

A Digitalização do Acervo Documental da Hemeroteca Municipal de Lisboa: uma primeira abordagem ao suporte electrónico, a partir do jornal *Os Ridículos* *

Por Álvaro Costa de Matos**

Mais do que uma intervenção teórica de fundo sobre a problemática da digitalização, é nossa intenção explicar, problematizar e apresentar aqui aquilo que, nesta área, está a ser desenvolvido e experimentado na Hemeroteca Municipal de Lisboa.

Para o efeito, estruturámos esta comunicação da seguinte maneira: a) num primeiro ponto esboçaremos, ainda que muito sumariamente, algumas considerações sobre o enquadramento teórico da digitalização; b) num segundo ponto ocupar-nos-emos da descrição técnica do jornal *Os Ridículos*, jornal que está a funcionar como uma espécie de “cobaia” desta nossa primeira abordagem ao suporte digital; c) num terceiro ponto abordaremos os problemas encontrados e a metodologia seguida na digitalização; d) num quarto e último ponto, apresentaremos alguns resultados do que tem vindo a ser feito.

II. A DIGITALIZAÇÃO

Porquê a digitalização? A Hemeroteca Municipal de Lisboa, como qualquer outra biblioteca, arquivo ou centro de documentação, debate-se diariamente com um problema gravíssimo que é a deterioração das suas colecções, nomeadamente das suas espécies bibliográficas mais

* Comunicação apresentada ao Colóquio “**Bibliotecas e Novas Tecnologias. Como combater a exclusão dos info-pobres?**”, no painel “Digitalização e difusão do património cultural” (Forum Lisboa, 11 a 13 de Outubro de 2000). Publicada em *Bibliotecas e Novas Tecnologias: como combater a exclusão dos info-pobres?* Actas, Lisboa, CML, 2001, pp. 77-85.

** Coordenador da Hemeroteca Municipal de Lisboa.

antigas, que, por isso mesmo, mais facilmente se degradam (lembro que esta biblioteca tem periódicos desde o início do século XVIII até à actualidade).¹ Debate-se, portanto, com uma questão que é de crucial importância: preservar uma herança/memória que é colectiva, isto é, garantir a sua sobrevivência para o futuro. Para atenuar este problema, e digo atenuar porque, nalguns casos, mais não podemos do que adiar uma morte inevitável, para atenuar este problema, dizia eu, começou-se, desde 1995, a transferir a informação para um suporte alternativo, o microfilme – transferência que naturalmente obedece a certos critérios, visando o longo prazo. O recurso a esta solução, que tem recebido recentemente um forte empurrão, quer através de maiores apoios financeiros quer através da cooperação com outros equipamentos culturais (é o caso, por exemplo, dos protocolos estabelecidos com os nossos arquivos municipais e com o Gabinete de Estudos Olisiponenses e das negociações que estão a ser feitas com a Biblioteca Nacional), tem-nos permitido salvar algumas colecções duma morte súbita, garantindo simultaneamente o acesso à informação.

No entanto, e como resultado de uma revolução tecnológica que não pára de nos surpreender, fomos obrigados a repensar esta política de preservação, no sentido de actualizá-la com as novas tecnologias postas à disposição dos profissionais da informação. A Hemeroteca não podia ficar indiferente a esta evolução. Com efeito, são muitos os produtos informáticos quer de leitura quer de reprodução de informação, e são muitas as vantagens que resultam da sua utilização, como iremos ver. Em função disto, entendeu-se então explorar e potenciar as virtudes destas novas tecnologias, nomeadamente do suporte digital, quer na preservação quer no acesso em rede à informação. Mas esta opção não

¹ Para saber mais sobre esta biblioteca, v. Matos, Álvaro Costa de, “Hemeroteca Municipal de Lisboa: a biblioteca dos «sete ofícios»”, in *Biblioteca, revista das bibliotecas municipais de Lisboa*, N.º 1-2 (Dez. 1998), pp. 88-90; *Ibidem*, N.º 9-10 (Jan. 2002), p. 72-73

colocou um ponto final na microfilmagem. O que estamos a tentar fazer é conciliar, sempre que considerarmos pertinente, as vantagens dos dois processos, pois entendemos que eles não são incompatíveis, antes complementado-se. Por outras palavras, pegamos em ferramentas novas, com extraordinárias capacidades, ferramentas que a tecnologia nos faculta e continuará a facultar, que, associadas ao microfilme, visam a satisfação dos mesmos objectivos: a preservação, a difusão e o acesso. Um exemplo: muito brevemente, parte da digitalização do nosso acervo documental será feita a partir de microfilmes já existentes. Esta estratégia não invalida o contrário, isto é, a digitalização directa dos documentos, como é o caso que aqui trazemos. Temos, portanto, duas abordagens que podem ser seguidas: uma baseada numa “solução híbrida”, digitalizando a partir do microfilme; outra baseada numa digitalização directa do documento original. Em ambas se satisfaz a preservação e o acesso em rede.

Bom, mas tratemos agora especificamente da digitalização, outra das alternativas que, fruto da tal revolução tecnológica, tem sido muito utilizada na transferência de suporte do conteúdo intelectual dos documentos. Por digitalização leia-se, numa forma muito simples, uma “fotografia electrónica” capturada directamente do documento original ou de um filme. As vantagens do suporte digital são muitas: a) extraordinária fidelidade ao original, como iremos ver depois na demonstração – na definição, na cor, nas manchas, na visualização das características do papel; b) manutenção da qualidade inicial independentemente do número de cópias; c) distribuição em rede com uso múltiplo.² Estas vantagens são as mais óbvias, mas temos outras não menos importantes: a) melhoria da comunicação da própria biblioteca; b) permissão da consulta à distância; c) salvaguarda dos

² CABRAL, Maria Luísa, “Microfilmagem & Digitalização: a coexistência pacífica”, in *Páginas a&b*, N.º 2 (1998), p. 42, 47.

originais. Quando associado a bases de dados bibliográficos ou sujeito à intervenção do OCR oferece-nos extraordinárias capacidades na difusão e no acesso à informação, facilitando e alargando esse acesso numa forma que podemos considerar revolucionária. É aqui que reside a sua enorme diferença em relação aos processos tradicionais, tornando a digitalização numa forma privilegiada para a comunicação e para a recuperação da informação.

Tem-se apontado algumas desvantagens da digitalização, como a rápida evolução tecnológica, as restrições legais (incluindo aqui as questões relacionadas com os direitos de autor), a falta de apoios institucionais, a dificuldade de armazenagem, a durabilidade da informação em suporte electrónico, para só falar das mais importantes.³ Mas estas desvantagens não constituem, em nosso entender, factores bloqueadores da digitalização. Pegando, por exemplo, na questão da rápida evolução tecnológica ou no suposto problema da armazenagem ou ainda na durabilidade, constatamos que a primeira não significa necessariamente incompatibilidade de soluções, que a segunda poderá ser contornada com um bom sistema de *jukeboxes* e que a terceira resolve-se com o disco óptico, que em condições adequadas garante uma durabilidade muito grande. O que mais nos preocupa é sem dúvida a eventual falta de apoio financeiro e a questão dos direitos de autor, mas que uma legislação liberal poderá contornar.

Foi a partir destes pressupostos, da ponderação dos prós e contras, que a Hemeroteca se lançou na digitalização do seu acervo documental, escolhendo com “laboratório de ensaio” para um trabalho que se pretende mais sistemático e generalizado uma colecção muito particular e muito peculiar, o jornal *Os Ridículos*. É sobre ele que muito rapidamente iremos falar.

³ Idem, *Ibidem*, p. 47.

II. O JORNAL OS *RIDÍCULOS*

Porquê *Os Ridículos*? Entendeu-se pegar neste jornal por vários motivos. Porque se trata de uma colecção única⁴, de um valor histórico e patrimonial extraordinário, reunindo ainda milhares de provas dos cortes de censura feitos durante a Primeira República e, sobretudo, durante a Ditadura Militar e o Estado Novo.⁵ Estamos, portanto, perante uma fonte de inquestionável valor histórico para o estudo do final da Monarquia Constitucional e daqueles três regimes políticos nas suas múltiplas vertentes: política, económica, social e cultural, incluindo aqui a moral, a estética, a religião e os costumes. O investigador poderá assim comparar estas provas com os textos finais, isto é, com os textos publicados pelo jornal e, desta forma, descortinar, através de um jornal humorístico e satírico, aquilo que a Primeira República, a Ditadura Militar e o Estado Novo entendiam não dever ser publicado, penetrando no âmago destes regimes políticos, definindo-os. Para estes regimes, com destaque para o Estado Novo, por razões que aqui não interessa aprofundar, tão importante era dizer à população o que pensar como apagar tudo sobre o que não devia pensar. O investigador poderá ainda perceber como é que a imprensa escrita lidou com este fenómeno, com as suas implicações logísticas, económicas e sociais, e que estratégias ela experimentou para atenuar ou contornar a acção do lápis azul. Poderá finalmente detectar as continuidades e as rupturas existentes entre os sistemas políticos tratados pelo jornal.

Por outro lado, preservam-se e salvaguardam-se os originais destas provas, algumas com mais de 70 anos de existência, deixando de haver um contacto directo entre a fonte, o documento, e o leitor. Mas,

⁴ Única, pois não temos conhecimento da existência nas bibliotecas e arquivos portugueses duma colecção d'*Os Ridículos* com as respectivas provas que foram submetidas á acção do lápis azul.

⁵ Ao todo, temos aproximadamente 22 mil provas de textos e desenhos censurados.

para além das provas, preserva-se ainda o próprio jornal, isto é, os números que foram publicados, muitos com mais de cem anos de existência, e que por causa disso e da constante consulta e reprodução a que foram e são sujeitos correm o sério risco de serem destruídos. O leitor passará agora a utilizar um suporte de leitura alternativo para a consulta destas provas e destes jornais, que neste caso será o computador, suporte que lhe permitirá um acesso rápido, “simpático” e muito eficiente, quer na leitura quer na reprodução ou importação do documento.

Por último, pegou-se nesta colecção porque se entendeu que esta poderia constituir um desafio interessante, pelos vários problemas que coloca, funcionando assim, como já se disse, como um “laboratório de ensaio” para uma acção/abordagem que se pretende mais estruturada e generalizada.

Esta colecção reúne quatro séries: uma primeira, publicada entre 3 de Outubro de 1895 e 1 de Junho de 1898, num total de 279 números; uma segunda, dada à estampa entre 12 de Abril de 1905 e 8 de Junho de 1963, num total de 5.972 números; uma terceira, de 2 de Setembro de 1967 a 31 de Julho de 1975, com 248 números; finalmente, a última, de 1 de Dezembro de 1978 a 4 de Outubro de 1984. Ao todo, temos 117 volumes, 81 dos quais com provas de censura. É nas últimas três séries que temos as provas dos cortes de censura, ainda que apresentadas de diferentes maneiras. Enquanto que nos jornais publicados durante a Primeira República (1910-1926) poucas são as provas que aparecem soltas, predominando os cortes na página inteira, nos jornais publicados durante a Ditadura Militar e o Estado Novo temos o inverso, com todas as provas de censura devidamente anexadas aos respectivos jornais. Por outro lado, deparámo-nos ainda com diversos tipos de provas: textos ou desenhos parcialmente censurados; textos e desenhos totalmente censurados e que, conseqüentemente, não foram publicados;

primeiras páginas censuradas. Ora, como devem calcular, esta diversidade de situações obrigou-nos a uma igual diversidade de soluções na metodologia utilizada para a digitalização destes documentos, tendo sempre como principal critério orientador a recuperação da informação acumulada. Tratemos agora dos problemas e da metodologia.

III. PROBLEMAS E SOLUÇÕES

Com já se disse, a apresentação das provas de censura não é uniforme ao longo da colecção. Temos também diversos tipos de provas. Mas os problemas não ficam por aqui: há ainda o tamanho das provas e das páginas do jornal (por exemplo, temos artigos censurados em folhas com um formato A2 ou A3), a qualidade do texto impresso e dos desenhos (alguns, poucos, dos documentos mais antigos têm que ser restaurados enquanto que noutros temos problemas de legibilidade da informação), a questão das cores e da sua manutenção (nas provas da Primeira República os cortes são a vermelho, nas relativas ao Estado Novo temos o famoso lápis azul), os manuscritos que acompanham os desenhos, que requerem, portanto, uma intervenção paleográfica. E depois temos o problema da descrição bibliográfica de toda esta documentação, da sua recuperação, da sua transferência para suporte digital, o que remete para a questão da resolução do texto e da imagem e do espaço que esta ocupa quer na rede da biblioteca, quer na rede mais vasta das bibliotecas municipais, quer ainda no serviço *on-line*. Temos ainda a questão da “fusão” do suporte digital com a descrição bibliográfica, a apresentação final e, por último, o armazenamento e a salvaguarda de toda esta informação. Como vemos, os problemas são muitos. Qual foi então a metodologia adoptada para os resolver e que neste momento está a ser usada?

Esta colecção oferece-nos dois núcleos documentais. A saber: as provas dos cortes de censura que serão tratadas com os respectivos textos finais e toda aquela informação que não foi sujeita à acção da censura e que constitui o grosso da colecção, isto é, aproximadamente 250.000 imagens. Em função desta divisão, adoptou-se, para o primeiro núcleo documental, a seguinte metodologia de trabalho:

a) numa primeira fase, procede-se à verificação de toda esta documentação (aqui detectar-se-á as eventuais faltas da colecção e avaliar-se-á o seu estado de conservação) e à sua desencadernação/desmanche, o que facilitará sobremaneira o tratamento digital, com implicações óbvias na qualidade da reprodução obtida;

b) numa segunda fase, procede-se, se necessário, ao restauro e remendagem quer das provas dos cortes de censura quer dos textos finais;

c) numa terceira fase, trata-se da descrição bibliográfica de toda esta informação, incluindo aqui naturalmente a sua indexação a partir da linguagem e dos *thesaurus* utilizados pela Hemeroteca. Optou-se aqui por catalogar também todas aquelas provas que foram integralmente censuradas e que, conseqüentemente, foram e são do desconhecimento do público, permitindo, desta maneira, o acesso desta informação, que é vital, ao leitor/investigador. Feito isto, poder-se-á recuperar a informação a partir das seguintes pesquisas: através de uma pesquisa simples, utilizando os índices criados, através da booleana, usando certos operadores, e, finalmente, através da pesquisa em texto livre, executando pesquisas em campos não indexados e, desta forma, satisfazendo qualquer necessidade de informação independentemente do conteúdo do índice de pesquisa;

d) numa quarta fase, digitaliza-se, isto é, procede-se à transferência do conteúdo intelectual dos documentos para o suporte

digital. Quando falo em documentos falo nas provas de censura e respectivos textos finais, o que permitirá ao leitor/investigador a tal comparação acima referida. Ocupamo-nos aqui, portanto, do tratamento da imagem e do texto e, como tal, importa definir os destinos e, em função destes, a qualidade da resolução da reprodução digital. Esta qualidade depende naturalmente dos programas de tratamento de imagem utilizados, do *scanner* que faz a captura, mas também do perfil do próprio operador, ou seja, da sua sensibilidade para o valor dos documentos que está a tratar. No nosso caso teremos uma matriz desta reprodução (com 400 *dpi's* de resolução) que ficará armazenada e salvaguardada em disco óptico, e, a partir desta, duas cópias: uma resolução intermédia (com 200/300 *dpi's*) que será utilizada na rede local (é durante a feitura desta resolução intermédia que entra o OCR, ferramenta que, como sabemos, faz o reconhecimento dos caracteres, isto é, lê a imagem/fotografia e transforma-a em texto, permitindo-nos dispor agora de um ficheiro em texto - o *txt* - que, depois de tratado pelos linguistas, será empregado na construção de um novo *thesaurus*, muito mais geral); e uma resolução para o serviço *on-line* (que terá 75 *dpi's*) - o programa que suporta este serviço e que disponibiliza igualmente a nossa base de dados bibliográficos, recebe as imagens em formato *tiff* ou *jpegg*, mais “pesados”, e transforma-as em formato *gifg*, mais “leve” e, conseqüentemente, de mais rápido acesso e impressão ou importação, ainda que o leitor/investigador possa escolher uma visualização da imagem em formato *tiff*, portanto, num formato com uma maior resolução, ainda que mais lento.

Uma nota sobre as vantagens da utilização do OCR. A utilização desta ferramenta abre-nos o acesso a um outro caminho na recuperação da informação. Isto é, permite-nos, com o auxílio do tal *thesaurus* geral que é construído, uma pesquisa integral ou difusa, leia-se uma pesquisa que oferece uma grande flexibilidade, facultando-nos

múltiplas aproximações à palavra que está a ser utilizada na pesquisa e, simultaneamente, permitindo-nos avaliar o comportamento do motor de pesquisa que lhe está associado;

e) depois do tratamento do texto e da imagem, segue-se a transferência desta informação para o nosso programa de gestão de informação documental, ou seja, procede-se ao “entrelaçamento” do suporte digital com os respectivos registo bibliográficos;

f) por último, temos o armazenamento de toda esta informação, que será feita, como já se disse, em disco óptico, e que incluirá a matriz, a versão intermédia e a versão disponibilizada na Internet. Porquê o disco óptico e não o cd ou o dvd? Porque o primeiro não só nos oferece uma durabilidade muito grande, em condições adequadas, como maiores garantias em termos de salvaguarda da informação.

O segundo núcleo documental, constituído, como vimos, pela maior parte da informação, isto é, pelos textos e desenhos que não foram objecto da intervenção da censura, segue a mesma metodologia de trabalho.

BIBLIOGRAFIA CONSULTADA

ASSOCIAÇÃO DE MUNICÍPIOS DO DISTRITO DE SETÚBAL – *As Tecnologias de Informação nas Bibliotecas Públicas*, Setúbal, AMDS, 1997;

AZEVEDO, Ana, “A biblioteca virtual ou precisando olhares sobre o futuro”, in *Páginas a&b*, N.º 1 (1997), pp. 95-111;

BETHENCOURT, Francisco, “As bibliotecas digitais”, in *Páginas a&b*, N.º 2 (1998), pp. 35-39;

BLÁZQUEZ, Jesús, “Aplicaciones y usos de Internet en las bibliotecas y centros de documentación, el caso del Centro Washington Irving”, in <http://usuarios.bitmailer.com/jblazquez/ponenci2.html> (consulta em Agosto de 2000);

IDEM, “Sobre las bibliotecas y Internet”, in <http://usuarios.bitmailer.com/jblazquez/press1997.html> (consulta em Setembro de 2000);

CABRAL, Maria Luísa, “Microfilmagem & Digitalização: a coexistência pacífica”, in *Páginas a&b*, N.º 2 (1998), pp. 41-52;

CAÇADOR, Fátima, “Uma ronda pelo Portugal Digital”, in *Público* (12 Fev. 2001), supl. *Computadores*, p. 2;

CALDEIRA, Alfredo, “Projecto de Digitalização do Arquivo Mário Soares”, in *Páginas a&b*, N.º 2 (1998), pp. 73-92;

DÖRR, Marianne, e WEBER, Hartmut, “Digitisation as method of preservation?”, in *Páginas a&b*, N.º 2 (1998), pp. 7-34;

FERREIRA, José Rincon, “A biblioteca digital”, in <http://www.usp.br/geral/infousp/rincon/rincon.htm> (consulta em Abril de 2000);

HARDEN, Sheila e HARDEN, Robert, “Net notions for librarians”, in <http://dspace.dial.pipex.com/town/square/ac940/nenotes.html> (consulta em Setembro de 2000);

Internet library for librarians: the most popular information resource site for librarians, in <http://www.itcompany.com/info retriever> (consulta em Julho de 2000);

MATOS, Álvaro Costa de, “Hemeroteca Municipal de Lisboa: a biblioteca dos «sete ofícios»”, in *Biblioteca, revista das bibliotecas municipais de Lisboa*, N.º 1-2 (Dez. 1998), pp. 88-90;

MINISTÉRIO DA CIÊNCIA E DA TECNOLOGIA/MISSÃO PARA A SOCIEDADE DA INFORMAÇÃO – *Livro Verde para a Sociedade da Informação em Portugal*, Lisboa, MCT/MSI, 1997;

MUÑOZ, Jordi Serrano, e ALARCÓN, Adoració Pérez, “Tecnologías de la información aplicadas a los servicios bibliotecarios”, in <http://www.ucm.es/info/multidoc/revista/num8/bib-vir.html> (consulta em Setembro de 2001);

RODRIGUES, Eloy, “Bibliotecas Virtuais e Cibertecários: o Futuro já começou”, in *Cadernos BAD*, N.º 3 (1995), pp. 23-34;

SILVA, Armando Jorge, “Bibliotecas e sistemas de informação electrónicos em rede”, in *Páginas a&b*, N.º 1 (1997), pp. 43-72.