

ENSAIO

BANCOS DE DADOS E PESQUISA QUALITATIVA EM HISTÓRIA: REFLEXÕES ACERCA DE UMA EXPERIÊNCIA

Edson Armando Silva*

Escrever sobre o uso do computador na pesquisa histórica (mas não somente nela) implica em algumas graves dificuldades, que devem ser explicitadas já de início. A primeira delas é que vivemos atualmente em um mundo em que a velocidade da informação é perturbadora. Mesmo os mais atentos às novidades não conseguem acompanhar com eficiência as novas técnicas e programas diariamente apresentados ao mercado. Novas versões de softwares enchem as prateleiras das lojas especializadas e as páginas da internet, oferecendo constantemente novos recursos. Muitos procedimentos laboriosamente desenvolvidos são substituídos por novos recursos nas versões atualizadas dos programas com os quais trabalhávamos há alguns meses. No processo de relatar, a escrita congela a experiência que vem à luz já defasada em relação às últimas novidades. O texto, pouco tempo depois de publicado, já ganha um ar de ultrapassado, submetendo o autor a uma sensação desagradável de obsolescência.

A decisão de narrar a experiência, mesmo conhecendo de antemão a sua parcialidade e fugacidade, se prende à convicção de que técnica e teoria são indissociáveis. Silenciosamente, a mudança de procedimentos vai alterando o “*savoir faire*” e repercute no ofício do historiador. Seria imperdoável a ele, profissional da temporalidade, desconsiderar o impacto da tecnologia da distribuição social do trabalho e do conhecimento. As transformações técnicas não são e nunca foram

*. Professor do DEHIS/UEPG. Doutorando em História na UFF. Integrante do NUPEHC.
E-MAIL: edasilva@uepg.br

inócuas. Elas resultam de um processo social de distribuição de trabalho e conhecimento mas alteram estas mesmas relações modificando a própria constituição do humano enquanto ser social. Como esperar que o ofício de historiador permaneça imaculado em relação às mudanças tecnológicas que se processam na nossa cultura?

Não se tratam de mudanças epidérmicas. As mudanças na linguagem do historiador, impostas pelas novas mídias, e as mudanças nos procedimentos de pesquisa e na organização das fontes ameaçam a constituição do próprio campo e inquietam os historiadores mais atentos. As reações são por vezes de um otimismo excessivo que tende a ver os recursos informáticos como bruxarias que resolvem de maneira quase mágica nossos problemas ou, no pólo oposto, como ameaça demoníaca de dissolução do saber erudito tradicional. A sensação de exterioridade do historiador em relação aos procedimentos informatizados faz com que ele os represente ora como Merlin ora como Mefisto¹.

Para romper esta dualidade de representações talvez seja útil a publicação de algumas experiências que sirvam de motivo de reflexão sobre o impacto da informática no campo do historiador. Não se tratam de modelos a serem aplicados acriticamente. São experiências que apresentam possibilidades e limites e que podem eventualmente orientar o historiador que deseja utilizar novas técnicas em suas pesquisas, na escolha do equipamento, do software ou no desenvolvimento de procedimentos. Mas a eficácia das técnicas utilizadas não depende apenas da escolha do equipamento ou dos programas. A definição dos procedimentos e sua correta configuração podem significar uma diferença radical de produtividade.

O uso da informática na pesquisa histórica não é novidade. Já nos principais congressos internacionais de historiadores se introduziam discussões sobre o uso de computadores na pesquisa². Edward Shorter, em 1971, publicou um manual intitulado “The Historian and the computer”³ e um dos mais populares manuais de metodologia em História

¹. Ciro Flammarion Cardoso escreveu um artigo intitulado “No limiar do século XXI” na revista **Tempo**, Rio de Janeiro, vol 1 n. 2, 1996, p. 7-30, no qual ele defende a idéia de que as novas tecnologias ameaçam o registro escrito e erudito da linguagem.

². Apenas para estabelecer alguns marcos podemos citar a V Conferência internacional de História Econômica em Leningrado (1970); o congresso internacional de ciências históricas em Moscou (1970) e a Primeira conferência sobre História e Computação (1973).

³. SHORTER, Edward. **The Historian and the Computer - A practical guide**. Prentice Hall, Nova Jersey, 1971.

no Brasil⁴, do final década de setenta, fala de cartões perfurados para a tabulação de dados históricos.

O desenvolvimento da tecnologia dos computadores pessoais, com o conseqüente barateamento dos equipamentos, disponibilizou o seu acesso a um número maior de pesquisadores. Criaram-se softwares de banco de dados relativamente amigáveis e baratos. Entretanto, a influência das técnicas informatizadas nos procedimentos de pesquisa histórica, mesmo não sendo recente, foi superficial. Essas tecnologias eram aplicáveis, com ganhos de produtividade, apenas em fontes seriais.

Os historiadores que se dedicavam à história econômica ou demográfica foram os primeiros que se beneficiaram das novas tecnologias. E mesmo eles tinham dificuldades de encontrar séries suficientemente numerosas, rigorosamente datadas, com os dados homogêneos, comparáveis e coerentes com as questões formuladas. Historiadores que trabalhavam com fontes qualitativas de natureza textual tinham ainda mais dificuldade em adaptar as suas fontes à tecnologia disponível. A necessidade de dominar ou desenvolver softwares adequados a documentos que apresentavam uma grande variedade de formatos desencorajava os historiadores, uma vez que apresentava uma relação custo/benefício negativa.

O processo de adaptar os documentos a bancos de dados do tipo relacional implica num pesado processo de formalização dos documentos. É necessário adaptar a forma original do documento a um modelo reconhecível pelo computador. Dessa maneira, se bem realizado, o processo resultaria na transformação de um documento em uma seqüência de enunciados adaptados ao modelo. Levando em conta a equivalência funcional dos termos na linguagem, é possível supor a manutenção do sentido da afirmação, mas se perde a disposição das palavras na frase e das frases no discurso, colocando em risco a apreensão da relação lógica das frases entre si.

O processo implica então em uma reconstrução do documento muitas vezes perigosa, porque os modelos são teoricamente informados e pode-se pretender demonstrar o que **já estava presente na construção do modelo**. Isso expressa uma tautologia ou significa uma trabalhosa constatação do óbvio. Por essas razões, o uso de banco de dados permaneceu restrito a historiadores cujas fontes na origem já

⁴. CARDOSO, Ciro Flamarion e BRIGNOLI, Hector Perez. **Os Métodos da História**, Rio de Janeiro: Graal, 1979.

apresentavam determinados padrões, como por exemplo registros de casamento ou relações de mercadorias importadas.

A grande vantagem do tratamento informatizado, que é a possibilidade de tratar um grande corpo documental, se dissipava diante das dificuldades e problemas da formalização dos documentos em bancos de dados do tipo relacional, tornando muito mais produtiva a análise intuitiva fundamentada em exemplos singulares bem escolhidos. Apesar da informatização crescente em nossa cultura, a categoria de software mais difundida entre historiadores foi o “processador de texto” que simplifica o processo de revisão e reorganização da escrita.

O declínio da história quantitativa marcou também o declínio do uso dos bancos de dados pelos historiadores. A década de 90, se por um lado marca o crescente distanciamento da história quantitativa, marca também, se considerarmos as experiências apresentadas em congressos internacionais de história e informática como indicadores, a ampliação do uso da informática em outros campos da história. No campo da história da arte deu-se o desenvolvimento de sistemas de informações iconográficas através de processos de digitalização; diversas instituições passam a divulgar os seus acervos através de sistemas de informações que incluíam estratégias de inteligência artificial; amplia-se o uso de simulação e realidade virtual para divulgação do conhecimento histórico⁵.

Apesar da maioria dos historiadores se restringir ao uso do processador de texto, diversas instituições se encarregaram de introduzir no campo historiográfico técnicas avançadas de computação. Entretanto, a reflexão sobre o uso de bancos de dados na pesquisa histórica deve observar uma diferença entre os objetivos, possibilidades e limites no âmbito das instituições responsáveis pela guarda e publicidade de documentos históricos e pesquisadores individuais.

Quando para uma instituição se coloca o problema de disponibilizar as informações sob sua guarda através de sistemas informatizados, os procedimentos e cuidados obedecem a objetivos imediatos diferentes daqueles do historiador ao organizar suas informações. Charles Dollar afirmava em 1989 que eram inúmeros os

⁵. FIGUEIREDO, Luciano R. História e Informática: o uso do computador. In CARDOSO, Ciro F e VAINFAS, Ronaldo (Orgs). **Domínios da História: Ensaio de teoria e metodologia**. Rio de Janeiro: Campus, 1997.

projetos em andamento para converter fontes impressas ou manuscritas em formato digital⁶.

As questões que as instituições se colocam sobre o uso de sistemas de informação digitalizada dizem respeito à unicidade dos registros digitalizados em uma base de dados compartilhada; à autenticidade das informações digitalizadas uma vez que a tecnologia disponível permite uma recriação do documento sem que permaneçam vestígios; e à preservação ao longo do tempo de documentos eletrônicos face à velocidade do desenvolvimento tecnológico. Dollar responde a essas questões apontando para o desenvolvimento de tecnologias de autenticação eletrônica, o que garantiria a unicidade e a autenticidade dos documentos, e para procedimentos de recopagem sistemática e migração periódica para novas plataformas tecnológicas, que juntamente com procedimentos mínimos de normalização garantiriam a legibilidade e recuperabilidade das informações digitalizadas.

Os objetivos e prioridades do pesquisador, que procura responder a questões mais específicas, são diversos. Trata-se de desenvolver ou adaptar um software à documentação com a qual ele se defronta. Deve-se levar em conta a facilidade de aprendizado, a flexibilidade e a possibilidade de operação com dados produzidos em plataformas diversas. No desenvolvimento da pesquisa, a flexibilidade e facilidade de operação é mais importante do que garantir a recuperabilidade a longo prazo ou a imutabilidade dos documentos. Ao historiador envolvido em um programa de pesquisa interessa muito mais a capacidade do programa de relacionar informações e recuperar documentos que respondam às questões formuladas por ele.

Alguns historiadores, juntamente com equipes interdisciplinares, se lançaram ao desenvolvimento de programas para historiadores e arquivistas. A revista *Ler História* n.º 23, publicada em 1992, descreve duas experiências realizadas por arquivos portugueses. Entretanto, a experiência mais conhecida talvez seja a desenvolvida por Manfred Thaller no Instituto Max-Planck em Göttingen na criação do programa KLEIO. Este programa representou um avanço conceitual em relação aos bancos de dados relacionais para o trabalho do historiador. O programa se baseia no conceito de base de dados orientada à fonte, de maneira que

⁶. DOLLAR, Charles. Tecnologias da informação digitalizada e pesquisa acadêmica nas ciências sociais e humanas: o papel crucial da arquivologia. In **Estudos históricos**. Rio de Janeiro, vol. 7, n. 13, 1994, p.65-79.

não é necessário pré-definir uma estrutura rígida ou uma rigorosa hierarquia entre os elementos do documento.

Dessa maneira, o programa abre a possibilidade de identificar relações imprevistas no momento da criação da base de dados. Isso se torna possível porque cada grupo de elementos pode ser arbitrariamente relacionado a outros grupos. A recuperação dos elementos pode ser feita através de “redes semânticas”, isto é, é possível recuperar documentos que contenham determinadas expressões em qualquer grupo dentro da base⁷. Thaller construiu uma base de dados que, devido à sua flexibilidade no trato com a informação, tornou úteis expressões como “quinta feira santa do ano de 1632” sem que seja necessário “traduzi-la” para uma estrutura estranha ao documento. Para facilitar ainda mais o trabalho com as fontes, o programa inclui utilitários que auxiliam no trato com sistemas de datação, sistemas de medidas, sistemas monetários, informações toponímicas, etc.. Segundo Thaller:

In Kleio a historical source is administered by transcribing various parts of a source, assigned to individual elements of a database, as literally as possible. All knowledge about the meaning of the transcribed items is administered separately in a layer of the system which is specifically dedicated to the administration of knowledge. Any query the user makes, any command that has to access the data which are stored in the transcribed source, is interpreted according to the knowledge stored about the source.⁸

Apesar das qualidades do programa, a sua vulgarização entre historiadores encontra obstáculos. A distribuição é difícil e a atualização constante do programa para a manutenção da comunicabilidade com outros sistemas é dispendiosa. Isso somado ao pequeno conhecimento de informática dos historiadores e à dificuldade de manutenção de um esquema de suporte e treinamento torna duvidosa a manutenção do programa por um longo período. Portanto, muitos historiadores e instituições hesitam em investir na criação de uma base digital fundada numa plataforma tecnológica tão específica e com demanda restrita.

⁷. THALLER, Manfred. What is “Source Oriented Data Processing”; what is a “Historical Computer Science”? In **Historical informatics: an essential tool for historians**. Atlanta, Georgia, October 14th, 1994.

⁸. THALLER, Manfred, op. cit., p. .38.

Não se pretende aqui defender um monopólio das casas comerciais de software. Algumas equipes conseguiram superar as barreiras colocadas pelo mercado. Um caso paradigmático é o desenvolvimento do sistema operacional LINUX. O núcleo do programa, desenvolvido por um estudante finlandês, foi distribuído gratuitamente pela internet. Em torno do programa se reuniu uma comunidade que passou a desenvolver coletivamente o programa, resolvendo falhas, adaptando-o a novas plataformas, criando listas de discussão para auto-ajuda. Essas estratégias foram eficientes para gerar um sistema não-proprietário, seguro, distribuído gratuitamente e com suporte para quem se propõe a desenvolver aplicativos ou trabalhar com o sistema.

A questão que se coloca é se é possível na comunidade de historiadores criar um movimento semelhante. Para exemplificar as dificuldades que se apresentam é interessante relatar a experiência com um programa de reconstrução de famílias que teve aplicação entre historiadores da população e demógrafos: o SYGAP. O programa era fruto da produção conjunta de um grupo de pesquisadores ligados ao Departamento de Demografia da Universidade de Montreal e do Grupo Rendu-Osler, que aglutina a Maison Rhône-Alpes des Sciences de l'Homme, o Centre Pierre Léon e o Institut Européen des Génomutations.

Apesar do depoimento de um de seus usuários ser bastante favorável, afirmando que “o SYGAP representa um investimento de tempo e equipamento extremamente rentável, agilizando sobremaneira o trabalho do historiador de populações”, o projeto foi abandonado e os dados produzidos por este sistema são compatíveis somente com o padrão Dbase III⁹.

A alternativa a estas dificuldades é a adaptação de softwares comerciais às necessidades dos historiadores. Um dos programas que melhor se ajusta a estas tarefas é o AskSam ©. Este programa é um banco de dados que opera de forma estruturada e livre e que pode trabalhar com textos de até 16.000 linhas por documento¹⁰.

A flexibilidade do programa permite a construção de procedimentos semelhantes aos apresentados no Kleio. Embora não seja um software desenvolvido especialmente para historiadores e arquivistas,

⁹. PEREIRA, Marco Aurélio Monteiro. Reconstrução de famílias no microcomputador: o SYGAP. In **Revista Brasileira de Estudos de População**. Campinas, vol 12 n. 1-2, jan/dez 1995, p.193-200.

¹⁰. Maiores informações sobre o programa podem ser obtidas no site <http://www.AskSam.com>

eles podem fazer um bom uso dele em virtude da sua flexibilidade e facilidade de operação.

Da mesma maneira que a maioria dos softwares comerciais, pode-se contar com uma equipe de desenvolvimento e suporte que auxilia o historiador nos procedimentos de formalização dos documentos. Embora ele não tenha recursos de autenticação digital que respondam às questões de unicidade na guarda dos documentos, o que pode ser visto como um ponto fraco no aspecto institucional, ele é perfeitamente adaptável ao trabalho do pesquisador. O programa vem com uma coleção de filtros, (além de outros que podem ser acrescentados) que permitem com facilidade a transformação de bases de dados disponíveis em diversos formatos em bases de dados AskSam.

O trabalho de construção de fontes auxiliares de informações, como por exemplo índices que auxiliem o trabalho com sistemas de datação, informações toponímicas, sistemas de pesos e medidas e sistemas monetários, pode ser coletivamente organizado sem necessidade de conhecimento em linguagens de programação. Estabelecido um padrão tecnológico entre um grupo de pesquisadores, as informações poderiam ser alimentadas e consultadas através da internet. Dessa maneira, mesmo sem familiaridade com técnicas de programação, diversos historiadores, que trabalhem temas ou períodos semelhantes, poderiam disponibilizar uns aos outros suas anotações e descobertas de maneira a aumentar o volume de informações disponíveis sobre um determinado tema.

É claro que isso implicaria numa dinâmica de pesquisa diferente da que nós estamos habituados. O trunfo do pesquisador deixa de ser o conhecimento da localização dos documentos ou a posse da informação para ser a interpretação dessas informações.

Finalmente, gostaria de exemplificar o uso deste banco de dados com a pesquisa que estou atualmente desenvolvendo. As questões centrais envolvem elementos da identidade de um grupo religioso. Além da trajetória institucional básica a principal preocupação é estabelecer uma conexão entre os desafios institucionais e as possibilidades de construções identitárias. Dessa maneira foi necessário estabelecer relações entre a trajetória institucional e os discursos dos agentes e suas biografias.

O desafio era então organizar uma enorme quantidade de informações, que vão desde artigos publicados em periódicos que expressam juízos e revelam visões de mundo com as trajetórias biográficas de mais de seiscentos personagens. Isto está sendo possível

com o AskSam. Diversas buscas e relatórios podem ser realizados mesmo quando previstos na estrutura inicial. Assim é possível recuperar, por exemplo, toda frase que contenha a palavra “obediência” ou relacionar todo parágrafo onde seja citado o “frei Anastácio”. Esta maneira flexível de tratar os dados permite elaborar relatórios complexos que articulem campos estruturados e informações textuais. Isso torna possível construir relatórios que busquem, através de redes semânticas, por exemplo, noções de autoridade presentes em missionários franciscanos nascidos na Alemanha na segunda metade do século XIX.

É claro que a possibilidade de produzir relatórios complexos não basta. A atividade de pesquisa reclama muito mais nossa capacidade de fazer perguntas do que recursos técnicos que agilizem as respostas. Entretanto, a popularização dessas tecnologias tende a retirar a ênfase do domínio da informação para destacar as possibilidades de construção. A principal revolução é a associação dessas tecnologias de bancos de dados com as novas formas de comunicação. A pesquisa poderia então deixar de ser uma atividade solitária para se firmar como construção coletiva.

O Dicionário Histórico e Geográfico dos Campos Gerais, proposto pelo departamento no qual trabalho, exemplifica a idéia de associação de uma base de informações produzida e gerenciada pelo AskSam e sua divulgação através da Internet. O Dicionário optou por ser um veículo (de divulgação e produção de conhecimento) multimidiático, ou seja, ele será publicado em diversas mídias, tanto através de textos impressos como *on-line* (na internet). Essa característica permite a ele se propor como uma obra aberta e ilimitada, porque a mídia na qual ela será veiculada permite uma constante revisão atualização e ampliação do seu conteúdo. Isto a torna também uma obra interativa, uma vez que o público terá condições de opinar e contribuir com o seu conteúdo.

Muitas vezes os próprios pesquisadores têm dificuldades para encontrar informações sistematizadas sobre áreas de conhecimentos que não estejam exatamente dentro da sua especialidade. Ao mesmo tempo em que o Dicionário cria um espaço para sistematização e divulgação do conhecimento produzido na Universidade sobre a região, abre um canal de diálogo com todas as pessoas interessadas em conhecer e discutir a região dos Campos Gerais.

Finalmente, é necessário destacar que ainda não é possível prever as novas tendências na área de História e Informática a longo prazo. Entretanto, a simples constatação da velocidade das transformações nos permite afirmar, sem medo de errar, que estamos no limiar de um mundo

no qual nosso ofício deverá se modificar profundamente e que a informatização de nossa cultura deverá influir profundamente nossa maneira de pensar e produzir história.