

Sistemas Informáticos para análise de dados demográficos: uma abordagem histórica

VII Congresso da ADEH
Granada, Abril de 2004

Antero Ferreira
anteroferreira@neps.ics.uminho.pt

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.¹ Não é de estranhar que perante tais dificuldades os historiadores demógrafos começassem 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

¹ Sobre estas limitações metodológicas ver Jacques Dupâquier (1984). *Pour la demographie Historique*. Paris: PUF.

na exploração dos registos paroquiais e das genealogias”². 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.”³ Apontava nessa oportunidade como principais limitações à utilização da informática, a necessidade de formalizar os ficheiros, os 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.

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 démographie Historique*. Paris: PUF.

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.”*⁴ 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 actual⁵ 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

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).

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 conseqüentemente 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.*”⁶

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

⁶ H. Charbonneau e J. Légarè (1979). O Uso de Computadores em Demografia Histórica. S. Paulo: Vozes, p. 100

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.”*⁷ 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

⁷ H. B. Newcombe; J. M. Kennedy, S. J. Axford & A. P. James (1959). Automatic Linkage of Vital Records. *Science*, 130(3381), 954-959.

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 Code⁸ 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

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.

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 adequado⁹. 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 rigor¹⁰.

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

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.

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 investigações¹¹.

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

¹¹ Na bibliografia apresentamos um conjunto de títulos representativo deste processo em países como Portugal, Espanha, Noruega, Dinamarca, Brasil.

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 paróquia¹². 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.

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.

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 imaginação 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) e o Projecto BALSAC (www.uqac.ca/balsac), na Inglaterra, o Projecto de Macfarlane

sobre Earls Colne (linux02.lib.cam.ac.uk/earlscolne), em Portugal o projecto do NEPS, (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 panaceia 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 reproduzindo, 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. "Rebordãos 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 Demografia 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 LÉGARÈ, 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 dépoulement 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.

NEWCORBE, 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