893 de matrimonio y 29.789 partidas de defunción<sup>4</sup>. La información proporcionada por estas partidas además de permitir identificar a los individuos, a sus padres y, en el caso de las partidas de nacimiento, a sus abuelos, aporta información acerca de la alfabetización, la edad, el oficio y los orígenes, entre otros datos<sup>5</sup>, como se puede observar en las Tablas 1, 2 y 3.

**Tabla 1.** Información de los Nacimientos Incluida en la Base de Datos

<b>Nacido</b>	<b>Padres</b>	<b>Abuelos</b>
Código numérico identificativo	Nombre	Nombre
Día	Segundo apellido	Edad
Mes	Edad	Localidad de origen y provincia
Año	Estado civil	Observaciones
Nombre	Oficio	
Primer apellido	Localidad de origen y provincia	
Segundo apellido	Observaciones	
Sexo		
Legitimidad		
Observaciones		

**Tabla 2.** Información de los Matrimonios Incluida en la Base de Datos

<b>Datos del matrimonio</b>	<b>Cónyuges</b>	<b>Padres</b>
Código numérico identificativo	Nombre	Nombre
Día	Primer apellido	Observaciones
Mes	Segundo apellido	
Año	Edad	
	Estado civil	
	Localidad de origen y provincia	
	Lugar del que es vecino	
	Oficio	
	Observaciones	

Tabla 3. Información de las Defunciones Incluida en la Base de Datos

<b>Difunto</b>	<b>Cónyuge</b>	<b>Padres</b>
Código numérico identificativo	Nombre	Nombre
Día	Primer apellido	Localidad de origen y provincia
Mes	Localidad de origen y provincia	Oficio
Año	Oficio	Observaciones
Nombre	Observaciones	
Primer apellido		
Segundo apellido		
Sexo		
Legitimidad		
Edad (años, meses y días)		
Estado civil		
Localidad de origen y provincia		
Oficio		
Expresión literal de la causa de muerte		
Observaciones		

Posteriormente a la informatización del Registro Civil en fichas, se llevó a la práctica un proceso de validación interna de la información, cotejando dichas fichas electrónicas con las actas microfilmadas. La revisión manual permitió recuperar numerosa información, que debido a errores en las actas originales o bien en la lectura y transcripción de los datos, hubiera sido excluida de la base de datos. Este es el caso de las dificultades para entender la grafía de los nombres y apellidos, de las localidades y de los diagnósticos de las causas de defunción. Con el objetivo de rescatar la mayor información posible, las localidades recogidas en las partidas del Registro Civil han sido comparadas con Nomencladores y mapas de la época. En cuanto a las causas de defunción, su tratamiento ha sido diferente, ya que ha sido necesario identificar las distintas enfermedades en diccionarios médicos, para posteriormente codificarlas según su correspondencia con las categorías de la Segunda

Nomenclatura de la Primera Clasificación Internacional de Enfermedades de Bertillon (1899) y sus equivalentes en la Clasificación Modificada de McKeown (1978)<sup>6</sup>.

## 2.2. El Padrón Municipal de Habitantes.

La información facilitada por los Padrones de Habitantes ha complementado la base de datos creada a partir del Registro Civil de Aranjuez. La bondad de esta fuente es que su perspectiva de análisis es el hogar, entendiéndolo como unidad de coresidencia, lo que aporta un punto de vista más complejo a la información ya obtenida<sup>7</sup>. Además de proporcionar datos nominales de los integrantes de un hogar, el Padrón contiene información relativa a la edad, parentesco, oficio y estado civil, entre otros datos disponibles para los distintos miembros del hogar en cuestión, como se puede observar en la Tabla 4.

Tabla 4. Información del Padrón Recogida en la Base de Datos

Código numérico identificativo	Fecha de nacimiento	Localidad y provincia de nacimiento
Número de folio	Edad (años, meses y días)	Localidad y provincia de residencia
Distrito y sección	Parentesco	Año de llegada
Nombre y número de la calle	Estado civil	Año, meses o días de residencia
Número de hogar	Instrucción	Clasificación como habitante
Nombre	Formación	Estado de residencia
Primer apellido	Oficio	Ausencia
Segundo apellido	Localidad de oficio	Observaciones
Sexo	Sueldo	

Con este propósito de contar con un enfoque de estudio a nivel de hogar, fueron informatizados los Padrones realizados durante el período que abarca esta investigación. Estos padrones han permitido integrar en la base de datos de Aranjuez: 7.657 sujetos en el caso del Padrón de 1877, 9.543 en 1903, 10.019 en el de 1915, 20.475 en el Padrón del año 1945, 27.197 en 1960 y 31.004 en el caso del Padrón de 1975.

Al igual que se hizo con la información informatizada proveniente del Registro Civil, los Padrones también han requerido un proceso de validación interna manual, cotejando los datos informatizados con los padrones microfilmados o en papel según fuera el caso. La revisión de los Padrones ha puesto de manifiesto la inconsistencia de parte de los datos derivada de errores en los Padrones originales, dificultades en la lectura e inexactitudes en la transcripción de los mismos. En este sentido, uno de los mayores inconvenientes del Padrón ha sido el derivado de los errores y omisiones de los datos de residencia. Por este motivo y con la finalidad de poder localizar espacialmente los hogares e individuos de la base de datos, ha sido necesario cotejar y solventar los errores relativos a la residencia consultando mapas y callejeros de la ciudad en los diferentes momentos padronales.

### 3. Metodología utilizada en el estudio. El proceso de Acoplamiento.

La base de datos elaborada integra por tanto la información relativa a los acontecimientos vitales inscritos en el Registro Civil de Aranjuez: nacimientos, matrimonios y defunciones, así como los datos facilitados por el Padrón Municipal de Habitantes.

Partiendo de una perspectiva en la que el sujeto tiene interés de estudio tanto desde su análisis individual, como de integrante de una unidad familiar<sup>8</sup> y participante del entorno social, el individuo es localizado en todos los acontecimientos vitales en los

que participa<sup>9</sup>. Con este propósito se ha llevado a cabo el proceso de acoplamiento, que permite identificar a los individuos tanto en los sucesos inscritos en el Registro Civil como en los diferentes padrones, mediante la utilización de la información nominativa (nombres y apellidos) recogida en la base de datos.

No obstante, previamente al proceso de acoplamiento del Registro Civil con los padrones, fue necesario realizar un proceso de acoplamiento interno del Registro Civil, con el objetivo de asignar identidades a los sujetos protagonistas de los nacimientos, los matrimonios y las defunciones.

#### 3.1. El proceso de acoplamiento interno del Registro Civil. La creación de la Tabla de Relaciones.

El proceso de acoplamiento interno del Registro Civil, parte de la identificación de los sujetos (Egos) protagonistas de los nacimientos, matrimonios y defunciones inscritos en Aranjuez, así como del tipo de relación que el individuo mantiene con otros sujetos que figuran en las partidas, como son sus padres, abuelos y cónyuge/s.

En este paso, la identificación de un individuo a través de su información nominativa, junto con los datos nominales disponibles de sus padres y/o esposo/a, permitía establecer un vínculo entre los diferentes sucesos vitales y corroborar que se trataba de un mismo individuo<sup>10</sup>.

A partir de estos resultados se creó la Tabla de Relaciones (Figura 1), que integra la información relativa a los diferentes sucesos vitales en los que el sujeto participa, éstos son: nacimientos, matrimonios y defunciones, y con la que es posible reconstruir la familia o familias de un individuo a lo largo del período que abarca la investigación. Con este objetivo se ha asignado a cada individuo un código unívoco (NIP) que permita localizarle en cualquier acontecimiento vital, tanto en los que aparece como protagonista (ID), como en los que aparece en di-

ferentes posiciones con respecto a otros individuos de su familia.

Figura 1: La Tabla de Relaciones.

	NIP	sex	idnac	idmat	iddef	hijpad	hijmad	nfraphij	nframhij	nfrappa	nfrapma	nframpa	nframma	conyug
	1	1	6	2832	26667	60862	74397	6	6	11601	11601	2707		2702
	2	1	8		1075	56058	56202	8	8					
	3	1	7		41	59875	60408	7	7					
	4	2	9	1511	8738	60534	56259	9	9				7761	41695
	5	1	10		421	60892	61148	10	10					
	6	2	11											
	7	2	12			58102	74290	12	12					
	8	1	13	1788	18221	58590	74310	13	13	11643	11643	11643		48554
	9	1	15		504	59880	55208	15	15					
	10	2	14		714	59106	61649	14	14					
	11	1	16		6752	55944	53087	16	16					
	12	1	17			68975	74255	17	17					
	13	1	18			68976	59976	18	18					
	14	1	19			68977	74284	19	19					
	15	2	21			68978	74387	21	21					
	16	1	20		30	53740	74291	20	20					
	17	2	22			40752	55258	22	22					
	18	1	23			41365	74391	23	23					
	19	1	24		11965	41355	55861	24	24					
	20	1	25			68979	74408	25	25					
	21	2	26		233	54954	74407	26	26					
	22	2	27	1655	19647	68980	60002	27	27			8681	8681	41825
	23	1	29	1645	23518	59463	74346	29	29	8744	8744	316		313
	24	2	28			68981	74399	28	28					
	25	2	30		43	68982	74343	30	30					

Nota ilustrativa: El individuo con NIP 1, que es un varón y que tiene registrado su nacimiento con el ID 6, su matrimonio con el ID 2.832 y su defunción con en el ID 26.667, es hijo del varón 60.862 y de la mujer 74.397. Figura como hijo en la fratría paterna y materna 6, y como padre, en la fratría paterna 11.601 y en la materna 2.702, ya que se casa y forma una familia con la mujer 2.702.

Tabla 5: Información Recogida en la Tabla de Relaciones

NIP: Código de identificación del individuo (ego)
Sexo del individuo (ego)
IDNac: Número de la ficha de nacimiento en que aparece el individuo (ego)
Año del nacimiento del individuo (ego)
IDMat: Número de la ficha del primer/segundo/tercer/cuarto matrimonio en que aparece el individuo (ego)
Año del primer/segundo/tercer/cuarto matrimonio del individuo (ego)
IDDef: Número de la ficha de la defunción en que aparece el individuo (ego)
Año de la defunción del individuo (ego)
Código del padre del individuo (ego)
Código de la madre del individuo (ego)
Número de la fratria paterna en la que el individuo (ego) aparece como hijo
Número de la fratria materna en la que el individuo (ego) aparece como hijo
Número de la fratria paterna en la que el individuo (ego) aparece como padre
Número de la fratria paterna en la que el individuo (ego) aparece como madre
Número de la fratria materna 1 en la que el individuo (ego) aparece como padre
Código del primer/segundo/tercer/cuarto conyuge del individuo (ego)

Como se puede advertir en la Figura, la información que contiene la Tabla de Relaciones es muy extensa<sup>11</sup> y da cuenta de todos los sucesos vitales en los que los individuos inscritos en el Registro Civil participan. En la Tabla 5 se explicita toda la información que recoge la Tabla de Relaciones.

La Tabla de Relaciones ha permitido reconstruir las familias de los diferentes individuos de la base de datos, así como la creación de genealogías<sup>12</sup> u otras bases de datos derivadas. Por el momento se ha creado la llamada Tabla de Familias (Figura 2), que reúne la información de las diferentes familias que un sujeto puede formar a lo largo de su vida, designadas mediante un código numérico (NIPFam).

Figura 2: La Tabla de las Familias

NIPFam4	NIPFam3	NIPFam2	NIPFam1	NIPFamOr	NIP	sex	idnac	idmat1	idmat2	iddef	hijpad	hijmad	nfraphij	nfre
		2295	2294	2293	41146	1		892	1044					
			2585	15508	41148	1		893		18273				
				15529	41170	1		916						
				15500	41137	1		883						
			2516	15556	41200	1		947						
			2594	15527	41168	1		914		17934				
				15549	41191	1		938		5803				
			2893	15550	41192	1		939						
			2571	15551	41193	1		940		7108				
				15552	41194	1		941						
				15553	41195	1		943						
			2638	2637	41197	1		944	1365	14324				
				15547	41189	1		936						
			2623	15555	41199	1		946		21175				
			2660	15546	41188	1		935		6313				
			2790	15557	41201	1		948		13889				
			2688	15558	41202	1		949		14049				
			3152	15559	41203	1		950						
			3030	15560	41204	1		951	1257					
			2927	15561	41205	1		952	1150	15047				
			2317	15562	41206	1		953						
			2669	15563	41207	1		954		8218				
			2797	15636	41288	1		1042		17624				
			2998	15554	41198	1		945						
				15538	41179	1		926						

Nota ilustrativa: El individuo con NIP 41.146 que es un varón, perteneciente a la familia 2.293, tiene dos matrimonios: el 892, por el que forma la familia 2.294 y, el 1.044, por el que forma la familia 2.295.

Debido a la complejidad de las formas de organización familiar que ha permitido reconstruir la base de datos, ha sido necesario que las familias cuenten con varios códigos identificadores, dependiendo del número de familias que un individuo forme a lo largo de su vida en Aranjuez. Así tenemos un NIPFamOr cuando se trata de la familia de origen del individuo, NIPFam1 que hace referencia a la primera familia de formación del sujeto, NIPFam2 a la segunda familia de formación y, NIPFam3 y

NIPFam4 cuando se trata de la tercera y cuarta familia de formación respectivamente.

Con el objetivo de integrar en la *Tabla de Relaciones* la información disponible de los sujetos del Padrón, se ha llevado a cabo un proceso de acoplamiento entre el Registro Civil y los distintos padrones, que se detalla a continuación.

### 3.2. El proceso de acoplamiento entre el Registro Civil y los padrones.

Si bien en el caso del acoplamiento interno del Registro Civil, la vinculación entre los diferentes sucesos vitales en los que un individuo participaba partía de la identificación del Ego a través de su información nominativa y de la disponible de sus padres y/o esposo/a, en el caso del acoplamiento del Registro con el Padrón Municipal de Habitantes no sucede de la misma forma. El reto que plantea la nueva vinculación, es la dispar estructura de estas fuentes.

Como se ha dicho anteriormente, el Registro Civil contiene tanto información de los sujetos protagonistas como de sus familias, en los padrones, por su parte, la única información disponible como tal es aquella perteneciente al individuo como sujeto, es decir, sólo contamos con la información nominal (nombres y apellidos) del sujeto aspirante a una identidad.

Con la finalidad de que el acoplamiento de la información del Registro Civil con la de los Padrones aporte unos resultados los más fiables posibles, se han llevado a la práctica estrategias que permitan solventar la falta de información de los padres y/o esposo/a en los padrones. Estas estrategias harían hincapié en la información relativa a la edad del individuo, al sexo y a la familia a la que pertenece.

En lo que respecta a la edad de los sujetos, se ha calculado la diferencia entre la edad del sujeto en el padrón correspondiente y la del Registro Civil. Para ello ha sido necesario calcular la edad del sujeto en el Registro Civil, debido a que esta variable no estaba disponible. De este modo, se ha tomado como referencia la fecha del Padrón en cuestión a 1 de julio, y se ha calculado la edad que el sujeto del Registro Civil tendría a 1 de julio del año del Padrón que en ese momento nos ocupa.

En cuanto a la información concerniente al sexo, parece que la condición de que el sexo sea igual es

obvia, pero debido a la calidad de las fuentes, en muchos casos la demografía histórica se plantea la dificultad de que el sexo de los sujetos está omitido o bien es erróneo. En el caso de la investigación de Aranjuez, la información relativa a finales del siglo XIX y los primeros años del siglo XX es la que presenta mayor número de inconvenientes en lo que al sexo se refiere, debido sobre todo a confusiones relacionadas con la asignación del sexo en las fuentes originales. Estos problemas han sido solventados mediante una revisión manual de los casos erróneos teniendo en cuenta el resto de información incluida, como la profesión, entre otros datos.

En cuanto a la información disponible de la familia, en los padrones no existe dicha información como tal, pero si disponemos de la información de los hogares. El Padrón Municipal de Habitantes viene estructurado por hogares, aunque si bien en todos los casos dichos hogares no se corresponden con una familia<sup>13</sup>, si podemos identificar hogar-familia en buena parte de los casos. De esta forma, y teniendo en cuenta que la organización de la información del hogar parte del cabeza de familia, que normalmente se trata de un varón, y se estructura mediante relaciones de parentesco con respecto a éste (esposa, hijos, padres y sirvientes, entre otros), hemos podido reconstruir un buen número de familias en los padrones.

Otra de las tácticas destinada a facilitar la identificación de los individuos en la base de datos, ha sido la conversión fonética de los campos nominales. Si bien el proceso de acoplamiento parte de la correspondencia del nombre y apellidos de los individuos, en todos los casos dicha correspondencia no es exacta, debido a errores ortográficos, omisión de nombres y/o apellidos compuestos y a confusiones derivadas de similitudes fonéticas. Por ello, la base de datos ha sido explotada a partir de un sistema de homologación y codificación fonética que se explicita en la Tabla 6<sup>14</sup>. Este procedimiento reduce el número de errores y permite establecer el mayor número de vínculos posibles entre los individuos de la base de datos<sup>15</sup>.

Tabla 6: Conversión Fonética de Nombres y Apellidos de la Base de Datos

Forma Original	Conversión Fonética	Forma Original	Conversión Fonética
ca	ka	i	y (si i va entre vocales)
ce	ze	i+vocal	y (siendo i la vocal inicial)
ch	Ç	ll	y (si ll va seguida de vocal)
ci	zi	ll	l (si no va seguida de vocal)
co	ko	ph	f
cu	ku	que	ke
enrr	enr	qui	ki
ge	je	rr	r (cuando rr es inicial)
gi	ji	sce	ze
gua	wa	sci	zi
h (muda)	(se elimina, excepto si está intercalada)	ss	s
hua	wa	tt	t
hue	we	v	b
hui	wi	y+consonante	i

Fuente: Reher, González Quiñones y Sanz Gimeno, 2001.

A partir de estos pasos necesarios, se ha abordado el acoplamiento entre el Registro Civil y los diferentes Padrones elaborados en Aranjuez durante el período de estudio. Dicho proceso de acoplamiento tiene dos niveles de detalle, el acoplamiento automático y el acoplamiento manual.

### 3.2.1. El proceso de acoplamiento automático.

En el acoplamiento automático la identificación de un individuo parte de la correspondencia exacta entre nombre, apellido 1 y apellido 2 del sujeto en el Registro Civil y en el Padrón correspondiente.

Debido al volumen de individuos insertos en la base de datos y a la proliferación de nombres y apellidos comunes, cabe la posibilidad de que dicha equivalencia exacta entre los campos nominales, nos aporte más de un aspirante para una única

identidad<sup>16</sup>, en este caso, el acoplamiento realizado no se consideraría automático, sino que sería revisado manualmente para elegir entre los candidatos el más óptimo mediante la visualización de su información en un formulario de Access.

Con el objetivo de que los resultados aportados por el acoplamiento automático sean fiables, dicho acoplamiento debe cumplir dos requisitos: que los sujetos vinculados tengan el mismo sexo y, que la diferencia de edad entre el individuo vinculado del Registro Civil con el del Padrón esté comprendida entre los más / menos 10 años de edad.

Teniendo en cuenta estos criterios los resultados obtenidos en el proceso de acoplamiento automático han sido muy satisfactorios, como se puede observar en la Tabla 7. En torno al 42% de los individuos de cada Padrón han sido localizados automáticamente en el Registro Civil, vinculando tan sólo la información nominativa disponible para cada sujeto.

Tabla 7: Resultados del Proceso de Acoplamiento Automático

Padrones	Registros	Vínculos	Porcentaje
1877	7657	3101	40,50
1903	9543	4167	43,67
1915	10019	4416	44,08
1945	20475	8825	43,10
1960	27197	11942	43,91
1975	31004	10838	34,96



Si bien los resultados obtenidos en el proceso de acoplamiento automático han demostrado la calidad de la base de datos construida y las posibilidades de la conversión fonética a la hora de identificar individuos, el proceso de acoplamiento manual se convierte en un reto, debido a la limitada información disponible para los diferentes sujetos que componen la base de datos del Registro Civil y de los Padrones.

### 3.2.2. El proceso de acoplamiento manual.

En el acoplamiento manual la identificación de un individuo parte de la correspondencia entre los campos nominales del Registro Civil y del Padrón oportuno, pero dicha correspondencia no es exacta como en el caso de los acoplamientos automáticos, sino que tan sólo coinciden dos de los campos nominales, siendo las combinaciones posibles:

- § nombre y primer apellido coincidentes.
- § nombre y segundo apellido coincidentes.
- § primer y segundo apellido coincidentes.

El acoplamiento resultante daría lugar a varios aspirantes para cada identidad y manualmente habría que escoger al mejor candidato entre los posibles, teniendo en cuenta los requisitos de sexo igual y diferencia de edad de más / menos 5 años.

La tabla derivada del acoplamiento entre el Registro Civil y el Padrón en cuestión mediante dos campos nominales, denominada 2 Vínculos, se insertaría en un formulario diseñado en Access. La visualización de la información de la tabla 2 Vínculos que se muestra en la Figura 3 se realiza en dos niveles, en la parte superior se recoge la información relativa al Registro Civil y en la inferior la del Padrón.

Figura 3. Formulario de Validación del Proceso de Acoplamiento entre el Registro Civil y el Padrón Municipal de Habitantes.

The screenshot shows a Microsoft Access form titled "formulario repe" with a menu bar (Archivo, Edición, Ver, Insertar, Formato, Registros, Herramientas, Ventana) and a toolbar. The form is divided into two main sections. The top section, titled "Miembros de la Familia del Registro Civil", contains fields for "NIP" (187), "NIPFam Orig" (897), "NIPFam1" (3963), "NIPFam2", and "NIPFam3". Below these is a list of family members with a dropdown menu showing "guillermo lanzada arellano", "luciana manuela lanzada arellano", and "antonio lanzada arellano". Below the list are fields for "Sexo" (1) and "Edad" (32), and three text boxes containing "francisco feliz", "lanzada", and "arellano". The bottom section, titled "IDPad903" (7538), contains a list of names: "felix lanzada arellano", "josefa mendez luengo", "julia lanzada mendez", "jose lanzada mendez", and "julio lanzada mendez". Below the list are fields for "Sexo" (1), "Edad" (31,8), and "DifEdad" (0), and three text boxes containing "felix", "lanzada", and "arellano". A large blue "V" button is visible on the right. The bottom status bar shows "Registro: 1 de 1" and "Registro: 18 de 4324".

Como se puede observar en la Figura 3, a pesar de que el proceso de acoplamiento se realiza con los campos nominales codificados, con el propósito de facilitar la labor de reconocimiento de los sujetos en una y otra fuente estadística, la información incluida en el formulario aparece decodificada con los nombres y apellidos originales.

Otra de las estrategias llevada a cabo para facilitar la búsqueda de las identidades correctas en una y otra fuente, y descartar los candidatos erróneos, es incluir los datos de la familia de los individuos vinculados. De esta forma, la visualización en el formulario de Access incluiría tanto la información nominal de la familia del sujeto en el Padrón, reconstruida a partir de los datos de parentesco de los diferentes hogares que integran los padrones, así como aquella referente a las distintas familias que los individuos pueden llegar a formar a lo largo de su vida y que han quedado registradas en la Tabla de Relaciones confeccionada a partir de la información del Registro Civil.

De esta forma la decisión puede tomarse de forma rápida y directa, indicando la validez del acoplamiento y, por tanto, la correspondencia de identidades, mediante la asignación de una “V” directamente en el formulario.

### 3.3. Resultados del proceso de acoplamiento entre el Registro Civil y los Padrones.

En la actualidad, el equipo de investigación está ultimando el proceso de validación del acoplamiento manual llevado a cabo entre el Registro Civil y los diferentes padrones comprendidos en el período de estudio.

Pese a que tan sólo se ha completado la validación del Registro Civil con los padrones de 1877, 1903 y 1915, los datos arrojan resultados muy satisfactorios. Como se puede observar en la Tabla 8, una media del 35% de los individuos empadronados en Aranjuez en estos años, han podido ser identificados en el Registro Civil a partir

del acoplamiento de dos de sus campos nominales.

Tabla 8: Resultados del Proceso de Acoplamiento Manual

Padrones	Registros	Vínculos	Porcentaje
1877	7657	2138	27,92
1903	9543	3780	39,61
1915	10019	3707	37,00

Teniendo en cuenta estos resultados y los obtenidos en el proceso de acoplamiento automático del Registro Civil con los diferentes padrones, el resultado final es muy positivo. En la Tabla 9 se presentan los resultados derivados de los procesos de acoplamiento automático y manual efectuados.

Tabla 9: Resultados del Proceso de Acoplamiento Automático y Manual

Padrones	Registros	Vínculos	Porcentaje
1877	7657	2138	27,92
1903	9543	3780	39,61
1915	10019	3707	37,00

Como se puede advertir en la Tabla, aproximadamente el 78% de los sujetos empadronados en Aranjuez a finales del siglo XIX y principios del siglo XX, han podido ser identificados en alguno/s de los sucesos vitales recogidos en el Registro Civil, ya sea en su nacimiento, matrimonio y/o defunción. Cabe esperar que los resultados del acoplamiento de los padrones de 1945, 1960 y 1975 con el Registro Civil, sean igualmente favorables, obteniendo porcentajes similares.

## 4. Perspectivas de Análisis

Los óptimos resultados que se están obteniendo en la investigación de Aranjuez, muestran las múltiples posibilidades que la vinculación de diversas fuentes a partir de datos nominales, tienen para la reconstrucción de familias y trayectorias de vida.

La relevancia de la voluminosa base de datos de Aranjuez elaborada mediante los diferentes procesos de acoplamiento que se han acometido, es la primera base de datos a nivel mundial en pro-

porcionar información acerca de los comportamientos demográficos de los individuos durante la transición demográfica.

Las posibilidades de investigación a partir de la base de datos son por tanto múltiples: permite evaluar el calendario y la intensidad de la caída de la fecundidad según las características socioeconómicas y culturales de las familias; analizar el descenso de la mortalidad infantil y general du-

rante la transición epidemiológica; observar la incidencia de ciertas enfermedades en determinadas familias; estudiar la relación existente entre la caída de la mortalidad y la fecundidad durante el período de la transición demográfica y la difusión espacial de los comportamientos demográficos; entre otras muchas perspectivas de análisis.

Estas y otras aplicaciones convierten la base de datos de Aranjuez en una valiosa fuente de estudio del período de la transición demográfica.

### Bibliografía

- AMORIM, Maria Norberta, 2001, "Micro-análise da transição demográfica numa parroquia açoriana", *Revista de Demografia Histórica*, XIX, II, ADEH, Zaragoza, Pp. 79-113.
- AMORIM, Maria Norberta, 1995, "Informatização normalizada de arquivos. Reconstituição de parroquias e historia das populações", *Boletín de la Asociación de Demografía Histórica*, XIII, II, ADEH, Zaragoza, Pp. 141-150.
- ARANGO, Joaquín, 1985, "La teoría de la transición demográfica y la experiencia histórica", *Revista Española de Investigaciones Sociológicas*, nº 10, Pp. 169-198.
- ARDIT, Manuel, 1999, "Microanálisis demográfico en larga duración. El caso de España", en González Portilla y Zárraga Sangroniz (eds.), *IV Congreso de la Asociación de Demografía Histórica. Demografía Urbana, migración y envejecimiento*, Lejona, País Vasco, Pp. 253-308.
- BERNABEU-MESTRE, José, RAMIRO FARIÑAS, Diego, SANZ GIMENO, Alberto, ROBLES GONZÁLEZ, Elena, 2003, "El análisis histórico de la mortalidad por causas. Problemas y soluciones", *Revista de Demografía Histórica*, XIX-I, ADEH, Pp. 167-192.
- CARRERAS, Albert (coordinador), 1989, *Estadísticas históricas de España. Siglos XIX-XX*, Fundación Banco Exterior, Colección Investigaciones, Madrid.
- COALE, Ansley J. and WATKINS, Susan Cotts, eds., 1986, *The decline of fertility in Europe*, Princeton University Press. Princeton.
- DíEZ NICOLÁS, Juan, 1971, "La transición demográfica en España", *Revista de Estudios Sociales*, nº 1, Pp. 89-158.
- DOPICO, Fausto, 1998, "El descenso de la mortalidad en España y en las regiones históricas (1860-1930)" en Dopico y Reher, *El declive de la mortalidad en España, 1860-1930*, Monografía nº 1, Asociación de Demografía Histórica, Huesca.
- FLEURY, Michael y HENRY, Louis, 1976, *Nouveau manuel de dépouillement et d'exploitation de l'état civil ancien*, Institut National d'Études Démographiques, París.
- INSTITUTO NACIONAL DE ESTADÍSTICA, varios años, *Censo de Población y Vivienda*, Instituto Nacional de Estadística, Madrid.
- INSTITUTO NACIONAL DE ESTADÍSTICA, varios años, *Movimiento Natural de la Población*, Instituto Nacional de Estadística, Madrid.
- INSTITUTO NACIONAL DE ESTADÍSTICA, varios años, *Padrón Municipal de Habitantes*, Instituto Nacional de Estadística, Madrid.
- IRISO NAPAL, Pedro Luis y REHER, David, 1987, "La fecundidad y sus determinantes en España, 1887-

- 1920”, *Revista Española de Investigaciones Sociológicas*, 39, Pp. 45-118.
- LIVI BACCI, Massimo, 1999a, *Historia de la población europea*, Crítica, Barcelona.
  - LIVI BACCI, Massimo, 1999b, *Historia mínima de la población mundial*, Ariel, Barcelona.
  - LIVI BACCI, Massimo, 1985, “Cambios de la fecundidad y la nupcialidad en España del siglo XVIII al XX”, en Espina, Fina y Lorente (compiladores), *Estudios de economía del trabajo en España*, Ministerio de Trabajo y Seguridad Social, Madrid.
  - MCKEOWN, Thomas, 1978, *El crecimiento moderno de la población*, Antoni Bosch, Barcelona.
  - NADAL, Jordi, 1966, *La población española (Siglos XVI a XX)*, Ariel, Barcelona.
  - OMRAM, A. R., 1971, “The epidemiologic transition: a theory of the epidemiology of population change”, *Milbank Memorial Fund Quarterly*, 49, Pp. 509-583.
  - REHER, David S., GONZÁLEZ QUIÑONES, Fernando y SANZ GIMENO, Alberto, 2001, “Procesos de modernización y trayectorias de vida. Propuesta para el análisis sociodemográfico a partir de datos locales en España”, *Pré-actas do VI Congresso da Associação de Demografia Histórica (ADEH)*, Instituto Politécnico de Castelo Branco, Castelo Branco (Portugal).
  - REHER, David S., 2000, “La investigación en demografía histórica: pasado, presente y futuro”, *Boletín de la Asociación de Demografía Histórica*, XVIII-II, ADEH, Pp. 15-78.
  - REHER, David S., 1999, “Back to the basics: mortality and fertility interactions during the demographic transition”, *Continuity and change*, 14: 1, Pp. 9-31.
  - REHER, David S., 1996, *La familia en España. Pasado y presente*, Alianza Universidad, Madrid.
  - REHER, David S. y VALERO LOBO, Ángeles, 1995a, *Fuentes de Información Demográfica en España*, Centro de Investigaciones Demográficas, Madrid.
  - REHER, David S., 1995b, “Wasted investments: some economic implication of childhood mortality patterns”, *Population Studies*, 49, Pp. 519-536.
  - REHER, David-Sven, POMBO, María Nieves y NOGUERAS, Beatriz, 1993, *España a la luz del Censo de 1887*, Instituto Nacional de Estadística, Madrid.
  - SANZ GIMENO, Alberto y GÓNZALEZ QUIÑONES, Fernando, 2001, “Las mujeres y el control de la fecundidad. Propuesta metodológica para su identificación durante la transición demográfica”, *Revista de Demografía Histórica*, XIX, II, ADEH, Zaragoza, Pp. 57-78.
  - SANZ GIMENO, Alberto, 2000, “Aranjuez, 1870-1970. Cien años de historia demográfica”, *Aranjuez. Studia*, nº 3, Ilustrísimo Ayuntamiento del Real Sitio y Villa de Aranjuez, Delegación de Servicios Culturales, Madrid.
  - SANZ GIMENO, Alberto, 1999, *La mortalidad de la infancia en Madrid*, Comunidad de Madrid, Consejería de Sanidad y Servicios Sociales, Madrid.

## Notas de Rodapé

\* La autora quiere expresar su agradecimiento a David S. Reher, Fernando González Quiñones y Alberto Sanz Gimeno, así como a todos los becarios y colaboradores que han participado en la creación de la base de datos de Aranjuez.

<sup>1</sup> Aunque a lo largo de los siglos XIX y XX la población de Aranjuez crece significativamente, dicho crecimiento no es sostenido, ya que se observan fluctuaciones substanciales en la tasa de crecimiento de la población. De este modo, entre 1787 y 1848 el crecimiento intercensal de la población era de un 0,5%, entre 1930 y 1940 aumentó hasta llegar a un 4,5% y, en el período comprendido entre 1981 y 1991 se produce un importante descenso, debido a la implantación de comportamientos demográficos modernos que hacen que esta cifra se situara en un 0,2% (Sanz Gimeno, 2000: 13).

<sup>2</sup> Si bien el Registro Civil y el Padrón constituyen el cuerpo principal de la base de datos elaborada, no es despreciable la utilidad del registro de las medidas antropométricas de los quintos, y de renta y patrimonio de otras fuentes anexas. La información facilitada por dichas fuentes además de enriquecer la base de datos, ha permitido contrastar y complementar datos erróneos u omitidos en las fuentes principales, como son por ejemplo el oficio o la alfabetización, entre otros.

<sup>3</sup> Se ha podido tener acceso a la información proporcionada por el Registro Civil, gracias a una autorización judicial y con el compromiso de no utilizar información nominativa alguna en la presentación y publicación de los datos.

<sup>4</sup> Las partidas que componen la información del Registro Civil de Aranjuez abarcan de enero de 1871 a abril de 1971 en el caso de los nacimientos, de enero del año 1871 a diciembre de 1970 en el de los matrimonios, y de enero de 1871 a noviembre de 1971 en el de las partidas de defunción.

<sup>5</sup> Si bien en buena parte de los casos, la información es completa tanto para los campos nominativos como para el resto de variables que comprenden las partidas, en algunos períodos la información de ciertas variables aparece omitida, como por ejemplo, la relacionada con los oficios y las localidades de origen de los padres y abuelos.

<sup>6</sup> Para mayor información acerca del tratamiento de las causas de defunción ver Bernabeu-Mestre, Ramiro Fariñas, Sanz Gimeno y Robles González (2003).

<sup>7</sup> El Padrón Municipal de Habitantes permite localizar tanto los individuos que comparten un domicilio, como los que coinciden en una determinada ubicación geográfica dentro de la localidad, como podrían ser por ejemplo, el barrio.

<sup>8</sup> El método tradicional que se utiliza para reconstruir familias parte de la identificación de las familias, para posteriormente ir incorporando los diferentes miembros que las componen.

<sup>9</sup> Los avances que Norberta Amorim y el Nucleo de Estudos de População e Sociedades de la Universidade do Minho vienen desarrollando para reconstruir parroquias mediante la creación de fichas biográficas individuales, han sido una fuente de inspiración relevante para la metodología de reconstrucción de familias del proyecto de Aranjuez.

<sup>10</sup> Para una información más detallada acerca del proceso de acoplamiento interno del Registro Civil, véase Reher, González Quiñones y Sanz Gimeno (2001).

<sup>11</sup> La Tabla de Relaciones creada reúne la información relativa a los 68.462 individuos que integran la base de datos de Aranjuez.

<sup>12</sup> El objetivo que se plantea con la construcción de genealogías es poder relacionar un individuo de la base de datos con los miembros de su grupo familiar ascendente, descendente y lateral. Las genealogías nos facilitarían por tanto, el análisis de los sujetos en función de las características de los miembros de su grupo de parentesco.

<sup>13</sup> Los miembros de conventos, monasterios, internados, asilos, cárceles, entre otra población institucional, son contemplados en el Padrón como hogares colectivos, si bien no existe relación de parentesco entre sus miembros. Otro de los problemas a este respecto es el caso de los duplicados, esto es, personas que tienen su hogar propio en Aranjuez y al mismo tiempo desempeñan un trabajo en otro hogar. Los sirvientes serían un ejemplo de esta situación, ya que en ocasiones figuran empadronados en el hogar propio, otras en el hogar de la familia con la que trabajan y en la mayor parte de las veces, en ambos hogares.

<sup>14</sup> El proceso de acoplamiento interno del Registro Civil también se llevó a cabo partir del tratamiento fonético de los campos nominales.

<sup>15</sup> Para ampliar información acerca del proceso de codificación seguido, consultar Reher, González Quiñones y Sanz Gimeno (2001)

<sup>16</sup> Existen en la base de datos individuos con información nominal idéntica y sin embargo se trata de sujetos diferentes. Este es el caso de los hermanos de aquellos niños que fallecían en los primeros días, meses o año de vida y que adoptaban por tradición el nombre del hermano difunto. Otro problema a este respecto es el de matrimonios duplicados tras la Guerra Civil española. Durante los primeros años de la posguerra numerosos matrimonios que se habían contraído durante la II República, volvieron a inscribirse en el Registro Civil debido a que el nuevo régimen les consideró ilegales. Estos “nuevos matrimonios” han sido eliminados de la base de datos, utilizando a efectos prácticos los inscritos durante la II República.

## Proyecto IPUMS, Integrated Public Use of Microdata Series: aplicabilidad a un estudio sobre la Emigración Española a Florida entre 1880 y 1920<sup>1</sup>

Albert Esteve Palós  
 Angels Tonents  
 Clara Cortina

### Resumo

Las constantes innovaciones tecnológicas producidas en campos como la obtención, procesamiento y difusión de datos abren un sinfín de oportunidades a la Demografía Histórica. El proyecto Integrated Public Use Microdata Series (IPUMS-USA), creado en la Universidad de Minnesota en octubre de 1997, hace un uso extensivo e intensivo de las nuevas tecnologías para poner a disposición gratuita de los investigadores muestras armonizadas de microdatos censales a través de internet ([www.ipums.org](http://www.ipums.org)). Las 25 muestras que actualmente son distribuidas para Estados Unidos proceden de 13 censos, publicados entre 1850 y 1990, y conforman la más substanciosa colección de datos para analizar las transformaciones demográficas, sociales y económicas de este país. Ante la gran variedad de formatos, variables y códigos empleados por los distintos censos, IPUMS

estandariza los formatos, armoniza los códigos, construye variables y, finalmente, elabora la documentación necesaria para facilitar el análisis de la información a los usuarios. A guisa de ejemplo, el estudio sobre la emigración española a Tampa, Florida, entre 1880 y 1920 realizado con datos de IPUMS (Esteve, A., Torrents, À., Cortina, C., 2003) ilustrará el proceso de solicitud, extracción y aplicabilidad de los mismos. Gracias a IPUMS, las características individuales y familiares de los españoles y sus descendientes directos en Florida han podido ser analizadas sin necesidad de acudir a los cuestionarios originales, tarea que hubiera implicado una significativa inversión de recursos y tiempo no al alcance de todas las investigaciones.

Palabras clave: microdatos censales, emigración, población extranjera, nupcialidad.

### Abstract

Historical census microdata are invaluable resource for historical demography research. Most of these data, however, remains unavailable to researchers. In the case of the United States, the Integrated Public Use of Microdata Series (IPUMS-USA), developed at the Minnesota Population Center, makes census microdata freely available to scholars in harmonized format with comprehensive documentation through a state-of-the-art and user friendly data access system. IPUMS-USA is now distributing 25 samples from 13 population censuses that cover a period of 150 years, from 1850 to 2000.

The study of the Spanish community in Florida between 1880 and 1920 based on IPUMS data will illustrate the overall process of obtaining census microdata and its applicability. Thanks to IPUMS, individual and family characteristics of the Spanish born and its descendants in Florida could be easily studied without being necessary to go to the original census forms, what, certainly, had been very costly.

Keywords: census microdata, migration, foreign population, marriage.

Les microdonnées des recensements historiques sont une inestimable ressource pour la recherche en matière de démographie historique. Et cependant la grand majorité de ces données sont inaccessibles aux chercheurs. Le projet Integrated Public Use of Microdata Series (IPUMS-USA) développé au Minnesota Population Center a fait des États-Unis une exception. Leurs microdonnées

ont été rendues gratuitement accessibles aux chercheurs moyennant un système facile et agile qui offre les données sous une forme già harmonisé et accompagnées de toute la documentation nécessaire. IPUMS-USA distribue à présent 25 échantillons appartenant a 13 populations recensées pour une période de 150 ans, de 1850 jusqu'à l'an 2000.

L'étude de la communauté espagnole de la Floride entre 1880 et 1920 utilise des données de l'IPUMS et permettra d'illustrer le processus d'obtention des données et leurs possibilités d'application. Grâce à l'IPUMS l'étude des caractéristiques individuelles et familiales des espagnols et leurs descendance en Floride à été rendu beaucoup plus facile évitant la consultation directe des formulaires des recensements.

Paroles clés: microdonnées du recensement, population étrangère, migration, nuptialité.

### El proyecto IPUMS: Integrated Public Use of Microdata Series

Los microdatos censales son una fuente de valor incalculable para la investigación en ciencias sociales (McCaa y Ruggles 2002). A diferencia de las encuestas, los microdatos censales ofrecen mayor densidad muestral, profundidad histórica y cobertura territorial. En la mayoría de países, los microdatos no se encuentran disponibles o su acceso es restringido, razones que explican su escasa utilización. Estados Unidos y Canadá son una excepción al respecto. Los microdatos de estos países están disponibles desde la década de los sesenta, hasta convertirse hoy en un componente indispensable de la infraestructura en ciencias sociales.

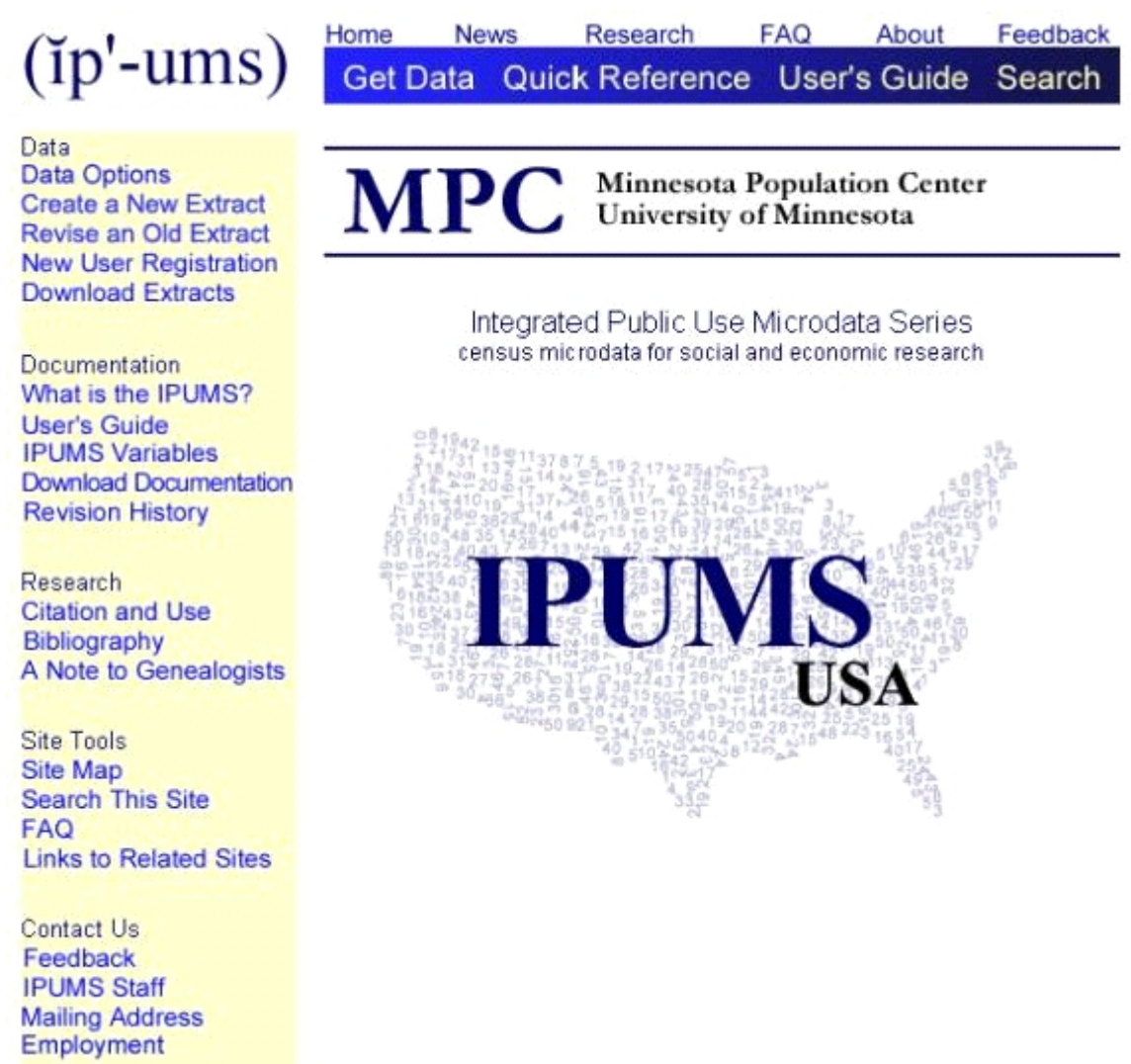
Las Series de Microdatos Censales Integrados de Uso Público (Integrated Public Use Microdata Series – USA) son parcialmente responsables del extendido uso de microdatos censales entre los demógrafos y otros investigadores interesados en el estudio de los Estados Unidos. IPUMS-USA, proyecto desarrollado por Ruggles, Sobek y otros (1997) en el Population Center de la Universidad de Minnesota, ha puesto a disposición gratuita de la comunidad científica series de microdatos armonizadas desde 1850 a 2000, debidamente documentadas, a través de un sistema sencillo y funcional de acceso a los datos. Desde su puesta en marcha preliminar en 1995, IPUMS ha sido una de las fuentes de datos demográficos más utilizadas en

el mundo entero. En 1998, por primera vez, se extendió el paradigma de IPUMS a los censos de Colombia, una experiencia piloto realizada en estrecha colaboración con el Departamento Nacional de Estadística de Colombia (DANE). Sin duda alguna, Col-IPUMS colocó la primera piedra de lo que más tarde sería una exitosa iniciativa de integración de microdatos a escala internacional: IPUMS-Internacional. En 1999, la agencia estadounidense National Science Foundation financió la propuesta IPUMS, dando vida a un proyecto que hoy, en su cuarto año, ha integrado con éxito datos de siete países (China, Colombia, Estados Unidos, Francia, Kenya, México y Vietnam), inventariado las muestras de microdatos en formato de cómputo existentes en el mundo para el periodo 1960-2000 (Hall, McCaa, y Thorvaldsen 2000), y preservado muestras de microdatos de más de cien censos. Además de estos logros, cabe señalar la buena acogida que entre los académicos está teniendo este proyecto.

Para los censos históricos de Estados Unidos, los publicados antes de 1930, IPUMS ha creado muestras de microdatos a partir de los cuestionarios censales originales, cuyo contenido estaba grabado en microfilms. La información de los cuestionarios seleccionados fue digitalizada por un equipo de personas que contaban con la ayuda de las más avanzadas tecnologías en este tipo de trabajos. El proceso de digitalización sigue aún día. IPUMS pretende ampliar la densidad muestral de la mayoría de sus censos históricos. En el caso del censo de 1880, ya está disponible el 100% de la población. Tradicionalmente, una de las principales limitaciones al uso de los microdatos está relacionada con el derecho a la confidencialidad de los mismos. Sin embargo, este es un problema menor cuando se trata de microdatos históricos, puesto que, en el caso de los Estados Unidos, transcurridos 72 años de su obtención, es posible proporcionar todo el detalle conceptual y geográfico contenido en los datos. Así pues, al ser comparadas con las muestras más recientes, las muestras históricas son más ricas en contenidos. En el contexto del North Atlantic Population Project ([www.nappdata.org](http://www.nappdata.org)), el

Population Center de la Universidad de Minnesota, junto con otras instituciones, pretende para el año 2005 difundir muestras de microdatos de censos históricos de países como Canadá, Gran Bretaña, Noruega o Islandia.

Imagen 1. Portada del proyecto IPUMS-Estados Unidos en Internet (www.ipums.org)



Principios: preservar, integrar y difundir

IPUMS es un proyecto cometido con tres grandes objetivos: la preservación, la integración y la difusión de microdatos censales. La consecución de estos objetivos requiere la utilización de distintos métodos, técnicas y habilidades.

Preservar.- IPUMS tiene entre sus objetivos principales la misión de inventariar los microdatos y documentos censales en el mundo que hayan sobrevivido hasta nuestros días. Este inventario constituye, por sí mismo, un extraordinario recurso para investigadores y archiveros de datos. En el caso particular de los censos históricos IPUMS ha creado sus propias muestras.



**Integrar (Homologar).** – La integración es el principal reto al que IPUMS debe enfrentarse, pero también su principal fortaleza, el hecho diferencial que singulariza a esta base de datos respecto a las demás. Durante la etapa de integración, los datos son procesados con cuatro finalidades distintas: i) garantizar la confidencialidad de los datos; ii) reformar, limpiar e imputar valores perdidos en la base de datos; iii) armonizar variables; iv) construir variables. De forma paralela al proceso de integración, IPUMS elabora documentación armonizada para cada una de las muestras y variables con el objetivo de proporcionar información relevante de todo el proceso al usuario final. Los desafíos y métodos asociados a estos objetivos serán tratados con mayor detalle en el próximo apartado.

**Difundir.** – Una plataforma de difusión eficiente es esencial para optimizar el uso de los microdatos integrados. IPUMS hace un uso extensivo e intensivo de las nuevas tecnologías para satisfacer tan importante objetivo. La difusión se realiza por internet, mediante un sistema que permite al usuario, una vez acreditado como  $ta^2$ , confeccionar su propia base de datos, escogiendo formatos, muestras, variables y casos específicos y accediendo a la documentación relacionada (Imagen 1).

#### Desafíos y oportunidades: métodos en la integración de microdatos

La realización de un proyecto como IPUMS supone armonizar datos procedentes de distintos institutos de estadística y de distintas décadas. Las muestras originales varían significativamente en cuanto a la calidad, formatos, conceptos tratados, y esquemas de clasificación de variables (Esteve y Sobek, 2003). IPUMS ha desarrollado una batería de técnicas para copar con tanta diversidad. A continuación, se exponen brevemente los procedimientos utilizados en cada una de las etapas:

a. Garantizar la confidencialidad de los datos. IPUMS emplea dos estrategias distintas para garantizar la confidencialidad de los datos: la firma

de un protocolo en el que el usuario, en el momento de registrarse, se compromete a preservar la confidencialidad de los mismos y la implementación de medidas de protección estadística. Estas medidas incluyen tareas como modificar aleatoriamente el orden de los hogares, combinar códigos que informan de características sensibles o que permitan identificar grupos de población muy pequeños, entre otras. Estas precauciones, sin embargo, no afectan a los microdatos históricos, puesto que es posible difundir todo el contenido censal a nivel individual.

b. Estandarizar formatos, limpiar e imputar valores perdidos de la base de datos. Los censos existen en una gran variedad de formatos que deben ser homogeneizados para su inserción final en IPUMS. Cada una de las muestras se adapta a un formato de tipo jerárquico basado en dos niveles: hogar y persona. El nivel de hogar contiene también toda la información relativa a la vivienda. Además, IPUMS realiza las comprobaciones necesarias para validar la consistencia interna de la base de datos e imputar, cuando es posible, los valores perdidos, tarea especialmente complicada cuando se trata de datos históricos, cuya calidad es, generalmente, inferior a la de los datos recientes.

c. Armonizar variables. La armonización de variables ocupa un lugar central en el proceso de integración. El objetivo que se persigue en esta fase es la creación de variables coherentes en el tiempo y el espacio que permitan la comparación entre países y en perspectiva histórica. La armonización de variables requiere trabajar con todas las bases de datos simultáneamente. Las variaciones en el fraseado de las preguntas, las clasificaciones empleadas por cada censo, y los significados culturales de los conceptos censales desafían el objetivo de la armonización. Por estas razones, IPUMS identifica y documenta todas las incompatibilidades. En el proceso de armonización intervienen tres componentes: recopilar la documentación existente, determinar la disponibilidad de variables, y diseñar un sistema de códigos armonizado. Ante cualquier variable, su codificación

final debe satisfacer dos requisitos: garantizar la máxima comparabilidad en el tiempo y en el espacio y, a la vez, retener todo el detalle contenido en las variables originales. Para ello, IPUMS utiliza un sistema de codificación compuesto de varios dígitos, todos ellos armonizados.

d. Construir variables. IPUMS construye nuevas variables para maximizar la utilización de los datos ya armonizados. IPUMS crea variables a nivel individual que informan de las relaciones entre miembros de la familia, facilitando así la tarea a los investigadores que quieran identificar hogares con determinada estructura. Para cada individuo, se construyen tres variables que identifican la posición de la madre, el padre y la pareja (en matrimonio o unión libre) dentro del hogar. Otro grupo de variables describe las características de la familia y el hogar a nivel individual y del hogar.

Los microdatos históricos de los censos de los Estados Unidos: una aplicación al estudio de la emigración española a Florida, 1880 - 1920

La segunda mitad del siglo XIX y las primeras décadas del XX fueron el escenario de intensas migraciones transatlánticas. Los europeos, primero los del norte y oeste (los “viejos inmigrantes”) y más tarde los del sur, entre ellos los españoles, centro y este del continente (los “nuevos inmigrantes”), se desplazaron de forma masiva hacia América. España es un país de emigración tardía concentrada en las primeras décadas del siglo XX, muy especialmente entre 1900 y 1913, y con una procedencia regional clara: Galicia, Asturias, Cantabria y el País Vasco como principales regiones emigratorias. En lo que se refiere a las destinaciones, los españoles que emigraron hacia América se dirigieron principalmente al Río de la Plata (Argentina y Uruguay), Brasil y Cuba, seguidas en un segundo término por México y Puerto Rico (Sánchez-Albornoz 1988). La emigración española a América ha sido abordada en su conjunto (Sánchez-Albornoz 1988, Sánchez Alonso 1995 y Yáñez 1988) o parcialmente (Klein 1992, Sánchez Alonso 1992 y Maluquer 1996).

Entre el conjunto de destinaciones americanas, la emigración a Estados Unidos ha despertado un interés menor (Rueda 1993, Carbó 1990b y Fernández Shaw 1987), por ser una destinación poco frecuente. Esto no significa, sin embargo, que no existieran en Estados Unidos asentamientos en los que la población de origen español tuvo un rol destacado, es el caso de Tampa, Florida.

Este desplazamiento de millones de españoles hacia América, que Sánchez-Albornoz (1988) denomina la emigración en masa, fue registrado por el Instituto Geográfico y Estadístico (IGE) desde 1882 a partir del movimiento anual de entradas y salidas de pasajeros de los puertos españoles. Las series obtenidas gracias a esta fuente cifran en alrededor de 3 millones los españoles que dejaron el país con destino a América, aunque estimaciones posteriores sitúan esta cifra en poco más de 4.7 millones (Yáñez 1988). En cualquier caso, las deficiencias que presentan las series del IGE son notables. Más allá de la tradicional dificultad para captar estadísticamente el fenómeno de la migración, las estadísticas españolas registraban exclusivamente las salidas desde puertos españoles, dejando fuera de recuento aquellos españoles que llegaron a sus destinos desde otros países, por ejemplo Francia, Portugal o Cuba. Las deficiencias en el registro se ponen de manifiesto al comparar las cifras de salidas con las de llegadas registradas, por las estadísticas americanas (Sánchez Alonso 1995, p.113-117). En el caso de los españoles en Florida, los vínculos existentes entre Cuba y la comunidad española en Tampa obligan a considerar especialmente el problema del subregistro, puesto que Cuba constituía un paso intermedio entre España y los Estados Unidos para una buena parte de los llegados<sup>3</sup>. Para salvar estas limitaciones, algunos autores han recurrido a otras fuentes, como en el caso de Carbó (1991a), que utiliza datos de la Oficina de Inmigración y Nacionalización estadounidense y los censos de población. La publicación de las muestras de microdatos de los Censos históricos de Estados Unidos amplía substancialmente la información disponible de los emigrantes a este país.

Precisamente, en este artículo, pretendemos mostrar la enorme potencialidad de los microdatos censales como fuente para el análisis, en nuestro caso, de las características demográficas y pautas matrimoniales de la comunidad española en Tampa, Florida. Los datos utilizados proceden de los censos estadounidenses de 1880, 1900, 1920 y, en particular, del Hispanic Oversample del censo de Estados Unidos de 1910.

#### La obtención de los datos

El sistema de extracción de datos de IPUMS permite a los usuarios confeccionar a medida sus propios pedidos de datos, como si de un vestido se tratara, según los intereses y necesidades específicas de cada investigación, evitando la descarga de

toda la información.

Para solicitar datos el usuario, una vez registrado, selecciona la opción **Crea un Extracto** de la página principal de IPUMS-USA. A continuación describimos el proceso de solicitud de los datos, con especial énfasis en la selección de los censos y de las variables.

#### Selección de los censos

En la primera etapa del proceso de extracción de datos se definen las características generales de la muestra de datos a solicitar. El usuario tiene la posibilidad de escoger una o más opciones para los siguientes criterios: i) estructura de los datos: rectangular o jerárquica; ii) tipo de archivo: SPSS, SAS o Stata; iii) y los censos.

Imagen 2. Muestras censales estadounidenses disponibles. IPUMS

Samples help

*The following are all of the available reglg samples*

- [1850 Free Population](#)
- [1860 Free Population with Black Oversample](#)
- [1860 Free Population](#)
- [1870 General](#)
- [1870 General with Black Oversample](#)
- [1880 General](#)
- [1900 General Sample - Minnesota \(preliminary\)](#)
- [1910 General](#)
- [1910 Black Oversample](#)
- [1910 Hispanic Oversample](#)
- [1920 General](#)
- [1940 General](#)

Los datos seleccionados en esta investigación proceden de los censos de 1880, 1900, 1910 y 1920 (Imagen 2). Los microdatos de 1880 incluyen el 100% de los individuos censados. En cambio, la densidad muestral de las muestras de 1900 y 1920 es del 1%. Dado que la población española, aún siendo numerosa en Tampa, no era muy numerosa en el conjunto de Estados Unidos, la muestras de 1900 y 1920 no permiten analizar en detalle la población de origen español en este país, aunque si que proporcionan un imagen nítida y útil para conocer su distribución por estados, edades, sexo y año de llegada. Existe una muestra general del Censo de 1910 para el conjunto de los Estados Unidos, aunque su densidad es menor a las anteriores, 1/ 750. De este mismo censo, sin embargo, disponemos del Hispanic Oversample (Gutmann, M.P. et al. 1999), cuya generosa densidad muestral permitió aproximarnos a la comunidad de origen español y sus descendientes directos en Estados Unidos, eliminando así, aunque sólo para un área específica, los temores de escasa confiabilidad estadística que suscitaban las muestras antes citadas. La riqueza de esta muestra permite el análisis pormenorizado de la población de origen mexicano, cubano y español y descendientes directos de

éstos para 57 condados repartidos entre los estados de Arizona, California, Florida, Kansas, Nuevo México y Tejas. En el caso de la población de origen cubano y español, la muestra incluye uno de cada dos hogares en los que residiera una persona hispana, que cumpliera uno de estos tres requisitos: i) nacida en Cuba o España; ii) de padre y/o madre nacido en Cuba o España iii) que hablara español; iv) identificada como raza española o cubana; v) con apellido hispánico. La mayoría de los españoles, aproximadamente la mitad de los que residían en Florida, vivían en el condado de Hillsborough, más concretamente en Ybor City, Tampa. Así pues, el Hispanic Oversample de 1910 informa del 50% de la población hispana de origen español y de los que conviven con ésta.

Según el censo de 1880, en el estado de Florida residían 498 españoles y descendientes directos de estos. Se trata, por tanto, de un contingente relativamente escaso. Veinte años más tarde, la muestra del 1% de 1900 arroja un total de 47 personas, que extrapolado al conjunto de la población nos daría aproximadamente como resultado una población de 4.700 personas. Entre 1880 y 1900, la comunidad española en Florida se habría multiplicado por 10. Finalmente, en la muestra del

Tabla 1. La comunidad española en Florida según las muestras censales seleccionadas. Distribución por sexo, edad y origen o ascendencia española.

Censo	Edad	Nacidos en España		Padre y/o madre español		Total
		Hombres	Mujeres	Hombres	Mujeres	
1880 (100%) General Florida	0-15	3	3	83	62	151
	16-64	90	19	113	92	314
	65 o +	6	4	10	13	33
	Total	99	26	206	167	498
1900 (1%) General Florida	0-15	1		5	8	14
	16-64	13	3	9	7	32
	65 o +	1				1
	Total	15	3	14	15	47
1910 (50%) Hispanic Oversample Hillsborough (FL)	0-15	28	12	310	274	624
	16-64	545	163	111	169	988
	65 o +	6	2	1	3	12
	Total	579	177	422	446	1624
1920 (1%) General Florida	0-15	2		18	12	32
	16-64	21	9	12	10	52
	65 o +	2				2
	Total	25	9	30	22	86

Fuente: Elaboración propia a partir de los microdatos censales

1% de 1920 aparecen 86 españoles y descendientes directos, aproximadamente el doble que en 1900. Pese a la imprecisión de los datos, los resultados obtenidos, como mostraremos más adelante, son coherentes con la historia de la comunidad española en Florida.

Los datos de 1910 no pueden compararse a los anteriores, puesto que no hacen referencia al mismo territorio. La comunidad española en Tampa estaba formada por aproximadamente 3.248 personas (1.624 x 2). La mitad de la comunidad española en Florida residía en Tampa. De aquí su importancia y representatividad en el conjunto de la emigración española a Florida.

### Selección de variables

En la segunda etapa del proceso de selección se escogen las variables con las que se desea trabajar (Imagen 3). Sólo están disponibles las variables<sup>4</sup> presentes en los censos seleccionados en la etapa anterior. La presencia de una variable se denota con el símbolo X. Para acceder a la documentación asociada a cada variable, los metadatos, basta con hacer doble clic en el nombre de la variable y se abre una ventana con la información correspondiente<sup>5</sup>.

Imagen 3. Selección de las variables. IPUMS

Core Demographic Variables (Person)						
Variable Name	Variable Description	Case Selected	1880	1900	1910	1920
<input checked="" type="checkbox"/> <a href="#">RELATE</a>	Relationship to household head -- General	<input type="checkbox"/>	X	X	X	X
<input checked="" type="checkbox"/> <a href="#">RELATE</a>	Relationship to household head -- Detailed	<input type="checkbox"/>	X	X	X	X
<input checked="" type="checkbox"/> <a href="#">IMPREL</a>	Imputed relationship to household head	<input type="checkbox"/>	X	X	X	X
<input checked="" type="checkbox"/> <a href="#">AGE</a>	Age	<input type="checkbox"/>	X	X	X	X
<input checked="" type="checkbox"/> <a href="#">SEX</a>	Sex	<input type="checkbox"/>	X	X	X	X
<input checked="" type="checkbox"/> <a href="#">RACE</a>	Race -- General	<input type="checkbox"/>	X	X	X	X
<input checked="" type="checkbox"/> <a href="#">RACE</a>	Race -- Detailed	<input type="checkbox"/>	X	X	X	X
<input checked="" type="checkbox"/> <a href="#">MARST</a>	Marital status	<input type="checkbox"/>	X	X	X	X
<input checked="" type="checkbox"/> <a href="#">DURMARR</a>	Duration of current marital status	<input type="checkbox"/>	.	X	X	.
<input checked="" type="checkbox"/> <a href="#">MARRNO</a>	Times married	<input type="checkbox"/>	.	.	X	.
<input checked="" type="checkbox"/> <a href="#">CHBORN</a>	Children ever born	<input type="checkbox"/>	.	X	X	.
<input checked="" type="checkbox"/> <a href="#">CHSURV</a>	Children surviving	<input type="checkbox"/>	.	X	X	.
<input checked="" type="checkbox"/> All Core Demographic Variables						
Ethnicity/Nativity Variables (Person)						
Variable Name	Variable Description	Case Selected	1880	1900	1910	1920
<input checked="" type="checkbox"/> <a href="#">BPL</a>	Birthplace -- General	<input type="checkbox"/>	X	X	X	X
<input checked="" type="checkbox"/> <a href="#">BPL</a>	Birthplace -- Detailed	<input type="checkbox"/>	X	X	X	X
<input checked="" type="checkbox"/> <a href="#">MBPL</a>	Birthplace of mother -- General	<input type="checkbox"/>	X	X	X	X
<input checked="" type="checkbox"/> <a href="#">MBPL</a>	Birthplace of mother -- Detailed	<input type="checkbox"/>	X	X	X	X
<input checked="" type="checkbox"/> <a href="#">FBPL</a>	Birthplace of father -- General	<input type="checkbox"/>	X	X	X	X
<input checked="" type="checkbox"/> <a href="#">FBPL</a>	Birthplace of father -- Detailed	<input type="checkbox"/>	X	X	X	X
<input checked="" type="checkbox"/> <a href="#">NATIVITY</a>	Nativity recode	<input type="checkbox"/>	X	X	X	X
<input checked="" type="checkbox"/> <a href="#">CITIZEN</a>	Citizenship status	<input type="checkbox"/>	.	X	X	X
<input checked="" type="checkbox"/> <a href="#">YRIMMIG</a>	Year of immigration	<input type="checkbox"/>	.	X	X	X
<input checked="" type="checkbox"/> <a href="#">YRSUSA1</a>	Years in the United States	<input type="checkbox"/>	.	X	X	X
<input checked="" type="checkbox"/> <a href="#">YRSUSA2</a>	Years in the United States, intervalled	<input type="checkbox"/>	.	X	X	X
<input type="checkbox"/> <a href="#">MTONGUE</a>	Mother tongue -- General	<input type="checkbox"/>	.	.	X	X

Como se observa en la Imagen 3, todas las variables de carácter demográfico fueron incluidas. Éstas son sexo, edad, relación con la persona principal del hogar, estado civil, duración del matrimonio, entre otras. El segundo bloque de variables seleccionadas incluye datos sobre etnia y origen de los individuos, tales como el lugar de nacimiento del sujeto, de sus padres, el año de llegada a los Estados Unidos, o su condición de ciudadano o no.

La Tabla 2 muestra un ejemplo concreto de lo que son los microdatos. Los datos proceden del censo de 1900. La variable Serial identifica los hogares. Todas las personas que residen en un mismo hogar tienen el mismo código. De los tres hogares presentes, dos de ellos contienen población de origen español. Si tomamos como referencia el

segundo hogar (serial = 1406501) y observamos simultáneamente todos sus datos, podemos extraer información adicional a cerca de su historia migratoria y familiar. Se trata de un hogar de dos unidades familiares, pues en él residen dos parejas casadas. Los cuatro miembros del hogar llegaron a Estados Unidos el mismo año, en 1890, diez años antes de ser censados. Como conocemos la duración del matrimonio para cada una de las parejas, sabemos que la primera de ellas llegó casada a los Estados Unidos y también sabemos que tuvo una hija en Cuba, gracias a lo cual podemos afirmar que hubo una estancia previa en Cuba antes de su llegada a Estados Unidos. Repitiendo este ejercicio para todas las familias, podemos ir deshilando la información contenida en los microdatos, como si de una madeja se tratara.

Tabla 2. Muestra de microdatos del Censo de 1900

year	serial	pernum	perwt	relateg	age	sex	raceg*	marst*	durmar	bpld	mbpld	fbpld	yrimmig
1900	1406402	1	100	Head/Householder	40	M	W	MP	11	Missouri	Missouri	Missouri	N/A
1900	1406402	2	100	Spouse	33	F	W	MP	11	Florida	Florida	Germany	N/A
1900	1406402	3	100	Child	8	M	W	NM	99	Florida	Florida	Missouri	N/A
1900	1406402	4	100	Other non-relatives	18	F	B	NM	99	Florida	Georgia	Florida	N/A
1900	1406402	5	100	Other non-relatives	50	M	B	MA	8	N.Carolina	S.Carolina	S.Carolina	N/A
1900	1406501	1	100	Head/Householder	49	M	W	MP	28	Spain	Spain	Spain	1890
1900	1406501	2	100	Spouse	56	F	W	MP	28	Spain	Spain	Spain	1890
1900	1406501	3	100	Child	20	F	W	MP	0	Cuba	Spain	Spain	1890
1900	1406501	4	100	Child-in-law	18	M	W	MP	0	Spain	Spain	Spain	1890
1900	1406601	1	100	Head/Householder	40	M	W	MP	2	Spain	Spain	Spain	1895
1900	1406601	2	100	Spouse	38	F	W	MP	2	Spain	Spain	Spain	1899
1900	1406601	3	100	Child	1	F	W	NM	99	Cuba	Spain	Spain	1899

\* MP=married, spouse present; MA=married, spouse absent; NM=never married; W=white; B=black

### La comunidad española en Ybor City, Tampa: características demográficas

Tampa tenía en 1910 una población de 39.599 habitantes según el Censo. La población española de origen representaba el 5,9 % del total (Tabla 3). La mayoría de estos españoles residían en Ybor City, un asentamiento vinculado al establecimiento de la

industria tabaquera, dónde convivían con los cubanos e italianos en un contexto de segregación laboral y residencial respecto al resto de habitantes, blancos y negros, de la ciudad de Tampa. La localización geográfica y su característica especialización económica en la industria tabaquera (Carbó 1990a) explica la particular composición por origen de los habitantes de Ybor City (Tampa).

Tabla 3. Distribución por origen de la población de Tampa, 1890 – 1940 (%)

		1890	1900	1910	1920	1930	1940
Nacidos en los EEUU	Blancos	29,99	23,91	30,40	35,99	47,35	53,86
	Afroamericanos	19,79	22,89	22,60	23,66	23,26	25,03
Extranjeros	Anglosajones	17,31	22,84	24,99	21,88	15,95	11,89
	Cubanos	29,40	18,46	9,75	7,10	5,62	3,56
	Espanoles	2,83	5,03	5,90	5,59	3,80	2,79
	Italianos	0,68	6,87	6,36	5,78	4,02	2,88
Total		100 (8245)	100 (19141)	100 (39599)	100 (48741)	100 (91015)	100 (93215)

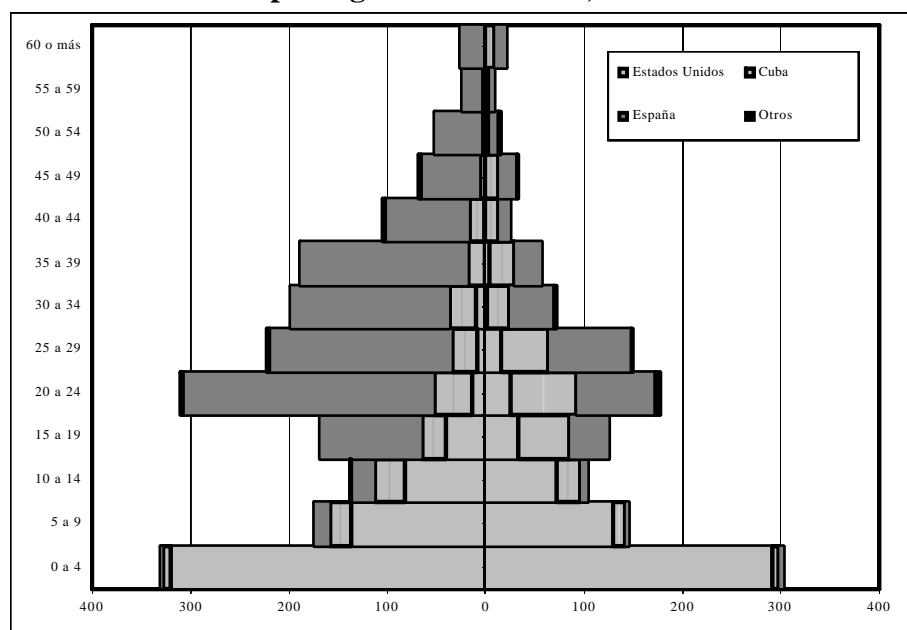
Fuente: Carbó (1990a)

Las cifras de Carbó (1990a) sólo tienen en cuenta los españoles de origen y no sus descendientes directos. En cambio, los microdatos del Censo de 1910 nos permitieron incluir a las segundas generaciones, formando parte de lo que hemos bautizado como comunidad española en Ybor City (Tampa), que reúne a un total de 3.248 personas cuya distribución por sexo y edades, en forma de pirámide de población (Figura 1), es altamente explicativa.

La Figura 1 informa de la distribución por sexo, edad y lugar de nacimiento de la comunidad española en Ybor City. Esta pirámide es el reflejo de la historia migratoria de estos españoles,

migración reciente y joven. La mayor afluencia masculina se traduce en significativos desequilibrios por sexo a partir de los 20 años. La natalidad de la comunidad española era elevada. La mayoría de los hijos de españoles nacieron en Estados Unidos. No menos importante, sobretudo a partir de los 10 años, son los nacidos en Cuba que emigraron a Ybor City, acompañados de sus padres o protagonizando ellos mismos el desplazamiento. La presencia de segundas generaciones nacidas en Cuba se observa en todas las edades, pues la presencia de españoles en este país es anterior a la emigración a Estados Unidos. La relación con Cuba es una característica que distingue Florida de las demás destinaciones dentro de Estados Unidos.

**Figura 1. Pirámide de población de la comunidad española en Tampa por lugar de nacimiento, 1910**

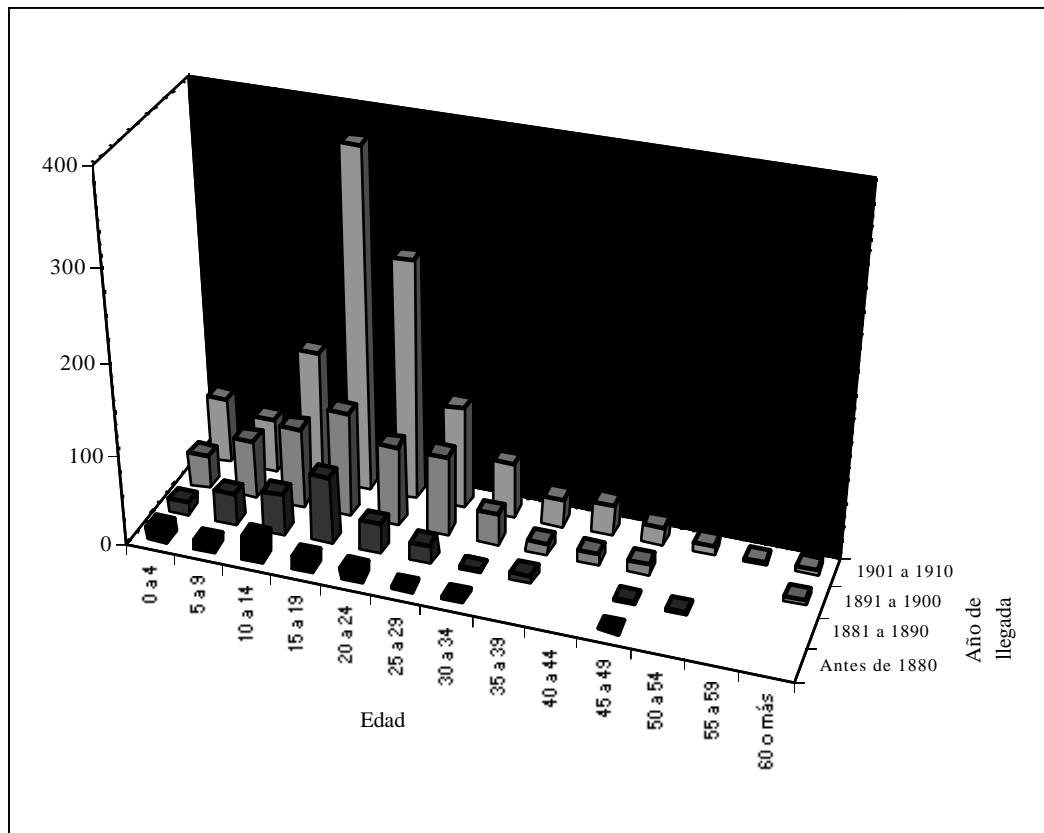


**Fuente:** Elaboración propia a partir de la Hispanic Oversample del Censo de los Estados Unidos de 1910

La distribución de los emigrantes españoles en función del año de llegada y la edad a la llegada a Estados Unidos nos muestra su distribución por edades e intensidad de cada periodo (Figura 2). En ningún caso se pretende reconstruir a partir de los datos del Censo de 1910 los flujos migratorios del pasado, puesto que se desconocen las salidas de observación debidas a la mortalidad o a la migración. Sin embargo, al tratarse de una migración reciente, joven y de carácter estable hacia una ciudad en expansión, consideramos oportuno presentar este

gráfico. La distribución por edades de los emigrantes se mantiene constante y de acuerdo al perfil clásico de la migración laboral: preponderancia de adultos jóvenes (15-30 años) que arrastran a las edades infantiles (0-14 años) y poca presencia de individuos en edades maduras y avanzadas. La mayoría de los españoles en Tampa en 1910 habían llegado durante la década anterior. De hecho, el número de españoles llegados se incrementa de manera paralela al desarrollo de la industria tabaquera.

Figura 2. Distribución por año y edad de llegada de la población española residente en Tampa en 1910



Fuente: Elaboración propia a partir de la Hispanic Oversample del Censo de los Estados Unidos de 1910

La comunidad española en Ybor City, Tampa, contaba con un elemento distintivo y cohesionador clave: su ocupación laboral en el sector tabaquero. Los microdatos del censo indican que hasta un 85% de los hombres eran empleados por la industria tabaquera, el 15% restante estaba a cargo funda-

mentalmente de los establecimientos comerciales, hoteleros y de restauración. Este elemento es importante porque introduce un alto grado de cohesión interna a la vez que da a la comunidad española respetabilidad en el conjunto de la ciudad de Tampa. Los trabajadores españoles constituyeron además un foco obrerista destacado con altos nive-



les de sindicación y con predominancia de las corrientes anarquistas. Desarrollaron también una red asociativa con fines cooperativos y culturales entre las que destacan el Centro Español y el Centro Asturiano. Los italianos y los cubanos disponían también de sus propios centros, de tal manera que existía poca interrelación entre las distintas comunidades. Los matrimonios entre latinos en 1910 eran escasos según Mormino y Pozzeta (1998, p.253). Ahora bien, el desequilibrio de efectivos por sexo favorable a los hombres planteaba un escenario en el que no todos podían formar pareja con mujeres de su misma comunidad. Los efectos de este desequilibrio en los niveles de soltería y homogamia en hombres y mujeres son explorados a continuación. ¿Influyó positivamente esta situación en la formación de matrimonios mixtos, especialmente entre hombres españoles y mujeres de distinto origen y ascendencia? Si es así, ¿se estrecharían en un futuro los vínculos sociales entre las distintas comunidades latinas?

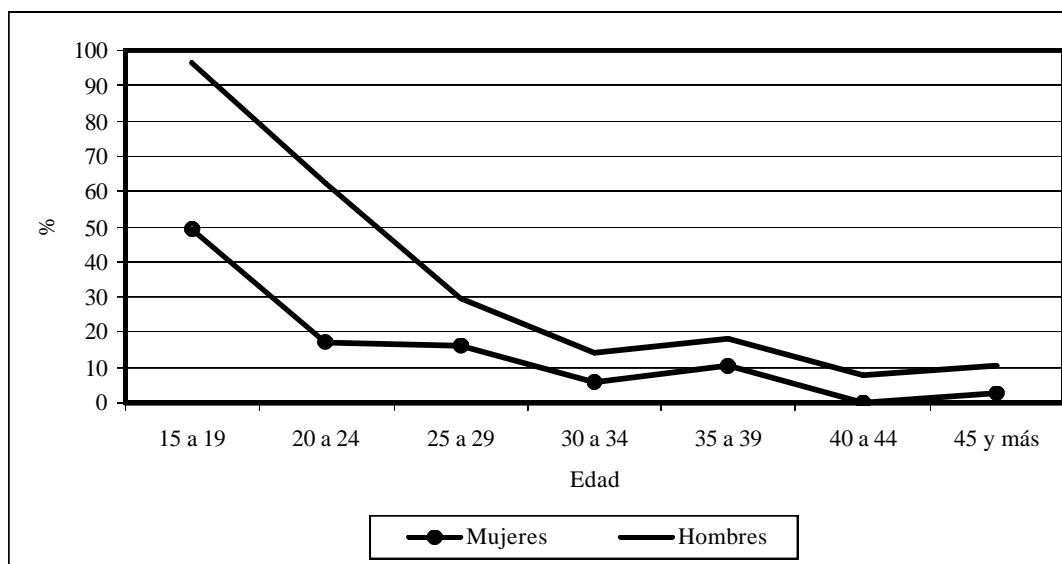
### Pautas matrimoniales

El estudio del comportamiento matrimonial de una población en toda su casuística supone tener en cuenta una amplia cantidad de factores, que es

posible desvelar gracias a los cruces entre variables que permiten los microdatos. Por ejemplo, conociendo la edad de llegada a los Estados Unidos y la duración del estado civil actual, es posible determinar, para los casados una sola vez<sup>6</sup>, si llegaron casados o no. A su vez, para los que llegaron casados, la comparación del año de llegada de ambos cónyuges informa sobre si éstos emigraron conjuntamente o uno de los cónyuges pospuso su llegada. El origen de estas parejas pudo ser determinado a partir de los microdatos, observando el lugar de nacimiento del sujeto, su pareja y sus ascendientes respectivos. Esta información permitió diferenciar entre matrimonios homogamos y heterogamos (entendiendo por homogamos aquellos constituidos por dos cónyuges españoles de origen o ascendencia, y heterogamos aquellos matrimonios en los que uno de los miembros no tiene ni origen ni ascendientes directos españoles).

Ante tanta complejidad, difícil de sistematizar en esta primera aproximación, estimamos conveniente acotar el terreno para centrarnos en los efectos de los desequilibrios entre efectivos de hombres y mujeres sobre la soltería y la formación de la pareja según el origen de los cónyuges, distinguiendo entre parejas homogamas y heterogamas.

Figura 3. Soltería por edad y sexo de la comunidad española en Tampa, 1910



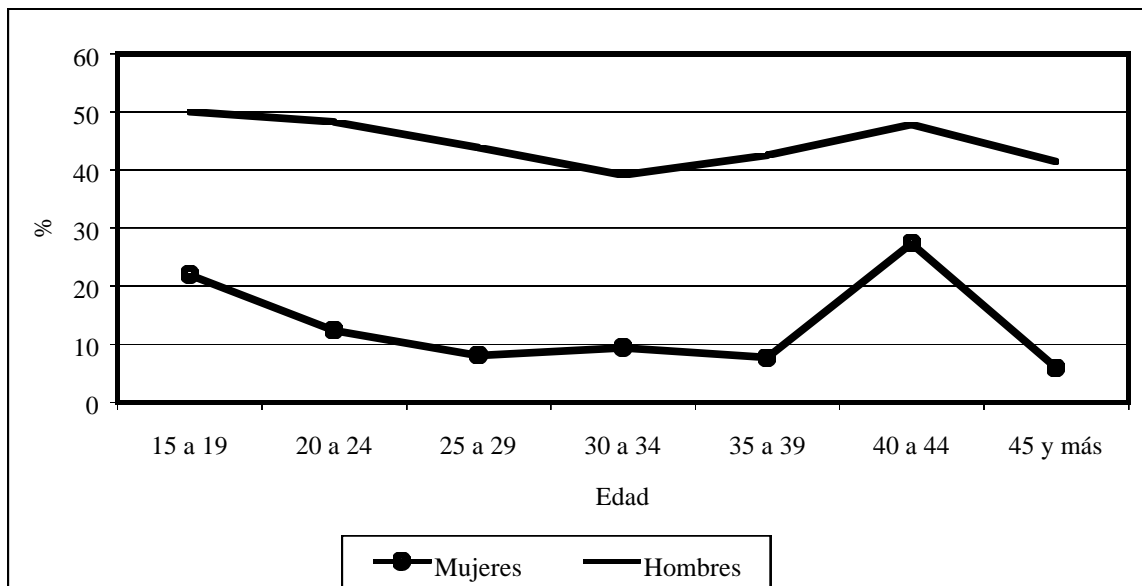
Fuente: Elaboración propia a partir de la Hispanic Oversample del Censo de los Estados Unidos de 1910

La distribución por edad y sexo de la soltería, expresada como la proporción de solteros/as respecto el total de hombres y mujeres respectivamente, informa para ambos sexos de la disminución de la soltería con la edad. Como muestra la Figura 3, el calendario nupcial presenta un desfase por género, siendo esta disminución más temprana entre las mujeres. La mitad de las mujeres entre 15-19 años ya se había casado alguna vez. Esta cifra sorprende si se compara con la proporción de solteros a las mismas edades, que supera el 95%. Las diferencias por sexo en la proporción de soltería no desaparecen con la edad, aunque tienden a

disminuir. La incidencia de la soltería entre las mujeres es menor a la de los hombres a todas las edades.

Este desequilibrio de efectivos entre sexos influye también en la formación de matrimonios heterogamos. El análisis de la heterogamia matrimonial es habitual en los estudios de población extranjera, al ser considerada una variable ilustrativa del grado de integración, forzada o voluntaria, de dicha población (Gray 1989, McCaa 1993, Stier 1994).

Figura 4. Heterogamia por edad y sexo de la comunidad española en Tampa, 1910



Fuente: Elaboración propia a partir de la Hispanic Oversample del Censo de los Estados Unidos de 1910

La heterogamia, expresada como la proporción de hombres/mujeres heterogamos/as sobre el total de casados/as, es a todas las edades inferior entre las mujeres (ver Figura 4), lo que significa que éstas se casaron mayoritariamente con hombres de su mismo origen o ascendencia. En cambio, no ocurre lo mismo entre los hombres, quienes formaban matrimonios mixtos en una proporción mayor a la de las mujeres. Por ejemplo, la proporción de mujeres heterogamas entre 25 i 29 años era inferior al 10%, mientras que la de los hombres a la misma edad rozaba el 45 %. La incidencia de la

heterogamia no presenta un patrón claro por edades. Entre los hombres, ésta se mantiene más o menos constante, indicando que todas las generaciones de españoles presentes en 1910 habrían encontrado dificultades parecidas en el momento de contraer matrimonio con mujeres de su misma comunidad. Sin embargo, debemos asumir con cautela la explicación anterior, dada la imposibilidad de reconstruir a partir de los datos de 1910 la situación que cada hombre vivió en el momento de formar pareja.

Las significativas diferencias entre hombres y mujeres en cuanto a la incidencia de la heterogamia son resultado de la desigualdad de efectivos por sexo. La práctica inexistencia de mujeres de origen o ascendencia española en matrimonios heterógamos es una prueba de que éstas son la opción deseada entre la población masculina de su mismo origen. En respuesta a estas desigualdades, los hombres

contrajeron matrimonio con mujeres de distinto origen. El 48 % de los hombres casados fuera de su comunidad lo hicieron con cubanas o hijas de cubanos nacidas en Estados Unidos, un 24,5% con italianas o hijas de italianos y sólo un 13% con estadounidenses de padres también estadounidenses (Tabla 4). El 87,8 % de los mujeres heterógamas se casó con cubanos o hijos de cubanos y sólo el 6,06 % lo hizo con estadounidenses.

Tabla 4. Matrimonios heterógamos, lugar de nacimiento de la pareja. Tampa 1910

		Lugar de nacimiento de la pareja				
		Total	Cuba	Italia	Estados Unidos	Otros
Hombres		350	168	86	48	48
	%	100	48.00	24.57	13.71	13.71
Mujeres		66	58	0	4	4
	%	100	87.88	0.00	6.06	6.06

Fuente: Elaboración propia a partir de la Hispanic Oversample del Censo de los Estados Unidos de 1910

En este tipo de análisis, los porcentajes tienen una capacidad descriptiva sin parangón entre el resto de indicadores, pero son poco informativos respecto de las interacciones entre los grupos. Un buen análisis de las pautas que gobiernan el mercado matrimonial, más allá de las restricciones que impone la desigual distribución de efectivos, exige conocer a todos los actores de este mercado. El Hispanic Oversample de 1910, como su propio nombre indica, sólo informa de la población hispana, sin incluir al resto de la población. Se trata, consecuentemente, de una imagen parcial, a partir de la cual no se puede examinar la interacción real entre las distintas comunidades. La muestra general del censo de 1910 disponible en IPUMS no es suficientemente representativa (1/750) para realizar este tipo de análisis. No obstante, próximamente IPUMS tiene previsto añadir a su sistema de extracción una muestra general de mayor densidad para este censo, que nos va a permitir responder alguna de las preguntas planteadas en esta investigación.

Las nuevas tecnologías pueden impulsar a la Demografía Histórica hacia una nueva etapa, en la que el desarrollo de la disciplina no ha de verse dificultado por las tradicionales dificultades asociadas a la captación y manejo de los datos. El paradigma IPUMS ha resuelto con éxito parte de éstas dificultades. Con la ayuda de la tecnología y el esfuerzo intelectual de un amplio equipo de investigadores, IPUMS ha resucitado, con finalidades académicas, los datos de los censos históricos de Estados Unidos. Desde su publicación, IPUMS-Estados Unidos ha sido la base de datos más consultada de este país. Entre sus usuarios encontramos mayoritariamente economistas y sociólogos, pero también demógrafos, historiadores y geógrafos. Según los resultados de una encuesta reciente, dirigida a los investigadores registrados de IPUMS, la mayoría de los encuestados señalaron al sistema de extracción de datos como el principal atractivo del proyecto, por su flexibilidad, agilidad y diseño innovador.

La utilización de los microdatos históricos de

Estados Unidos para el estudio de la emigración a Florida es sólo una de las muchas aplicaciones que pueden hacerse con esta fuente. Con este ejemplo, hemos querido mostrar como el análisis demográfico de las poblaciones históricas es hoy más posible que hace medio siglo, pese a estar más lejos temporalmente. Las innovaciones tecnológicas producidas en campos como la obtención, procesamiento y difusión de datos han abierto un océano de oportu-

nidades a la Demografía Histórica que, consecuentemente, redundarán positivamente en el mejor conocimiento de nuestro pasado. La experiencia de IPUMS debe ser aprovechada para emprender iniciativas similares, no necesariamente circunscritas al ámbito de los censos históricos, sino al conjunto de las fuentes históricas, tesoros aún por desenterrar, arrinconadas todavía en las salas más recónditas de las parroquias, archivos y museos, entre otros edificios.

## Bibliografía

- CARBÓ PEIRÓ, M., (1990a): Emigración vasca a los Estados Unidos. Su presencia y contribución al desarrollo del Estado de Florida desde finales del siglo XIX. Tampa 1886-1936. Barcelona: Universidad de Barcelona, p. 200, (mimeo).
- CARBÓ PEIRÓ, M., (1990b): La emigración española hacia los Estados Unidos de Norteamérica: 1898-1936. Universidad de Barcelona, Departamento de Historia de América. Tesis de licenciatura.
- ESTEVE, A., SOBECK, M. (2003): Challenges and Methods of International Census Harmonization. *Historical Methods* 36: 66-79.
- ESTEVE, A., TORRENTS, À., CORTINA, C. (en prensa): "La emigración española a Estados Unidos: una aproximación desde los microdatos censales de 1910". *Revista Scripta Nova*.
- FERNÁNDEZ-SHAW, C.M., (1987): Presencia española en los Estados Unidos. Madrid: Instituto de Cooperación Iberoamericana, Ed. de Cultura Hispánica.
- FERNÁNDEZ-SHAW, C.M., (1992): La Florida Contemporánea: Florida, tú eres. Madrid: Mapfre.
- GRATTON, B.; GUTMANN, M.P., (2000): "Hispanics in the United States, 1850-1990". *Historical Methods*, 2: p. 137-153.
- GRAY, A., (1989): "Measuring preference for in-marriage: a response to McCaa". *Population Studies*, 1: p. 163-166.
- GUTMANN, M.P.; FRISBIE, W.P.; BLANCHARD, K.S., (1999): „A New Look at the Hispanic Population of the United States in 1910“. *Historical Methods*, 1: p. 5-19.
- HALL, P.K., McCAA, R., THORVALDSEN, G., (2000): Handbook of international historical microdata for population research. Minneapolis: Minnesota Population Center.
- KLEIN, H.S., (1996): La inmigración española en Brasil (siglos XIX y XX). Colombes: Archivo de Indios.
- LICHTER, D.T.; LeCLERE, F.B.; McLAUGHLIN, D.K., (1991): «Local Marriage Markets and the Marital Behavior of Black and White Women». *American Journal of Sociology*, 4: p. 843-867.
- McCAA, R., (1993): "Ethnic intermarriage and gender in New York City". *Journal of Interdisciplinary History*, 24 (2): p. 207-231.
- McCAA, R., RUGGLES, S. (2003): The Census in Global Perspective and the Coming Microdata Revolution. In Vol. 13, *Nordic Demography: Trends and Differentials*, Scandinavian Population Studies, edited by J. Carling. Oslo: Unipub/Nordic Demographic Society, pp. 7-30.
- McDONALD, P., (1995): «L'équilibre numérique entre hommes et femmes et le marché matrimonial: le point sur la question». *Population*, 6: p. 1579-1590.
- MALUQUER, J. (1992) : Nación e inmigración: los españoles en Cuba. Gijón: Gran Enciclopedia Asturiana Silverio Cañada. p.168.
- MORMINO, Gary R.; POZZETTA, George E., (1998): The Immigrant World of Ybor City. Italians and Their Latin Neighbors in Tampa, 1885-1985. Gainesville: University Press of Florida.
- PÉREZ MURILLO, M.D., (1999): Cartas de emigrantes escritas desde Cuba. Estudio de las mentalidades y valores en el siglo XIX. Sevilla: Aconcagua Libros, Colección La otra América.
- RUEDA, G., (1993): La emigración contemporánea de españoles a Estados Unidos, 1820-1950. Madrid: Mapfre, p. 360.
- RUGGLES, S.; SOBEK, M. et al., (1997): Integrated Public Use Microdata Series: Version 2.0. Minneapolis: Historical Census Projects, University of Minnesota.
- SÁNCHEZ ALONSO, B., (1995): Las causas de la emigración española, 1880-1930. Madrid: Alianza.
- SÁNCHEZ ALONSO, B., (1992): Inmigración española en Argentina: siglos XIX y XX. Gijón: Gran Enciclopedia Asturiana Silverio Cañada, p.168.
- SÁNCHEZ-ALBORNOZ, N., (1988): "Medio siglo de emigración masiva de España a América" en SÁNCHEZ-ALBORNOZ, N. (ed.) *Espanoles hacia América. La emigración en masa, 1880-1930*. Madrid: Alianza, p.13-29.
- SCHOEN, R., (1983): "Measuring the thightness of marriage squeeze". *Demography*, 1: pp. 61-78.
- STIER, H.; SHAVIT, Y., (1994): "Age at marriage, sex-ratios, and ethnic heterogamy". *European Sociological Review*, 1: p. 19-87.
- YÁÑEZ, C., (1988): La construcción de las series anuales de la emigración española a América, 1860-1930. *Papers de Demografia*, 30. Bellaterra: Centre d'Estudis Demogràfics.

## Notas de Rodapé

<sup>1</sup> Este trabajo se inscribe en el marco del Proyecto BS02000-0477: Población y migraciones en Cataluña. Estudio territorial, histórico y prospectivo, financiado por el Ministerio de Ciencia y Tecnología. Programa Nacional de Promoción General del Conocimiento. Plan Nacional I+D+I 2000-2003.

<sup>2</sup> Cualquier investigador que desea obtener datos debe registrarse como usuario. El acceso es restringido a investigadores acreditados, que desean llevar a cabo una investigación bien definida y que aceptan regirse por las condiciones de utilización. Para solicitar acceso es preciso cumplimentar un formulario y aceptar las condiciones y términos de uso, mecanismo que IPUMS utiliza para garantizar la utilización responsable de los datos. Desde la página web principal de IPUMS se puede acceder a este formulario seleccionando la opción **Apply for Access**

<sup>3</sup> Carbó (1990a) califica a los españoles de Tampa de “fantasmas” para la historiografía española porque su procedencia de Cuba borraba su rastro.

<sup>4</sup> Algunas variables ofrecen la posibilidad de elegir sólo aquellos casos que cumplan un determinado requisito.

<sup>5</sup> Antes de solicitar definitivamente el extracto, el usuario puede revisar sus contenidos y realizar los cambios que considere oportunos. Actualmente, el tiempo que demora el sistema en ejecutar el pedido de datos no supera los treinta minutos desde el momento en que se solicitan. En este espacio de tiempo, el usuario es informado vía email de que su extracto se encuentra en el servidor de IPUMS, listo para ser descargado.

<sup>6</sup> Esta información la conocemos a partir de la variable “Número de veces casado” registrada en el Censo de 1910.