

# Digitalizando para Durar: a Experiência do Arquivo Público Mineiro

Carlos E. R. Lopes<sup>1</sup>, Eliane D. Amorim<sup>2</sup>, Fernanda M. Vieira<sup>3</sup>,  
Eduardo A. do Valle Jr.<sup>4\*</sup>

<sup>1</sup> Núcleo de Processamento Digital de Imagens (dlopes@dcc.ufmg.br)

<sup>2,3</sup> Arquivo Público Mineiro (apm.sec@mg.gov.br)

<sup>4</sup> Équipes Traitement des Images et du Signal (valle@ensea.fr)

## Resumo:

A partir dos anos 90, os arquivos e outros órgãos de documentação passaram a utilizar a tecnologia como forma de otimizar as ações de preservação e acesso. Assim, nos últimos anos, os projetos voltados para digitalização de acervos passaram a consumir parte significativa dos recursos liberados pelas agências de financiamento. Não obstante estes investimentos correm sérios riscos em virtude da fragilidade dos meios de armazenamento digitais, obsolescência tecnológica e principalmente da ausência de estratégias de preservação digital. Esta comunicação pretende transmitir as ações de preservação digital que vem sendo adotadas no Arquivo Público Mineiro.

**Palavras-chave:** Acesso, Arquivos, Digitalização, Metadados, Preservação digital.

## 1. Introdução

A tecnologia digital, como forma de otimizar ações de preservação e acesso, vem sendo adotada por instituições do mundo inteiro devido às inúmeras vantagens que oferece. O Brasil segue também essa tendência, e, recentemente, projetos voltados para digitalização de acervos documentais têm se beneficiado de significativos financiamentos oferecidos por agências governamentais.

A informação digital apresenta diversas vantagens: a perfeita replicabilidade da informação (cada cópia é fiel à versão original); a possibilidade de acesso remoto eficiente através de WEB; a proteção do original analógico (já que a grande maioria dos usuários se contenta com a cópia), dentre outras. No entanto, a obsolescência tecnológica e a fragilidade dos meios de armazenamento desafiam sua preservação.

A preservação digital deve ser entendida sob dois aspectos. Primeiro a conversão dos documentos analógicos, gerando cópias digitais que permitam o acesso à informação e evitem o manuseio direto e constante dos originais. Segundo, a garantia do acesso, pelas gerações atuais e futuras, aos registros digitais impedindo que a obsolescência tecnológica e a fragilidade dos meios de armazenamento tornem esses registros ilegíveis.

---

\* Eduardo A. do Valle Jr. é bolsista da CAPES - Brasil/Brasília.

Projetos que negligenciem a preservação a longo prazo podem tornar os registros irremediavelmente inacessíveis em menos de uma década. A perda do registro digital, além do desperdício de recursos, ameaça o acervo como um todo - uma vez que o documento original deverá ser submetido a nova digitalização (com resultantes riscos e desgastes). Se o original se houver envelhecido, danificado ou desaparecido, a perda da cópia torna-se crítica.

Mesmo com essa ameaça, parte significativa dos projetos de digitalização desenvolvidos nos últimos anos focou a *geração* de cópias digitais e ignorou sua *preservação*. Isso deve-se, em parte, pelo fato de as instituições terem em mente a preservação dos registros tradicionais, cujas ações podem ser iniciadas, sem risco de perda, décadas após sua produção. O mundo digital, no entanto, exige que as ações de preservação se iniciem no ato de produção do registro e prossigam em todo seu ciclo de vida. Decisões sobre padrões e formatos de arquivos, definição e criação de metadados, cuidados com manuseio e armazenamento, etc. devem ser tomadas desde o início de um projeto de digitalização.

## **2. Experiência do APM**

Criado em 1895, o Arquivo Público Mineiro é a instituição cultural mais antiga de Minas Gerais. O acervo sob sua guarda é composto de documentos textuais, fotografias, filmes, mapas e plantas de origem pública e privada que remontam ao início do século XVIII e cobrem os períodos colonial, imperial e parte do republicano.

A partir de 1999, a instituição vem utilizando a tecnologia digital como forma de otimizar as ações de preservação e acesso. Passados cinco anos, estão implantados ou em andamento vários projetos voltados para conversão de acervos ao formato digital e implantação de sistemas de recuperação da informação. Foram priorizadas as coleções atingidas por diversos fatores de degradação / deterioração (acidez dos papéis, tinta ferrogálica, manuseio, pragas e outros), os documentos muito acessados, e os acervos dispersos em diversas instituições. Para os documentos textuais, um sistema híbrido, associando microfilmagem e digitalização, foi adotado.

Não foram contempladas questões como autenticidade e assinatura - ações que, embora fundamentais na preservação de documentos que sejam digitais desde a

origem, são menos importantes em acervos digitais gerados a partir de documentos analógicos, cujos originais permanecem salvaguardados.

Dentre os projetos já concluídos, ou em fase final, podemos citar:

- Imagens do acervo fotográfico do APM: preservação e acesso em meio digital;
- Acervo do DOPS: acesso em meio digital;
- Projetos cooperativos Digitalização da Comissão Construtora da Nova Capital;
- Edição eletrônica da Revista do APM;
- Preservação e acesso aos acervos das câmaras municipais.

Esses projetos de digitalização resultaram, até o momento, em 350.000 imagens de documentos textuais, 11.000 fotografias e 200 imagens de mapas e plantas.

### **3. Aspectos Administrativos**

#### *Estratégia de Preservação*

A legibilidade de um documento digital, através dos anos e décadas após sua geração, se dá através de procedimentos de migração dos dados, emulação das tecnologias obsoletas e cuidadoso controle das mídias de armazenamento. Essas condições não podem ser atendidas sem uma documentação acerca do processo de digitalização dos documentos.

Essa documentação fornece todas as informações técnicas a respeito do registro digital - e é importante armazená-la de forma padronizada e independente de tecnologia. O conjunto dessas informações descritivas e explicativas do documento digital costuma ser chamado de *metadados*\*.

Os metadados adicionam valor ao documento, esclarecendo acerca do seu conteúdo, formato e história administrativa. No caso de metadados de preservação digital, são essenciais os dados sobre a imagem (formato do arquivo, resolução, ...), sobre o equipamento digitalizador (nome, modelo, configurações, ...), sobre o processo de aquisição (nome do operador, data da criação, *software* utilizado, ...). Essas informações permitem reconstruir o contexto original de criação do documento, e portanto, sua leitura, cópia para novas mídias, migração para

---

\* Expressão que significa, literalmente, *dados acerca dos dados*.

diferentes formatos de arquivo, emulação da tecnologia original, e outras operações de preservação digital.

Além da padronização do fluxo de trabalho, outra estratégia de preservação digital consiste no uso do padrão *XML*<sup>\*</sup> para preservação dos metadados. Atualmente o *XML* é o padrão de arquivo recomendado para preservação de arquivos de textos que não necessitam preservar sua identidade visual. Dessa forma, o Arquivo Público Mineiro desenvolveu junto com o Departamento de Ciência da Computação da UFMG um sistema capaz de converter ao padrão *XML* as informações existentes em diferentes bases de dados.

### *Fluxo de Trabalho*

Para padronizar o processo de geração de imagens no acervo do Arquivo Público Mineiro, um sistema de gestão do fluxo de trabalho<sup>†</sup> foi implantado. O sistema permite acompanhar a seleção do formato de arquivo, resolução, metadados de preservação, destinação dos diversos tipos de imagens gerados.

O fluxo de trabalho adotado no APM pode ser visualizado no esquema da Figura 1.

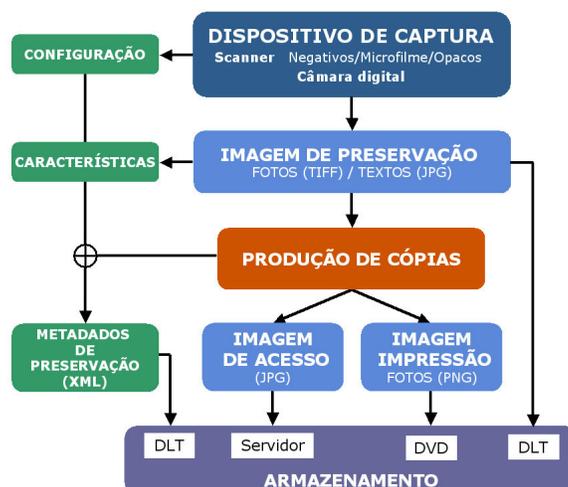


Figura 1: Fluxo de trabalho para digitalização adotado no APM.

O sistema de fluxo de trabalho é composto de 4 módulos, conforme a Figura 2.

<sup>\*</sup> Linguagem de marcação extensível, do inglês *eXtensible Markup Language*.

<sup>†</sup> Do inglês *workflow*.

O módulo Administrativo é o responsável pela gestão das tabelas de apoio aos outros módulos: dispositivos de captura, configurações dos fundos, mídias de armazenamento, ...

É utilizado pelo profissional responsável pelo projeto.

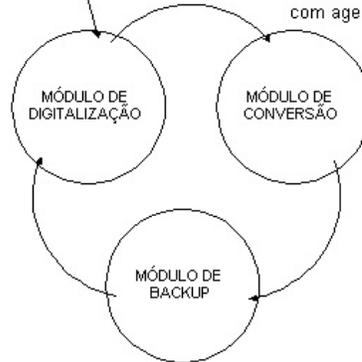


O módulo de Conversão executa a geração das imagens de acesso a partir da imagem master. As imagens geradas seguem as configurações estabelecidas no módulo administrativo.

É executado automaticamente de acordo com agendamento.

O módulo de Digitalização fornece para o usuário as diretrizes durante a digitalização e solicita as informações necessárias para formar os metadados de preservação digital do documento.

É utilizado pelo profissional responsável pela digitalização.



O módulo de Backup mantém o controle das imagens que já foram armazenadas nas respectivas mídias de acordo com as configurações no módulo administrativo.

É utilizado pelo profissional ligado ao suporte ou infra estrutura.

Figura 2: Os diferentes módulos do sistema de Fluxo de Trabalho.

O módulo **administrativo** contém informações básicas dos sistemas tais como dispositivos de captura, configurações dos fundos, padrões de digitalização, dispositivos utilizados para armazenamento das imagens, e outros. São extraídos neste módulo os metadados relacionados ao equipamento utilizado durante o processo de digitalização.

O módulo de **digitalização** orienta o operador durante todo o processo de digitalização, fornecendo informações quanto à manipulação dos documentos, configurações dos equipamentos, decisões a tomar, etc. Neste módulo são extraídos os metadados referentes à operação de conversão.

No módulo **produção de cópias** são realizadas as compactações automáticas das imagens para geração de cópias secundárias para acesso e impressão (discutidas no item *Cópias de Preservação x Cópias de Acesso*, adiante).

O módulo de **backup** possui uma interface que orienta o usuário quanto a definição das mídias de armazenamento, direcionando cópias de preservação acesso e

impressão para os dispositivos adequados (veja o item *Armazenamento e Distribuição*, adiante).

#### **4. Aspectos Técnicos**

##### *Formato do Arquivo de Imagem*

A correta escolha do formato de arquivo digital a ser utilizado torna mais fácil a tarefa de preservação de documentos digitais. Um arquivo digital de preservação deve considerar a acessibilidade e a utilidade do documento em longo prazo (próximas décadas). Formatos não apropriados podem resultar em dificuldades futuras em questões como legibilidade, capacidade de migração, dentre outros.

O instituto *TASI*<sup>\*</sup> cita alguns cuidados que devem ser observados no momento da escolha do formato para arquivos digitais. São eles:

- Preferir formatos não-proprietários, ou seja, aqueles cuja especificação seja pública e cujo desenvolvimento não esteja ligado a uma empresa específica. O uso de formatos proprietários torna às vezes impossível obter uma especificação oficial, além de introduzir o risco da descontinuidade do formato pela empresa criadora. Em optando por um formato proprietário, prefira aqueles amplamente documentados e usados.
- Utilizar formatos que sejam padrões industriais *de facto*. Uma vez que mais organizações, empresas e pessoas utilizam um determinado formato, sua vida útil é estendida, bem como o suporte técnico de que ele desfruta.
- Evitar a compressão dos dados. Sendo a compressão necessária, utilize formatos com compressão sem perdas ou privilegie a qualidade ao máximo em detrimento à redução do tamanho final do arquivo. Isso deve ser feito em consideração das necessidades futuras, ainda não previstas, da informação.
- Evitar a criptografia dos dados. Se for necessário fazer uma migração de formato posteriormente, a criptografia pode introduzir dificuldades muitas vezes intransponíveis.

---

<sup>\*</sup> *Technical Advisory Service for Images*, veja referência bibliográfica no fim

### *Formato dos Metadados*

A legibilidade dos metadados é ainda mais crítica que a dos próprios dados, pois em caso de obsolescência da tecnologia, ou perda do contexto dos dados, só será possível recuperar a informação se os metadados estiverem disponíveis. Parte dos metadados deve, inclusive ser preservada em forma tradicional (papel) de forma a poder recuperar os outros metadados, em caso de necessidade. Por exemplo, as mídias devem ter etiquetas explicativas com a data de sua geração, equipamento e *software* necessário para utilizá-las, e instruções de como recuperar os outros metadados.

Um formato preferido para o armazenamento de metadados é o de texto puro, pois é dos mais fáceis de recuperar em caso de obsolescência ou necessidade de migração. Infelizmente, texto puro não é estruturado, e muitos dados se representam melhor na forma de tabelas, hierarquias ou outras formas semi-estruturadas.

O padrão *XML*, linguagem de marcação extensível, permite combinar as vantagens do texto puro adicionadas à uma estruturação dos dados - e por isso, foi adotado no APM para representação dos metadados.

### *Cópias de Preservação x Cópias de Acesso*

Às vezes as necessidades de preservação digital entram em conflito com as de acesso. Por exemplo, as imagens digitais de preservação devem ter alta resolução e nenhuma compressão ou criptografia. Entretanto, as imagens de acesso devem ser pequenas, não precisam de tão alta resolução, e eventualmente precisam ser marcadas ou criptografadas, de forma a garantir sua proveniência e confidencialidade.

Para resolver esse conflito, utiliza-se um sistema de diferentes cópias digitais de um mesmo documento, cada uma para atender um propósito específico. Primeiro gera-se um máster digital de alta qualidade, e as imagens de acesso são geradas automaticamente a partir desse máster.

A imagem de **preservação** ou **máster** é aquela que representa o “original digital” da fotografia devendo possuir, portanto, qualidade elevada. Para as imagens de

preservação de fotografias, o APM escolheu o formato *TIFF*<sup>\*</sup>, que embora proprietário da *Adobe Systems Incorporated*, é extensamente documentado e armazena as informações sem compressão.

Para disponibilizar o acervo fotográfico via *Web* ou *Intranet*, o máster digital é muito volumoso. A partir do máster, um arquivo de **acesso** é gerado em baixa resolução e compressão com perdas (compressão *JPEG* usando o formato *JFIF*, não-proprietário e extensamente utilizado na *Web*)

As imagens de acesso não tem qualidade suficiente para serem impressas num livro ou revista. Para esse propósito, o arquivo máster é copiado em uma imagem de **impressão**, de resolução intermediária e comprimido sem perdas usando o formato *PNG*<sup>†</sup>, também não proprietário e muito usado na *Web*.

### *Armazenamento e Distribuição*

O armazenamento do acervo digital requer tanto planejamento e precaução quanto o de um acervo físico. De fato, dada a fragilidade dos suportes às variações de temperatura e umidade relativa, à poeira e outros poluentes, e mesmo a campos magnéticos, as exigências em termos de controle de depósito e manipulação podem ser mais severas para as mídias digitais que as convencionais.

A escolha do suporte deve levar em conta aspectos como finalidade do arquivo digital, longevidade e confiabilidade da mídia, longevidade e confiabilidade dos dispositivos de leitura, condições de armazenamento e custo.

Normalmente as imagens de acesso são armazenadas em discos rígidos, que podem ser associados em arranjos chamados *RAID*<sup>‡</sup>. Esses arranjos permitem diversos graus de redundância dos dados, de forma a permitir, em caso de falha de um dos discos, a recuperação completa da informação a partir dos outros. Foi essa a solução adotada no APM.

As imagens de preservação também podem ser armazenadas em grandes arranjos *RAID*, mas, dado o custo elevado de armazenar tanta informação *online*, normalmente escolhe-se gravá-las em fitas magnéticas ou discos ópticos. No APM,

---

<sup>\*</sup> Formato de Arquivo de Imagem Etiquetado, do inglês *Tagged Image File Format*

<sup>†</sup> Gráficos Portáteis em Rede, do inglês *Portable Network Graphics*

<sup>‡</sup> Arranjo Redundante de Discos Baratos, do inglês *Redundant Array of Inexpensive Disks*

as imagens de preservação são armazenadas em fitas *DLT\** do tipo *IV*, que apresentam baixo custo por *megabyte*, e uma grande longevidade e confiabilidade, tanto da mídia quanto do dispositivo de leitura/gravação.

Os discos ópticos são uma alternativa às fitas magnéticas, sobretudo pela sua popularidade, a facilidade de acesso não-seqüencial e a maior robustez. No APM, discos *DVD†* são utilizados para as imagens de impressão.

## **5. Conclusão**

A ausência de planos de preservação que garantam a longevidade dos registros digitais ameaça investimentos e esforços. Instituições que desenvolvem projetos voltados para conversão de acervos analógicos para o formato digital devem, além da responsabilidade de preservar os acervos originais, assumir um compromisso de preservação dos acervos digitais, sob pena de perder o investimento realizado.

Na área da preservação, a tecnologia digital foi incorporada como mais uma etapa do processo de preservação dos documentos sem substituir os procedimentos já consagrados como higienização, acondicionamento, microfilmagem e outros. No caso dos documentos textuais, o objetivo de garantir a preservação com qualidade arquivística determinou a escolha de implantação do sistema híbrido de reformatação para preservação, associando microfilmagem e tecnologia digital.

Não obstante, a maior mudança na área da preservação está relacionada a preservação dos registros digitais. No caso do APM, a padronização do fluxo de trabalho, a adoção de formatos de arquivos de imagens e metadados cuidadosamente selecionados, e o uso correto das mídias digitais apropriadas estão no centro da estratégia utilizada para preservação dos dados digitais para esta geração e as vindouras.

## **6. Bibliografia**

ASCHENBRENNER, Andreas. The Bits and Bites of Data Formats – Stainless Design for Digital Endurance. RLG News. 15 fev. 2004, vol. 08, nº 01.  
<http://www.rlg.org/preserv/diginews>

---

\* Fita Digital Linear, do inglês *Digital Linear Tape*

† Disco Digital Versátil, do inglês *Digital Versatile Disc*

- BROWN, Adrian (2003). Digital Preservation Guidance Note 3: Care, handling and storage of removable media. The National Archives. United Kingdom
- CONWAY, Paul. Preservação no universo digital / Paul Conway: Coordenação de Ingrid Beck; tradução de Rubens Ribeiro Gonçalves da Silva -- Rio de Janeiro: Projeto conservação preventiva em bibliotecas e arquivos: Arquivo Nacional, 1997.
- Digital Preservation Coalition. Disponível em <http://www.dpconline.org/text/about/copyright.html>
- GREEN, David L. et al (2003). The NINCH Guide to Good Practice in the Digital Representation and Management of Cultural Heritage Materials. Humanities Advanced Technology and Information Institute, University of Glasgow and National Initiative for a Networked Cultural Heritage, New York University.
- KENNEY, Anne R., RIEGER, Oya Y. e ENTLICH, Richard (2003). Moving Theory into Practice: Digital Image Tutorial. Cornell University Library, Research Department, Cornell University. Disponível em <http://www.library.cornell.edu/preservation/tutorial/contents.html>
- KRAEMER, Ligia Leindorf Bartz. Metadados como elementos para o tratamento padronizado de recursos virtuais: associação ao projeto bibliotecas virtuais temáticas, do programa PROSSIGA do IBICT. Dissertação de mestrado. Centro Federal de Educação Tecnológica do Paraná. Curitiba, 2001. pág. 89. Disponível em <http://www.ppgte.cefetpr.br/dissertacoes/dis2001.htm>. 07/07/2004.
- ROSS, Seamus e Gow, Ann (1999). Digital Archaeology: Rescuing Neglected and Damaged Data Resources. Humanities Advanced Technology and Information Institute, University of Glasgow.
- Technical Advisory Service for Images (2002). Choosing a File Format. Disponível em <http://www.tasi.ac.uk/advice/creating/format.html>
- VALLE, Eduardo A. Sistemas de Informação Multimídia na Preservação de Acervos Permanentes. Dissertação de Mestrado. Departamento de Ciência da Computação, Universidade Federal de Minas Gerais. Belo Horizonte, 2003.