

Projeto # 8
Memória Estatística do Brasil no acervo da Biblioteca do Ministério da Fazenda no Rio de Janeiro
Eustáquio Reis, IPEA-RJ

Objetivos

- Recuperar, digitalizar e divulgar na internet documentação e publicações contendo dados e estatísticas sobre a economia brasileira no Século XIX e primeira metade do Século XX existentes no acervo da Biblioteca do Ministério da Fazenda no Rio de Janeiro (BMF/RJ).

- Sistematizar e compilar bases de estatísticas econômicas, sociais e demográficas em níveis nacional, estadual e municipal para o Século XIX. Construir séries históricas os para principais indicadores econômicos e sociais, destacando-se estimativas de produto e renda, força de trabalho, estrutura ocupacional, distribuição de renda, estoques de capital e infra-estrutura, em particular de transportes, em níveis nacional, estadual e municipal para períodos anteriores a 1920, data do primeiro Censo Econômico do país.

Em termos amplos, o projeto visa a preservação científica, cultural e física do acervo da BMF/RJ cuja idade, raridade e precariedade do estado de conservação recomendam restrições do manuseio à finalidades bibliográficas específicas. O projeto possibilitará o acesso da comunidade científica e do público em geral às informações e obras raras do acervo existente na BMF/RJ.

Trata-se de projeto que já se encontra em desenvolvimento e que dá continuidade às linhas de pesquisas desenvolvidas pelo NEMESIS (Núcleo de Estudos e Modelos Espaciais Sistêmicos patrocinado pelo CNPq/PRONEX) sediado no IPEA/RJ.

Justificativas

A justificativa do projeto encontra-se, antes de mais nada, na carência e na dificuldade de acesso à informações estatísticas sistematizadas sobre aspectos econômicos, sociais e demográficos da história brasileira no Século XIX. Essa carência contrasta com a riqueza bibliográfica e científica do acervo da BMF/RJ que, como depositária das publicações e documentação do próprio Ministério da Fazenda e outros ministérios além de herdeira dos acervos das bibliotecas da Alfândega do Rio de Janeiro e dos extintos Instituto Brasileiro do Café (IBC) e do Instituto do Açúcar e do Alcool (IAA), conta com uma coleção de mais de 150 mil volumes, dos quais pelo menos 1.200 podem ser classificados como obras raras da história econômica e financeira do país.

Destacam-se nesse sentido as coleções históricas de anuários e relatórios estatísticos iniciando-se em meados do Século XIX que por sua idade, raridade, condições de conservação mas sobretudo pela riqueza de informações estatísticas sobre os mais diversos aspectos da história brasileira merecem e requerem preservação, recuperação e divulgação em meio eletrônico para permitir ampla disseminação entre acadêmicos e pesquisadores das diversas áreas, jornalistas, estudantes e do público em geral. Merecem menção especial os anuários estatísticos de finanças públicas desde 1830; de comércio exterior e transporte desde 1840; as estatísticas ferroviárias desde 1870, entre vários outros. Por fim, nota-se que para o período histórico de interesse para o projeto inexistia ainda um sistema estatístico nacional que se inaugura com a criação do Instituto Brasileiro de Geografia e Estatística (IBGE) em 1937 e da Fundação Getúlio Vargas (FGV) em 1945.

Em termos de preservação, cabe notar que após um século de exposição à umidade típica do Rio de Janeiro, o acervo bibliográfico da BMF/RJ foi posteriormente submetido ao ressecamento que caracteriza os ambientes com ar-condicionado. Essas condições climáticas adversas e o próprio tempo fizeram que parte considerável do papel impresso se tornasse especialmente acidificado, corroído e quebradiço implicando perda de informações históricas e estatísticas, sobretudo quando manuseado sem a devida cautela. Por fim, nota-se que, apesar da importância científica que esses acervos estatísticos possuem para a memória histórica do país, dificilmente apresentam atrativos para iniciativas de republicação em formato impresso.

As fontes de inspiração do projeto, mantidas as devidas proporções, encontram-se em projetos como: LexisNexis (<http://www.lexisnexis.com>) cujo objetivo é digitalizar, indexar e disponibilizar na internet 200 anos documentos não-sigilosos do governo dos EUA; Um Milhão de

Livros (http://www.library.cmu.edu/Libraries/MBP_FAQ.html) da Universidade Carnegie Mellon; e a Biblioteca Infinita do Google (Huttner 2005).

Linhas de ação

- Digitalização eletrônica e divulgação na internet do acervo científico e cultural sobre o Brasil existente na Biblioteca do Ministério da Fazenda no Rio de Janeiro (BMF/RJ). Ações específicas nesse sentido são:
 - Identificar e selecionar acervo bibliográfico e, em particular, as coleções de anuários e relatórios estatísticos relevantes para digitalização e divulgação na BMF/RJ
 - Catalogação e indexação bibliográfica do acervo bibliográfico e estatístico de interesse.
 - Digitalização eletrônica do material selecionado no acervo
 - Indexação eletrônica do material digitalizado de forma a permitir a pesquisa informatizada
 - Divulgar gratuitamente acervo bibliográfico e estatístico de interesse em formato Adobe Acrobat (pdf) ou similar pela internet para acadêmicos e pesquisadores das diversas áreas, jornalistas, estudantes e o público em geral.
 - Criar homepage do projeto com mecanismos de busca e recuperação
 - Compilar em formatos de bases de dados computacionais (planilhas, etc.) estatísticas históricas sobre as condições demográficas, sociais, econômicas e financeiras do país e, sobretudo, dos estados e municípios brasileiros durante o Século XIX e a primeira metade do Século XX quando inexistia um sistema estatístico nacional. Essa linha de ação vem complementar os esforços de sistematização das estatísticas históricas brasileiras ora em desenvolvimento em www.ipeadata.gov.br/Regional.
- Disseminação de tecnologias informatizadas de preservação de acervos bibliográficos e culturais implicando ações de:
 - Treinamento de pessoal de nível superior nas áreas de biblioteconomia, informática, história e ciências sociais em geral.
- Familiarização e treinamento de pessoal de nível médio com pesquisas científicas e tecnologias de ponta em informática.

Metodologia

- Identificação e seleção do acervo bibliográfico e estatístico de interesse.
- Catalogação e documentação bibliográfica do acervo selecionado.
- Digitalização de documentos:
 - Análise de documento: imagens individuais extraídas das páginas dos livros por meio de copiadora eletrônica (bookscanner) de alta resolução.
 - Limpeza para remover imperfeições tipográficas outros erros e centralização das imagens digitalizadas para exposição on-line. A pesquisa no texto continua sendo impossível.
- Conversão de páginas impressas em texto eletrônico pesquisável.
 - Por meio de programas computacionais de reconhecimento ótico de caracteres (OCR) as imagens digitalizadas podem ser convertidas em textos pesquisáveis. Sugere-se Abby Fine Reader.
- Transferência do documento digitalizado para servidor para que seu conteúdo seja passível de pesquisa na internet.
 - Identificação de estatísticas relevantes para transformação em base de dados computacionais.
 - Sistematização e compilação das bases de dados computacionais
 - O desenvolvimento do projeto até o momento utilizou para a digitalização o equipamento (bookscanner) Seleconta SC-7000 com tecnologia Fujitsu/Konica cujas especificações são: digitalização de documentos com tamanho grande (17x 23,25 polegadas); suporte de livros ajustável evitando tensões no dorso do livro; digitalização rápida 4,5 segundo/página para documentos de 8,5 x 11 polegadas; imagens nítidas e límpidas até 600-dpi de resolução;

compensação automática de curvatura nas páginas do livro mais grossos; detecção automática das bordas da área de digitalização; produto da digitalização diretamente gerado em pdf ou formato impresso

(<http://www.kmbs.konicaminolta.us/eprise/main/KMBS/Showroom/Models/0818302?Category=&PC> =>).

Essas características permitem a digitalização do conteúdo impresso de livros raros, em formato irregular ou estado precário de conservação com precisão e delicadeza, sem causar danos físicos aos originais. O equipamento é, portanto, crucial para a digitalização do acervo histórico da BMF/RJ composto, em sua grande maioria, por livros antigos e raros, em estado precário de conservação dos livros, em tamanhos e formatos irregulares, e com lombadas grossas, dentre outras características que dificultam o manuseio.

Na primeira fase do projeto testou-se com sucesso a velocidade, eficácia, confiabilidade e qualidade da SC-7000. De outubro de 2005 até abril de 2006 foram digitalizados 71 títulos ou obras (podendo incluir mais de um volume), totalizado aproximadamente 26.000 páginas. As imagens reproduzidas em pdf são de excelente qualidade como se comprova em www.nemesis.org.br/memoria onde encontra-se disponibilizado todo o material já digitalizado até o momento.

Deve-se notar, contudo, que o ritmo de andamento do projeto está aquém daquele que seria ideal para a digitalização em horizonte de tempo razoável -- ou seja, de 3 a 5 anos -- de um conjunto minimamente relevante de relatórios estatísticos do período histórico considerado. Tendo em vista dar maior agilidade ao projeto e torná-lo factível nesse horizonte de tempo, o cerne da proposta é a aquisição de outra unidade da SC-7000.

Para facilitar as tarefas de indexação e criação de arquivos pesquisáveis do projeto propõe-se a aquisição de algumas cópias do programa de reconhecimento óptico de caracteres (OCR) Abbyy Fine Reader 8.0 Professional Edition or Corporate (http://www.abbyy.com/finereader_ocr/) que permite converter e editar arquivos dispõe de funções de edição, paginação e geração de metadados. Esse programa servirá também para os trabalhos de pós-produção das publicações já digitalizadas até o momento.

Outra carência a se suprir no projeto diz respeito a estações de trabalho (workstation) para as tarefas de catalogação e pré e pós-produção das publicações digitalizados, bem como para desempenhar a função de servidores para arquivamento e de saída para internet.

Por fim, o desenvolvimento do projeto requer consultoria nas seguintes áreas:

- Ciências da informação (Biblioteconomia, Informática, etc.) para conceber e orientar a execução do projeto de indexação das publicações digitalizadas e selecionadas para compilação de bases de dados. Essa tarefa vem complementar as atividades de elaboração de metadados que se desenvolvem no Ipeadata.
- Informática e computação, em particular para o desenvolvimento de ferramentas de web; concepção e construção da homepage do projeto e dos programas de disponibilização das publicações eletrônicas que serão produzidas pelo projeto.

Indicadores de avaliação

A avaliação do projeto pode ser feita pelo acompanhamento contínuo do número de documentos digitalizados, indexados e disponibilizados na URL do projeto que, provisoriamente, encontra-se em www.nemesis.org.br/memoria. Esses indicadores poderão ser complementados pelo subsídios para as bases de dados estatísticas disponibilizadas em www.ipeadata.gov.br.

Instituições envolvidas e grau de envolvimento

O projeto está sendo desenvolvido conjuntamente pelo IPEA e a Biblioteca do Ministério da Fazenda. O coordenador do projeto é Eustáquio J. Reis do IPEA/RJ, que também coordena o NEMESIS. A execução do projeto está a cargo das equipes técnicas das bibliotecas do IPEA e do Ministério da Fazenda, ambas situadas no Rio de Janeiro, sendo que nessa última se situa fisicamente a execução do projeto. O projeto conta também com a colaboração da Coordenação de Informática do IPEA/RJ.

As tarefas da coordenação técnica incluem além da concepção e desenvolvimento do projeto, a identificação no acervo da BMF/RJ das publicações que merecem ser selecionadas para digitalização, indexação e disponibilizadas na internet. Em fase posterior, a seleção da

documentação estatística que deverá ser objeto de compilação em base de dados computacionais e incorporadas ao Ipeadata.

A coordenação da execução do projeto é de responsabilidade de Kátia de Oliveira e Vera Guilhon, bibliotecárias da BMF/RJ. Dentre as tarefas executivas destacam-se identificação e seleção do próprio acervo da BMF/RJ; a documentação e catalogação das publicações selecionadas para fins de digitalização; a digitalização, propriamente dita, com correção de erros e acertos de formatação, a estocagem em servidores e CDs.

O assessoramento do projeto em termos de informática e computação é feito por Jorge Morandi, coordenador de informática do IPEA/RJ. A seu cargo está toda a orientação em termos de infraestrutura física, programas de digitalização e de saída para a internet.

References

- Huttner, Joseph. Google Book Search Project. 7/12/2005.
<http://students.haverford.edu/jhuttner/Essays/Computers/GoogleBookSearch.htm>
- Greene, Kate. A Million Books. 8/5/2006.
http://www.technologyreview.com/read_article.aspx?id=16434&ch-infotech