

Digitalização de manuscritos históricos: a experiência da Casa Setecentista de Mariana¹

Alexandra Moreira

Mestre em ciência da informação pela Universidade Federal de Minas Gerais (2003); E-mail: xandramoreira@yahoo.com.br

Alcione de Paiva Oliveira

Doutor em informática pela Pontifícia Universidade Católica do Rio de Janeiro; E-mail: alcione@gmail.com

Fábio Faria Mendes

Doutor em ciência política pelo IUPERJ
E-mail: fmendes@ufv.br

Jonas Marçal de Queiroz

Doutor em história pela USP - prof. adjunto Dep. Artes e Humanidades da UFV/MG
E-mail: jonasqueiroz@yahoo.com

Vanuzza Braga

Mestranda em história, política e bens culturais no CPDOC/FGV
E-mail: vanuzabraga@hotmail.com

RESUMO

Este trabalho descreve a experiência do projeto de digitalização de manuscritos dos acervos dos Cartórios do 1º e 2º Ofícios do Arquivo Histórico da Casa Setecentista de Mariana-Iphan, apresentando as tecnologias e os recursos disponíveis atualmente para a tarefa de digitalização, suas características e deficiências. São mostrados também os critérios nas decisões adotadas e os metadados utilizados.

PALAVRAS-CHAVE

Perservação digital de manuscritos. Digitalização de manuscritos. Tecnologia de informação. Arquivos histórico da Casa Setecentista de Mariana/Iphan. Minas Gerais, Brasil.

Digitalization of historical manuscripts: experience of seventeenth century Mariana House

ABSTRACT

This article presents the technologies and resources currently available for the digitalization, their characteristics and weaknesses. It also outlines the criteria guiding the decisions taken during the project and the metadata incorporated into the project. This work describes the experience digitalization project of the manuscripts of the historical archive of the "Cartórios do 1º e 2º Ofícios Casa Setecentista de Mariana-Iphan".

KEYWORDS

Digital preservation of manuscripts. Digitalization of manuscripts. Knowledge management. Information technology. Historical Archives Seventeenth Century Mariana House/Iphan. Minas Gerais, Brazil.

INTRODUÇÃO

As informações registradas em papel estão limitadas a um espaço físico específico, o que dificulta o acesso a informações localizadas remotamente. O papel também está sujeito à deterioração pelo manuseio e pela ação de agentes ambientais. Com o intuito de facilitar o acesso aos documentos pelo público e contribuir para sua preservação, vários métodos têm sido testados. Entre as propostas mais recentes, existe a sugestão do uso da digitalização, entendida como um caminho para complementar soluções relacionadas tanto à preservação quanto ao acesso. Ela pode ser vista como uma forma de solução complementar de preservação porque torna possível restringir o acesso ao documento original, liberando para consulta apenas o material digitalizado.

Além disso, a possibilidade da transmissão das informações digitalizadas por meio de uma rede de computadores, como a Internet, permite maior acesso e agilidade às solicitações de conteúdo pelo público. Vale lembrar que, para o material digitalizado ser disponibilizado publicamente, é necessário obter as autorizações necessárias junto aos proprietários legais do acervo. Outro fator importante para a adoção da digitalização é o custo para montagem de um laboratório para este fim. Dependendo do material que será digitalizado, um laboratório de digitalização, em alguns casos, pode ser composto apenas por um microcomputador, atualizado dotado de gravador de mídia e de bom *scanner*, equipamentos que, dependendo da configuração, podem ser adquiridos por preços compatíveis com o orçamento de pequenas instituições culturais.

Apesar dessas vantagens, existem problemas com a adoção da digitalização em detrimento de outras técnicas de preservação de conteúdos, como a microfilmagem. Um microfilme, desde que adequadamente gerenciado, possui vida útil de até 500 anos. Já as mídias digitais como CDs e DVDs ficam em torno de algumas décadas. No entanto, este problema pode ser contornado em pouco tempo. Já existem notícias de lançamentos recentes de mídias digitais com expectativa de vida em torno de 300 anos (www.clubedohardware.com.br/noticias/1060).

Em relação à digitalização, o maior obstáculo é a obsolescência das mídias. Como a tecnologia evolui muito rapidamente,

¹ Trabalho financiado pela Fapemig

dispositivos capazes de ler determinada mídia permanecem um tempo relativamente curto no mercado e, em poucos anos, uma mídia que era a sensação no seu lançamento, torna-se impossível de ser lida por falta de dispositivo de leitura. Para contornar esta situação, é necessário o estabelecimento de rotinas permanentes de migração do suporte digital.

Este trabalho descreve a experiência do projeto de digitalização de manuscritos dos acervos dos Cartórios do 1º. e 2º. Ofícios do Arquivo Histórico da Casa Setecentista de Mariana-Iphan, coordenado pelo Departamento de Artes e Humanidades da Universidade Federal de Viçosa e financiado pela Fapemig, por meio do edital Fapemig no.009/2005: “USO DA TECNOLOGIA DIGITAL NO RESGATE DA IDENTIDADE HISTÓRICO-CULTURAL DE MINAS GERAIS”.

Espera-se que este relato ajude a nortear projetos semelhantes e que os critérios adotados possam auxiliar nas decisões difíceis que surgem em projetos desta natureza.

A próxima seção descreve o estado e a natureza dos documentos do acervo da Casa Setecentista de Mariana. Em seguida, são apresentados os conceitos da digitalização e suas principais técnicas. Os cuidados necessários para a manipulação adequada de mídias de armazenamento são discutidos na seção 4. Na seção 5, são apresentados os critérios de digitalização adotados para o acervo da Casa Setecentista e as razões da adoção dos critérios. Finalmente, a seção 6 apresenta as conclusões relativas a essa experiência de digitalização.

DOCUMENTOS DA CASA SETECENTISTA

Um conjunto de documentos representa a noção básica de um arquivo. Paes (2002) fornece quatro conceituações para o termo arquivo adotado pela comunidade arquivística, dentre os quais destacamos o que mais se relaciona com este trabalho:

Arquivo – designação genérica de um conjunto de documentos produzidos e recebidos por uma pessoa física ou jurídica, pública ou privada, caracterizado pela natureza orgânica de sua acumulação e conservado por essas pessoas ou por seus sucessores, para fins de prova ou informação (PAES, 2002, p. 24).

O arquivo histórico depositado na Casa Setecentista de Mariana-MG é um arquivo que guarda uma documentação de terceira idade ou permanente. Essas denominações, indicadas por Valette (1973), citado por Paes (2002, p. 21), estão relacionadas às funções que um arquivo desempenha, que são três. O arquivo de primeira idade ou corrente tem a função de atender a demanda da própria repartição ou instituição que os produziu. O arquivo da segunda idade ou intermediário tem a função de atender à uma demanda rara dos órgãos que os

gerou, é uma etapa transitória. Já o arquivo da terceira idade ou permanente, também mencionado por Belloto (2004) nas palavras de Paes é entendido da seguinte forma:

A função de um arquivo permanente é reunir, conservar, arranjar, descrever e facilitar a consulta dos documentos oficiais, de uso não-corrente, ou seja, concentrar sob sua custódia, conservar e tornar acessíveis documentos não-correntes, que possam tornar-se úteis para fins administrativos, pesquisas históricas e outros fins (PAES, 2002, p.121).

Nesse sentido, a casa tem como função disponibilizar informações contidas no acervo com a finalidade de servir de base do conhecimento histórico. O Arquivo Histórico da Casa Setecentista de Mariana possui valioso acervo documental, que pertenceu aos Cartórios do 1º. e 2º. Ofícios, abrigando rica documentação dos períodos Colonial, Imperial e da República, ou seja, de 1709 a 1956. Como a região de Mariana envolvia, nos períodos Colonial e Imperial, uma região substancialmente maior do que o atual território do município, a documentação abarca várias das cidades que compõem a região atualmente.

Este acervo documental, de cerca de 50 mil documentos ou, aproximadamente, 350 metros lineares, foi recolhido pelo antigo Dphan (Diretoria do Patrimônio Histórico e Artístico Nacional), atual Iphan, em 1948, e que o tombou. O Arquivo funcionou inicialmente em uma dependência do prédio da Câmara Municipal de Mariana, sendo transferido na década de 1960 para a atual Casa Setecentista. Trata-se de documentação de natureza jurídica de relevante valor histórico, que pode ser atestado a partir da análise de algumas características extrínsecas e intrínsecas do acervo.

Em relação às características extrínsecas, um aspecto é o estado do suporte dos documentos. No que se refere às características intrínsecas, o que importa é o conteúdo retratado na documentação e os aspectos lingüísticos da escrita da época. Em relação ao conteúdo do acervo, trata-se da documentação de um cartório de ofício do século XVIII, contendo assuntos relacionados a inventários, sesmarias, ações civis etc. Assim sendo, considera-se esta documentação histórica, uma vez que o conteúdo retrata eventos importantes do passado, permitindo entender parte da história do Brasil Colonial e do Brasil independente.

Na Casa Setecentista de Mariana-MG existem armazenados outros fundos, como o Fundo da Arquidiocese de Mariana e o Fundo da Câmara Municipal de Mariana. Toda esta documentação vem recebendo tratamento organizacional. A organização dos fundos armazenados na Casa Setecentista é apresentada no quadro 1:

Quadro 1

Organização da documentação depositada na Casa Setecentista de Mariana

Fundo	Grupo	Subgrupo
FÓRUM DE MARIANA	CARTÓRIOS DE OFÍCIO	AÇÃO CÍVEL CRIME EXECUÇÃO INVENTÁRIO JUSTIFICAÇÃO NOTIFICAÇÃO SESMARIA TESTAMENTARIA
	REGISTROS CARTORIAIS – LIVROS	ALFORRIAS– CARTAS AUDIÊNCIAS COMPRA E VENDA DE ESCRAVOS FIANÇAS DE CRIMINOSOS HIPOTECAS NOTAS PROCURAÇÕES QUERELAS REGISTRO DO TESTAMENTO
ARQUIDIOCESE DE MARIANA		
CÂMARA MUNICIPAL DE MARIANA	GUARDAMORIA	DATAS/ÁGUAS MINERAIS
	INCONFIDÊNCIA MINEIRA	

Fonte: elaborado pelos autores

A documentação citada recebeu esta identificação no ano 2000, quando a responsável pela Casa Setecentista, na ocasião, Vânia Carvalho Rôla Santos, identificou o acervo arquivístico, separando-o em fundos, séries etc. Além disso, estabeleceu a denominação jurídica dos termos utilizados para descrever o assunto contido nos manuscritos, identificou datas-limites para cada série e a dimensão total do acervo arquivístico em metros lineares.

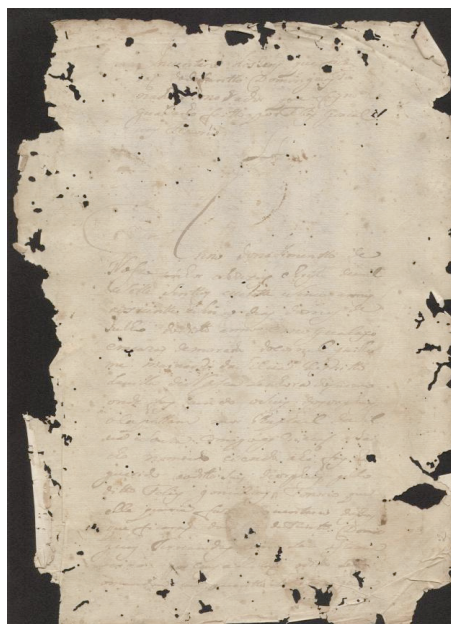
Embora a documentação esteja em condições satisfatórias de aclimação e acondicionamento, dentro dos parâmetros de conservação preventiva, cerca de 80% do arquivo do 2o. Cartório ainda não passou pelos tratamentos de higienização, acondicionamento e identificação necessários para a sua preservação a longo prazo. Alguns documentos encontram-se bastante deteriorados em função de ataques de pragas, de agentes ambientais e pelo manuseio. Na figura 1, pode-se observar o estado de um documento, sendo que grande parte dos documentos possuem estado de conservação semelhante.

De forma a iniciar uma mudança neste quadro, o projeto proposto tinha como metas higienizar, organizar, identificar e informatizar entre 10% e 15% do conjunto dos Inventários e Testamentos do Cartório do 2º Ofício da Casa Setecentista (cerca de 2000 autos) e digitalizar entre 10% a 15 % dos Inventários e Testamentos do Cartório do 1º Ofício da Casa

Setecentista (cerca de 450 autos). A digitalização tem a função de limitar o manuseio, diminuindo a deterioração dos originais.

Figura 1

Aspecto de um documento do acervo



CONCEITOS DE DIGITALIZAÇÃO

Nesta seção, são descritos os principais elementos relacionados com a digitalização. O objetivo não é ser completo, mas sim fornecer uma visão geral para servir de base para entender as decisões tomadas com respeito às técnicas de digitalização. Os conceitos discutidos nesta seção são formato de arquivo, profundidade do *bit*, resolução, compressão, alcance ou faixa dinâmica, tamanho do arquivo, dimensões do pixel e imagens digitais.

Formatos de arquivo

Os dados em um computador são armazenados em meio permanente em unidades chamadas de arquivos, que consistem em uma seqüência de *bits* (0 ou 1). A forma como esses *bits* são organizados é variada, implicando necessidade do uso de programas diferentes para sua leitura e gravação. Além disso, a organização do arquivo define algumas propriedades do arquivo, tais como número de cores, resolução, etc. Por convenção, é associada a cada tipo de arquivo uma extensão diferente, de forma a indicar o tipo de arquivo. A extensão é um sufixo de três ou quatro letras adicionado ao nome do arquivo após o caractere '.'. Assim sendo, usa-se a extensão "jpg" para os arquivos no formato JPEG, a extensão "gif" para os arquivos no formato *Graphical Interchange Format* etc.

Na escolha do formato de arquivo mais adequado, uma análise das necessidades da instituição e de seus usuários deve ser investigada. Para isso, é preciso verificar qual o formato mais adequado a cada ocasião e se existe a necessidade de gerar dois tipos de arquivo: um de acesso (para usuário) e outro de imagens originais para o armazenamento (denominado arquivo mestre ou *master*), ou os dois. Para essa escolha, é fundamental que certos quesitos sejam observados, como indicado por Kenney e Rieger (2003). Os quesitos que orientam a escolha do formato de arquivo de imagens originais são os seguintes:

- o arquivo a ser usado permite uma boa resolução;
- a profundidade de *bits* indicada permite a captura do documento em uma imagem com a qualidade estabelecida pela instituição;
- a capacidade de armazenamento de informações sobre cores é suficiente para o tipo de documentação que será digitalizada;
- os metadados documentam de forma precisa e clara os documentos;
- o formato é compatível para funcionar em várias plataformas computacionais;
- o formato não acarreta perda de informação;
- o formato quando descomprimido é amplamente suportado.-

Já no caso do formato do arquivo de acesso, espera-se que ele apresente determinadas características, tais como:

- possa ser comprimido, com pequenas perdas;
- seja lido, transmitido e apresentado de forma rápida;
- ocupe pouco espaço nos meios de armazenamento;
- seja amplamente suportado por ferramentas de apresentação e de manipulação para correção de falhas.

Imagens digitais

Uma imagem digital de um documento é o resultado da captura da imagem do mesmo na forma digital realizada por meio de uma câmera digital, ou por meio de um *scanner*. As imagens digitais podem representar fotografias, manuscritos, textos impressos etc. No contexto do projeto elas registram apenas manuscritos a partir do século XVII. Uma imagem digitalizada ou fotografada digitalmente é composta por uma matriz bidimensional de pontos. Estes pontos são denominados *pixels* (*picture element*). Os *pixels* assinalam um valor tonal que vai do branco ou preto a tons de cinza e colorido. Esses *pixels* são a representação do código binário, de zeros e uns, para ser interpretado por um programa de computador. Assim, com a interpretação pela máquina desse código de *bits*, é reproduzida uma versão analógica para visualização.

Profundidade do bit

A profundidade de *bits* é o número de *bits* que são usados para registrar as informações de um pixel. Quanto maior a quantidade de *bits*, maior será a variação de tons. Como um *bit* pode assumir apenas o valor 0 ou 1, com um *bit* é possível apenas distinguir dois tons, preto e branco (bitonal); com 8 *bits* é possível distinguir 256 tons (2^8), com 16 podemos distinguir 64.536 tons (2^{16}) e assim por diante. No entanto, quanto maior for o número de *bits* usado para armazenar um *pixel*, maior será o tamanho da imagem final. A figura 2 mostra o uso de dois *bits* para representar uma variação tonal na escala de cinza.

Figura 2

Com dois *bits* obtêm-se quatro variações

00	01	10	11

A profundidade de *bits* afeta também outro requisito da qualidade da captura de uma imagem: o alcance ou faixa dinâmica. O alcance ou faixa dinâmica relaciona-se com as variações tonais, isto é, escala de tons que varia do tom mais claro ao mais escuro. O alcance ou faixa dinâmica é afetado pela profundidade, em decorrência de cada *pixel* receber uma variação de tom. Quanto maior o número de *bits*, maior a possibilidade de se obter maior variação de tons.

Por essas razões, a profundidade de *bit* deve ser cuidadosamente estudada e definida em um processo de digitalização. Sua importância cresce em se tratando de manuscritos, uma vez

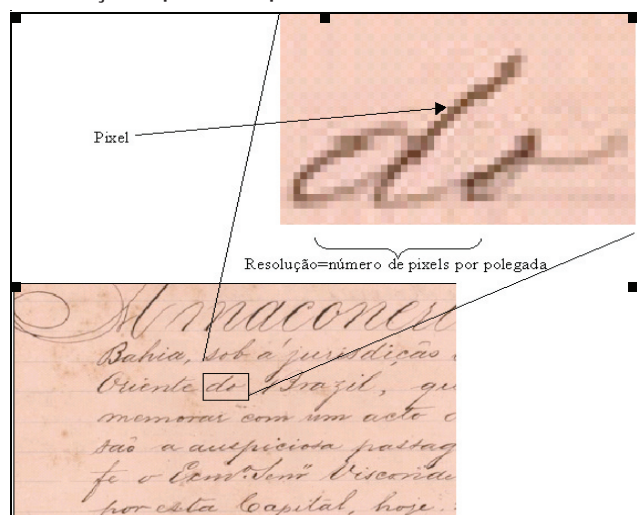
que, em função da imprecisão e variações naturais que ocorrem durante a escrita e pelo desgaste no tempo, alguns caracteres podem se tornar sutis, e apenas com a captura adequada de todas as nuances e tons de um manuscrito é possível mantê-lo legível. Contudo, não é preciso atribuir número de *bits* excessivo para aumentar o número de cores capturadas. O olho humano é capaz de distinguir por volta de 10 milhões de cores e, neste caso, 24 *bits* seriam mais do que suficientes para capturar todas as cores percebidas pelo ser humano.

Resolução

A resolução é outro indicador que determina a qualidade de um documento digital. A resolução é a capacidade de distinguir detalhes espaciais finos, expressa em *dpi* (*dots per inch*), e descreve o número de pontos (*pixels*) por polegadas. Quanto maior o número de *dpi* maior será a qualidade da imagem e maior será o tamanho do arquivo que a armazena. A resolução também pode ser expressa em *ppi*, (*pixel per inch*), quando se tratar de dispositivos como monitores e *scanner*, ficando o termo *dpi* reservado para impressoras, uma vez que *dots* está relacionado a pontos de tinta. A decisão sobre a resolução adotada para os arquivos mestres digitalizados deve ser realizada com cuidado para se obter um compromisso ótimo entre a qualidade e o espaço para armazenamento da imagem. A figura 3 mostra imagem com uma ampliação que permite identificar um *pixel* e o conceito de *dpi*.

Figura 3

Identificação do pixel e do dpi



COMPRESSÃO

Ao digitalizar um documento, o *scanner* varre suas informações ponto a ponto. Cada imagem possui grande quantidade de pontos, o que dificulta sua manipulação e armazenamento. Portanto, é necessário diminuir o tamanho do arquivo, comprimindo suas informações. Por meio de algoritmos de

compressão, é possível reduzir o tamanho do documento digital. Existem dois tipos de compressão: com perda e sem perda de informações. No uso de algoritmos de compressão sem perda, a redução do tamanho do arquivo não implica a perda de nenhum detalhe de informação, sendo possível obter a imagem original por meio de um algoritmo reverso. No entanto, estes algoritmos podem não obter uma taxa de compressão (relação entre o tamanho do arquivo original e o comprimido) muito alta. Os algoritmos de compressão com perda podem obter taxa de compressão maior, mas acarretam perda de informação, não sendo possível recuperar o arquivo original posteriormente. Dentre os algoritmos de compressão sem perda, podem-se citar o LZW, o JBIG e o ITU-T.6. Dentre os algoritmos de compressão com perda, podem-se citar o JPEG e o ImagePac. Muitos formatos de imagens já armazenam a imagem em modo comprimido, como o JPEG.

Alcance ou faixa dinâmica

O alcance ou faixa dinâmica é a percepção da diferença tonal, isto é, a diferença de tom entre as partes mais claras e as partes mais escuras. Quanto maior a faixa dinâmica, maior a capacidade de representar matizes. Esta propriedade aparece mais evidente em figuras ou fotos. Nelas, a exibição de tons que variam ligeiramente, como detalhes de sombra e de luz, podem ficar difíceis de perceber com o uso de uma faixa dinâmica limitada.

Tamanho do arquivo

A unidade de medida para calcular o tamanho de um arquivo é o byte. Cada *byte* é formado por uma seqüência de oito *bits*. Para o cálculo do tamanho de um arquivo a ser digitalizado em *bytes*, toma-se por base a medida do documento original e utiliza-se a seguinte fórmula:

$$\text{tamanho do arquivo} = (\text{altura} \times \text{largura} \times \text{profundidade de bits} \times \text{dpi}^2) / 8$$

Fórmula 1 – Tamanho do arquivo em *bytes*.

Se as dimensões em *pixels* são conhecidas, a fórmula obedece ao seguinte esquema:

$$\text{tamanho do arquivo} = (\text{altura em pixels} \times \text{largura em pixels} \times \text{profundidade de bits}) / 8$$

Fórmula 2 – Tamanho do arquivo em *bytes*.

Qualidade

Pode-se buscar melhor imagem por meio dos recursos do *software* de digitalização e ferramentas de edição de imagem. Eles permitem eliminar ou diminuir efeitos negativos da digitalização. Alguns exemplos seriam: pontos no fundo da imagem (*despeckling*), eliminar distorções na orientação da

imagem (*deskewing*) etc. Além disso, estes programas permitem aumentar a nitidez (*sharpening*), utilizar filtros personalizados e ajustar a profundidade. Mas sua aplicação deve ser realizada com cuidado, de preferência aplicados a apenas em imagens de acesso. Isso porque, dentro dos estudos arquivísticos, a manipulação da informação original pode levantar questões sobre a fidelidade e a autenticidade de um documento.

Cores

A cor é um aspecto difícil de equacionar na tarefa de digitalização. No caso deste projeto, como serão digitalizados manuscritos do século XVIII, o uso de cores nas imagens-mestre é justificado, uma vez que as cores podem registrar informações importantes para o pesquisador. Manuscritos não possuem uma escrita uniforme e a pressão e a quantidade de tinta usada varia a cada caractere registrado pelo escritor. Vários manuais de digitalização recomendam seu uso em se tratando de manuscritos, como o guia de digitalização da Universidade da Geórgia (2001) e os padrões de digitalização de material de herança cultural da Biblioteca do Estado de Maryland (2005, p.10). No caso de imagens para simples consulta ou transmissão em uma rede de computadores, pode-se adotar imagens em uma escala de tons de cinza. As cores podem ser influenciadas pelos seguintes fatores: nível de iluminação no momento da captura, profundidade de *bits*, capacidades do sistema de digitalização e a representação matemática da informação de cor no momento em que a imagem passa pela cadeia de digitalização de um espaço de cor a outro.

Rendimento e controle de qualidade do sistema

O rendimento de um sistema varia de acordo com requisitos que ele possui. Este rendimento pode ser determinado por meio de provas nas quais se verificam a resolução, a reprodução de tons, a qualidade das cores, o ruído e os artefatos. Além disso, equipamentos com as mesmas características podem, com o tempo, ter seu rendimento afetado. Portanto, é preciso estabelecer o nível de qualidade desejada na reprodução, mantendo atenção para obter a mesma qualidade em todas as reproduções. É necessário, também, realizar testes periódicos com o equipamento, com o objetivo de aferir o seu rendimento. Um teste que pode ser aplicado é, após determinado número de digitalizações, redigitalizar o primeiro documento para verificar se a nova digitalização tem a mesma qualidade da primeira.

CUIDADOS NA MANIPULAÇÃO DA MÍDIA DE ARMAZENAMENTO

Discos ópticos, mais especificamente CDs e DVDs, necessitam de cuidados na guarda e no manuseio. Todo o trabalho de digitalização dos conteúdos documentais poderá estar armazenado nestas mídias, embora sejam suportes apenas de disseminação e guarda provisória, não sendo, até o momento,

suportes que possam ser considerados de preservação. Por esse motivo, recomendações e dicas básicas no uso destes discos são importantes para que não se perca todo o trabalho executado com a perda de uma mídia. Assim, esta seção tem como objetivo orientar os técnicos no uso, no armazenamento e na conservação dos dados arquivados nos discos ópticos, cujas recomendações se baseiam no guia *Care and Handling of CDs and DVDs – A Guide for Librarians and Archivists*, elaborado por BYERS (2003). As recomendações indicadas a seguir servem tanto para CDs como para DVDs.

Recomendações relacionadas ao manuseio:

- verifique visualmente a superfície do disco. Se observar alguma mancha ou poeira, limpe-o antes de usá-lo;
- segure o disco pelas bordas ou pelo orifício central, nunca na superfície.
- não envergue nem dobre o disco;
- abra um pacote de mídias graváveis apenas quando estiver pronto para gravar o disco;
- escreva no disco apenas do lado da etiqueta, usando marcador ou caneta adequada para CDs. Nunca no lado espelhado, no qual o laser lê;
- não escreva no disco usando marcadores cuja tinta contenha solventes;
- não use lápis, caneta ou marcadores de ponta fina para escrever no disco;
- não utilize etiquetas adesivas e nem tente removê-las ou reposicioná-las se já estiverem coladas;
- não deixe discos soltos ou jogados em cima de mesas ou outras superfícies, guarde-os na capa imediatamente após o uso.

Recomendações relacionadas à limpeza:

- se for necessário, lavar. Ensaboe as mãos com sabão neutro ou sabonete e, apenas com a espuma, lave o disco com suaves movimentos em linha reta, partindo do centro para a borda. Lembre-se, nunca realize movimentos circulares. Enxágüe o disco segurando-o pelas bordas. Para secá-lo, utilize uma toalha de papel com movimentos em linha reta, partindo do centro para a borda;
- para remover sujeira, poeira, marcas de dedos, manchas e líquidos utilize um pano limpo de algodão, realizando movimento em linha reta do centro do disco para a borda;
- em caso de sujeira teimosa, use detergente específico para limpeza de CD/DVD ou mesmo álcool isopropílico ou metanol.

Recomendações relacionadas à guarda:

- mantenha discos em suas capas para minimizar os efeitos negativos de mudanças ambientais;
- armazene os discos na vertical, como se fossem livros numa estante, e em capas específicas para CDs e DVDs que não contenham PVC em sua composição. Evite guardá-los na horizontal por longo tempo (anos);

- armazene os discos em local fresco, seco e escuro e onde o ar seja limpo. Evite a umidade ou o calor extremos;
- não exponha os discos à luz do sol prolongada ou outras fontes de luz ultravioleta.

Uma dica muito interessante recomendada por BYERS (2003) refere-se à maior durabilidade dos dados gravados no disco. Quando desejamos que dados sejam preservados por longo prazo, é recomendado o uso de CDs e DVDs que tenham a camada refletiva de metal dourado. Ainda neste sentido, os CDs e os DVDs graváveis devem ser mantidos em uma temperatura entre 4° e 20°C e umidade relativa entre os 20% e os 50%. Os números aceitáveis são 18°C de temperatura e 40% de umidade relativa. Quanto mais frio, melhor para armazenamentos de longa duração. As dicas de conservação aqui mencionados não visam a indicar que os CDs e DVDs podem ser usados como mídia adequada à preservação de longo prazo, uma vez que isto não é aceito internacionalmente. É importante lembrar que CDs e DVDs graváveis podem ter duas finalidades em um arquivo: serão consultados freqüentemente ou serão mantidos guardados como cópia de reserva.

No primeiro caso, devem ser guardados em local de fácil acesso, sendo cada disco verificado visualmente após cada uso, visando a identificar possíveis defeitos ou contaminações. Já os discos para arquivo devem ser mantidos em ambiente controlado, longe das cópias usadas para consulta.

Com relação à longevidade das mídias, Byers (2003) aponta os seguintes valores estimados, para condições rigorosamente controladas de manuseio e armazenamento físico:

Quadro 2

Duração estimada para as mídias óticas

Tipologia dos Discos Ópticos	Durabilidade
CD-R, DVD-R, DVD+R	entre 100 e 200 anos.
CD-RW, DVD-RW, DVD+RW, DVD-RAM	≅ 25 anos.
CD-ROM, DVD-ROM	entre 20 e 100 anos.

PADRÕES ADOTADOS

Testes de tempo de digitalização

Para que fossem definidas a resolução e a profundidade de *bits* mais adequada ao projeto, utilizaram-se as recomendações de instituições ligadas à arquivística e à biblioteconomia. Alguns testes foram realizados para medir a qualidade da imagem, assim como tempo que seria ocupado pelo processo de digitalização. Os manuais de digitalização consultados (NINCH, 2002; KU *Digital Initiatives*, 2003; MDCHP, 2005) não recomendam uma resolução menor que 300 dpi para arquivos-mestre resultantes da digitalização de documentos manuscritos, sendo que consideram

o valor ideal 600 dpi. No que se refere à profundidade de *bits*, o recomendável é utilizar 24 *bits*, e, dependendo do caso, 8 *bits* (escala de cinza) podem ser aceitáveis. No caso do projeto, era preciso enquadrar as recomendações dos manuais às limitações de pessoal, físicas e orçamentárias, sem comprometer o objetivo da digitalização. Uma das principais limitações é o fato de possuímos apenas um *scanner* para o total de 30 mil varreduras. Nesse caso o fator tempo é um critério importante a ser analisado. Verificou-se que a digitalização em tons de cinza (8 *bits*) consumia 29 segundos, enquanto no uso de cores (24 *bits*) o tempo gasto era por volta de 1 minuto. Esse período é relativo apenas ao tempo de o *scanner* realizar a varredura no documento. É preciso contabilizar também o tempo gasto na colocação do papel e do armazenamento da imagem. O primeiro tempo independe da escolha da profundidade de *bits* e o segundo não apresentou valores significativos. Neste caso, o tempo do *scanner* não era o que mais pesava e, portanto não iria gerar alteração tão significativa no resultado final.

Além disso, estávamos seguindo regras de digitalização, com o objetivo de preservação e manutenção do tom “amarelado” que predomina nos manuscritos, como forma de assegurar princípios arquivísticos de fidelidade do documento. Outro critério adotado com vista à preservação foi a escolha da resolução de 300 dpi, que é a mínima recomendada para documentos manuscritos pelas instituições consultadas. Uma resolução maior resultaria em grande aumento do espaço ocupado, uma vez que o crescimento do tamanho do arquivo é quadrático em relação ao número de pixels. A tabela 1 mostra os resultados colhidos nos testes.

Tabela 1

Testes de digitalização

	resolução	cor	tempo	Tamanho do arquivo
Folha 1	300	24 bits	1 min.	19 MB. LZW
Folha 1	300	8 bits (cinza)	29 seg.	6 MB. LZW
Folha 2	300	24 bits	49 seg.	18,96 MB. LZW
*Folhas 3 e 4	300	24 bits	1 min. 8 seg.	36,57 MB. LZW
Folha 5	300	24 bits	51 seg.	18,47 MB. LZW
Folha 6	300	24 bits	50 seg.	19,20 MB. LZW
Folha 7	300	24 bits	51 seg.	18,51 MB. LZW

* As folhas 3 e 4 foram digitalizadas de uma única vez. Folha aberta.

Formatos de arquivo aplicados no projeto

Estabeleceu-se que seriam usados três formatos de arquivo no projeto. Para o armazenamento do arquivo-mestre, onde existe a necessidade de maior resolução e profundidade de *bits*, usou-se o formato tiff (*Tagged Image File Format*), com compressão de dados sem perda. Para a apresentação na Internet, onde existe a necessidade de um arquivo menor para facilitar a

transmissão pela rede, foram definidos os formatos JPEG (*Joint Photographic Experts Group*), com compressão de dados, e o formato PDF (*Portable Document Format*). Cada formato de arquivo será analisado a seguir com base nas informações fornecidas por Kenney e Rieger (2003, p.17), detalhando-se os critérios utilizados para a sua adoção no projeto.

O formato de arquivo TIFF utiliza como extensão “.tif” ou “.tiff”. Sua profundidade de *bits* vai de bitonal (1 *bit*) até um máximo de 64 *bits*. Pode ser usado sem compressão ou comprimido sem perdas (algoritmos ITU-T.6, LZW, etc), ou comprimido com perdas (algoritmo JPEG). É um padrão de fato. Possui um suporte limitado na Internet, sendo acessado por aplicações externas ao navegador. Seu suporte a metadados é limitado a um conjunto básico de rótulos etiquetados. O formato tiff, por não apresentar perdas e por possuir ampla escala de profundidade de *bits*, será usado no armazenamento do arquivo-mestre do documento digitalizado com intuito de conservação.

Outro formato selecionado é o JPEG (*Joint Photographic Expert Group*) /JFIF (*JPEG File Interchange Format*) com extensão “.jpeg”, “.jpg”, “.jif”, “.jif”. Sua profundidade vai da escala de cinza de 8 *bits*, ao colorido com 24 *bits*. Sua compressão se apresenta com perda (algoritmo JPEG) e sem perda. É padronizado pela norma ISO 10918-1/2. Possui suporte nativo na Internet por meio de todos os navegadores. O suporte de metadados é feito por meio de campo de texto livre para comentários. Pelo fato de possuir suporte em todos os navegadores e poder ser usado com compressão, este formato e o formato PDF foram selecionados para a exibição dos documentos na Internet.

O formato PDF usa como padrão a extensão “.pdf”. Sua profundidade de *bits* vai da escala de cinza de 4 *bits* até o máximo de 64 *bits*. Pode ser usado sem compressão ou comprimido sem perdas (algoritmos ITU-T.6, LZW), ou comprimido com perdas (algoritmo JPEG). É um padrão de fato, possuindo suporte limitado na Internet e sendo acessado por aplicações externas ao navegador. Seu suporte a metadados é limitado a um conjunto básico de rótulos etiquetados.

O quadro 3 sintetiza os formatos de arquivos e suas respectivas resoluções, adotados no projeto.

Quadro 3

Formatos de arquivos adotados no projeto

	Resolução	Profundidade de bits	Compressão	Formato do arquivo
Arquivo-mestre	300 dpi	24 bits	LZW sem perdas	TIFF
Arquivo para acesso em rede	de 75 a 125 dpi	8 bits	JPEG com perdas	JPEG e PDF

Regras de nomeação de arquivo e estrutura de diretórios

As imagens digitalizadas devem ser armazenadas em uma organização similar à organização física dos documentos originais. O exemplo a seguir mostra a estrutura de diretório adotada para o arquivamento da imagem digital:

Exemplo:

```

Diretório Raiz
  Projeto_digitalização_ufv
    documentos
      1º ofício
        Auto 92
          folha 1.tiff
          folha 2.tiff
          ...
  
```

Os manuscritos não são digitalizados em aberto, sendo mantida a seqüência formato livro (folha 1, folha 2, ...). Assim, mantém-se padrão e seqüência do texto para a posterior leitura pelo usuário.

Mídia de gravação

Na escolha de mídias para gravação das imagens digitalizadas, existia a opção do uso CDs ou os DVDs. Ambas as mídias são apropriadas para receber arquivos de grandes dimensões, como imagens. Após alguns cálculos sobre a quantidade de arquivos que seriam gerados, optou-se pelo uso de DVDs. Isso porque os CDs têm capacidade máxima de 800 megabytes, enquanto a capacidade típica de um DVD é de 4,7 gigabytes, ou quase 6 CDs. Se considerarmos o uso de um DVD de camada dupla, sua capacidade de armazenamento passa para 8 GB, o equivalente a aproximadamente 10 CDs. Como cada imagem digitalizada ocupa entorno de 20 megabytes, um CD poderia armazenar no máximo 40 imagens, o que é muito pouco para uma estimativa de 30 mil imagens que seriam geradas na digitalização dos documentos do 1º ofício. Neste caso, seriam necessários 750 CDs, se realizada apenas uma cópia, ou 1500, no caso do projeto, uma vez que foi adotado o critério de criação de duas cópias (uma para uso e outra para segurança). Portanto, para esta tarefa, pode-se utilizar 150 DVDs de dupla camada.

Rotina de digitalização

Este projeto tem por objetivo final a digitalização, mas algumas tarefas, como a higienização, devem ser realizadas antes do início da digitalização. Após a digitalização é necessário definir os passos para o armazenamento e rotinas para a atualização da mídia, de modo a evitar perdas por deterioração ou obsolescência tecnológica. A soma dos passos de cada etapa forma um processo de digitalização, armazenamento e manutenção. Este processo está registrado no *Manual de Digitalização para Manuscritos Depositados na Casa Setecentista de Mariana-MG* (MOREIRA, 2006).

Apresenta-se, a seguir, resumo dos passos que formam o processo estabelecido pelo projeto para a Casa Setecentista:

- **tratamento para eliminação de pragas** – técnica de congelamento dos documentos para a eliminação de pragas;
- **higienização dos documentos** – uso de pincéis em mesas apropriadas;
- **digitalização** – escaneamento dos documentos para a geração do arquivo mestre segundo os padrões estabelecidos (300dpi, 24bits);
- **gravação das imagens mestre** – formato selecionado (tiff com compressão sem perdas), obedecendo à regra de nomeação e a estrutura de diretórios previamente definidas;
- **registro** – preenchimento na planilha-metadados das informações do documento;
- **backup** – Quando o espaço ocupado pelos arquivos digitalizados se aproximar de 4,7 gigabytes, deve ser realizado o *backup* das imagens, gravando em mídia digital (correntemente DVDs). Devem ser feitas duas cópias, uma para manipulação e outra como reserva armazenada em local diferente da outra versão, preferencialmente em outro prédio.
- **armazenamento** – mídia em invólucro rígido e opaco. Uma etiqueta deve ser colocada no lado externo do invólucro com as seguintes informações: validade do suporte, identificação dos documentos;
- **registro em fichas** – registro em fichas de papel da existência do suporte em mídia digital.

Metadados

A organização e descrição de documentos arquivísticos é um processo que visa a estabelecer a proveniência e facilitar o acesso. DOLLAR (1994), em artigo sobre impacto das tecnologias em relação aos princípios e práticas de arquivo, indica a importância da organização física e intelectual:

A organização física nos dá comumente uma seqüência de grupos de documentos que se reflete num registro de localização de prateleira. A organização intelectual – apresentação de conexões lógicas e de relações entre grupos de documentos – é de longe mais significativa (DOLLAR, 1994, p.15).

A organização intelectual é estabelecida com a definição dos metadados. Os metadados são definidos como dados sobre dados. Isto porque os metadados descrevem as informações contidas em um documento. Essa descrição por meio de etiquetas permite localizar com maior facilidade alguns itens do documento. Por apresentar etiquetas para o conteúdo (produtor, título, datas e descrição sucinta do conteúdo), facilita encontrar o que se deseja com maior agilidade. Os metadados, organizados em uma estrutura, funcionam como

índices, facilitando as buscas tanto de clientes, quanto de organizadores. Os produtores/organizadores são beneficiados também pela descrição de rotinas e critérios fundamentais na organização.

Os metadados podem ser classificados em descritivos, estruturais e administrativos. Para o estabelecimento dos metadados descritivos a serem usados no projeto, decidiu-se selecionar um subconjunto dos que são sugeridos na International Standard Archival Description – ISAD(G) (2000). A ISAD(G) é uma norma estabelecida pelo Conselho Internacional de Arquivos que visa a descrever um conteúdo informacional aplicando-se um sistema de níveis. Esse sistema de níveis, denominado estrutura multiníveis, permite uma descrição detalhada e racional, facilitando o entendimento e o acesso mútuo. Além disso, respeita os princípios da proveniência e da organicidade. A norma ISAD (G) não é uma norma prescritiva, mas sim diretiva, e por este motivo a usamos como base para a escolha dos metadados que acreditamos serem adequados ao projeto.

Os metadados aplicados aos documentos arquivados na Casa Setecentista de Mariana- MG são de dois tipos: os de acesso público e os de acesso restrito. Os metadados de acesso restrito, tais como procedência e condições de reprodução, são utilizados pelos funcionários ou técnicos nas atividades de descrição do conteúdo da documentação, enquanto que os de acesso público, tais como idioma, título, âmbito e conteúdo, não possuem restrição de acesso. Outra classificação para os metadados proposta pela ISAD(G) é separação em áreas que agrupam os elementos de descrição comuns. No caso dos elementos adotados no projeto, a classificação fica da seguinte forma:

– **Na área de identificação** – indicada para o preenchimento de informações essenciais para identificar a unidade de descrição. Os campos relacionados a esta área são:

- código de referência;
- título;
- datas;
- nível de descrição;
- dimensão e suporte.

– **Na área de contextualização** – indicada para o preenchimento de informações sobre origem e custódia da unidade de descrição. Os campos relacionados e selecionados desta área são:

- nome dos produtores;
- procedência.

– **Na área de conteúdo e estrutura** – indicada para o preenchimento de dados sobre o assunto e organização da unidade de descrição. Os campos relacionados e selecionados são:

- âmbito e conteúdo;
- avaliação, eliminação e temporalidade;
- incorporações.

– Na **área de condições de acesso e uso** – indicada para o preenchimento de dados sobre a acessibilidade da unidade de descrição. Os campos relacionados e selecionados são:

- condições de reprodução;
- idioma.

– Na **área de controle da descrição** – para o preenchimento de dados sobre como, quando e por quem a descrição da unidade de descrição foi elaborada. O campo relacionado e selecionado é:

- nota do arquivista.

As áreas de fontes relacionadas e notas não foram selecionadas, mas nada impede que futuramente sejam incluídas.

Os metadados devem ser digitalizados em uma planilha para, posteriormente, serem inseridos em um banco de dados. O banco de dados será utilizado pelo sistema de consulta e por programas que converterão os metadados para o formato EAD¹ (*Encoded Archival Description*), padrão de marcação para descrição de documentos que é apoiado pela Sociedade Americana de Arquivistas e pela Biblioteca do Congresso.

CONCLUSÕES

Este artigo relata a experiência de digitalização dos documentos históricos da Casa Setecentista de Mariana. A microfilmagem, quando executada de acordo com as normas internacionais de preservação, permanece como o procedimento adequado para a preservação de documentos, em função de suas características de fidelidade e durabilidade. A digitalização isoladamente ainda não é vista como uma solução para preservação de documentos. Existem várias questões que devem ser respondidas antes que isto aconteça, por exemplo, como garantir que o documento digitalizado não sofra alterações no conteúdo, ou como garantir que o documento migre periodicamente de mídias para impedir que seu conteúdo se torne inacessível devido a obsolescência dos equipamentos.

As respostas a essas questões começam a surgir, tais como o uso de assinaturas digitais e o estabelecimento de rotinas de migração de mídias. No entanto, enquanto essas soluções não se consolidam, a digitalização já pode auxiliar indiretamente na preservação, uma vez que o acesso aos documentos originais pode ser limitado. Além disso, a digitalização permite uma ampliação do acesso à informação por meio da distribuição de cópias e a transmissão pela Internet.

A experiência da Casa Setecentista revela que estes benefícios podem ser obtidos com recursos relativamente limitados. No entanto, deve-se procurar obter equipamentos de procedência

confiável, com boa capacidade de memória e recursos para manipulação de imagens. O *scanner* deve possuir uma área de digitalização compatível com os documentos que serão digitalizados. É importante o estabelecimento de uma rotina de digitalização bem definida, assim como regras para manipulação das mídias e para migração dos arquivos digitalizados para mídias mais modernas. Finalmente, o estabelecimento dos metadados é uma etapa crítica no processo e deve ser feita com cuidado, uma vez que deles depende o acesso ágil aos documentos.

**Artigo recebido em 24/08/2007
e aceito para publicação em 16/05/2008**

REFERÊNCIAS

- BELLOTTO, Heloísa Liberali. *Arquivos permanentes: tratamento documental*. 2. ed. rev. e ampl. Rio de Janeiro: Ed. FGV, 2004. 320 p.
- BYERS, F. R. *Care and handling of CDs and DVDs: a guide for librarians and archivists*. Washington: NIST Special Publication, 2003. 48 p.
- DIGITAL LIBRARY OF GEORGIA. *Digital library of Georgia digitization guide [on line]*. 2001. Version 1.0. Disponível em: <<http://dlg.galileo.usg.edu/guide.html>>. Acesso em: 2008.
- DOLLAR, Charles M. O impacto das tecnologias de informação sobre princípios e práticas de arquivos: algumas considerações. *Acervo*, Rio de Janeiro, v. 7, n. 12, p. 3-38, jan. 1994.
- INTERNATIONAL COUNCIL ON ARCHIVES. ISAD(G): General International Standard Archival. 2nd ed. Ottawa: CIA/CDS, 2000. Disponível em: <[http://www.icacds.org.uk/eng/ISAD\(G\).pdf](http://www.icacds.org.uk/eng/ISAD(G).pdf)>. Acesso em: set. 2007.
- KENNEY, Anne R., RIEGER, Oya Y. *Moving theory into practice: digital imaging tutorial*. Ithaca, NY: Cornell University Library/Research Department, 2000-2003. 138 p. Disponível em: <<http://www.library.cornell.edu/preservation/tutorial/contents.html>>. Acesso em: set. 2004.
- KU DIGITAL INITIATIVES. *Recommended standards and best practices for digital projects*. 2003. 8 p. Disponível em: <<http://kudiglib.ku.edu/docs/DLI%20Standards.doc>>. Acesso em: 21 jun. 2004.
- MARYLAND DIGITAL CULTURAL HERITAGE PROGRAM (MDCHP). *Digitization standards for Maryland Digital Cultural Heritage Program Collections*. Baltimore, Maryland: Maryland Digital Cultural Heritage Program, 2005. Disponível em: <http://www.mdch.org/uploadedFiles/mdch/MDCH_Digitization_standards_2005.pdf>. Acesso em: 2006.
- MOREIRA, A. *Manual de digitalização para manuscritos depositados na Casa Setecentista de Mariana-MG*. Viçosa: [s.n.] 2006. 46 p. Relatório.
- NATIONAL INITIATIVE FOR A NETWORKED CULTURAL HERITAGE (NINCH); HUMANITIES ADVANCED TECHNOLOGY AND INFORMATION INSTITUTE (HATII). *The NINCH Guide to good practice in the digital representation and management of cultural heritage materials*. Scotland, UK: University of Glasgow, 2002. Disponível em: <<http://www.nyu.edu/its/humanities/ninchguide/XIV/>>. Acesso em: 2006.
- PAES, Marlena Leite. *Arquivo: teoria e prática*. 3.ed. rev. e ampl. Rio de Janeiro: Ed. FGV, 2002. 228 p.
- VALETTE, Jean-Jacques. *O papel dos arquivos na administração e na política de planificação nos países em desenvolvimento*. Rio de Janeiro: Arquivo Nacional, 1973. 63 p.

¹ EAD sítio oficial <http://www.loc.gov/ead/>