

Preservação de Documentos Digitais

Resumo da Teoria

Palestras proferidas por
Bodê, Ernesto
Abr./mai. 2009
Universidade de Brasília - Unb

1.0 Introdução

Nesse trabalho, apresentamos um resumo dos tópicos abordados nas palestras sobre preservação de documentos digitais. Essas palestras foram apresentadas na disciplina **Conservação e Restauração de Documentos**. Este resumo tem como objetivo servir como texto unificador; estabelecendo assim, mais claramente, as relações entre os diferentes tópicos abordados.

Os temas abordados nas palestras procuraram avançar paulatinamente em alguns problemas teóricos e práticos que a comunidade científica tem pesquisado, no que diz respeito à **preservação de documentos digitais**. Inicialmente, abordamos conceitos básicos: o conceito de documento eletrônico, digital e formatos de arquivo. Em seguida, estabelecemos uma breve discussão sobre a questão da autenticidade de documentos digitais. Para então apresentarmos aspectos relacionados à preservação do suporte físico de fitas magnéticas e discos CD's e DVD's. Finalmente, apresentamos estratégias atualmente utilizadas para implementar políticas de preservação corporativa de documentos digitais.

Estruturamos esse texto a partir dessa divisão de temas. Ao final, apresentamos uma conclusão geral sobre os assuntos abordados.

2.0 Conceitos Fundamentais

2.1 Documento, documento eletrônico e documento digital

Nessa primeira etapa, procuramos apresentar conceitos fundamentais que serviriam de base para as futuras discussões. Intencionalmente, procuramos fugir de definições de classes profissionais, como documento de museu, de biblioteca ou de arquivo. Há essa necessidade porque os documentos digitais estão presentes em todas essas três realidades, além de muitas outras. Os problemas são essencialmente os mesmos e parece ser mesmo vantajoso, pelo menos profissionalmente, abordar o problema de maneira ampla. Além disso, é preciso mencionar o fato de que a disciplina Conservação e Restauração de Documentos tem como público, pelo menos, graduandos em biblioteconomia e arquivologia.

Estranhamente, não parecem existir muitos textos que abordem o conceito de documento de maneira geral, apesar de facilmente encontrarmos trabalhos sobre o documento histórico, o documento arquivístico, o documento biblioteconômico e etc., sobre isso Raul Contreras escreveu:

“Cientificamente, historiadores, juristas e diplomatas definiram e caracterizaram [documento] nem sempre com total precisão ou amplitude porque tais definições foram formuladas a partir da ótica da especialização de quem fez a definição e também porque o documento é obra humana e não se reduz a um esquema único e rígido.” (CONTRERAS, 1983, p. 31)

Mesmo trabalhos específicos sobre a preservação de documentos digitais não se aprofundam numa discussão sobre documento e documento digital. Essa pode ser uma abordagem perigosa, um disco vinil é um documento eletrônico? Em tudo se assemelha a um disco CD (suporte plástico com conteúdo reproduzido através de um equipamento eletrônico), sendo esse último largamente aceito como documento eletrônico. Porém, os discos em vinil, apesar de documentos sonoros, não utilizam tecnologia digital. Assim, torna-se clara a necessidade de distinguir entre documento

eletrônico e digital, o mesmo aplica-se a documentos em suportes como fitas magnéticas (nem todas com tecnologia digital). Na busca de uma clara distinção conceitual, iniciaremos nosso raciocínio com o próprio conceito de documento.

Michael Buckland explorando o conceito de documento e seu desenvolvimento histórico nos relata que:

“Uma primeira evolução foi a extensão da noção de documento para além de textos escritos, um uso encontrado na maioria dos dicionários de Inglês e Francês. *“Qualquer expressão do pensamento humano”* foi uma definição freqüentemente usada para “documento”, entre documentalistas. Nos EE.UU, as frases “registro gráfico” e “livro genérico” foram largamente utilizadas. Isso foi conveniente para estender o escopo e incluir fotos e outros materiais áudio-visuais. Paul Otlet (1868-1944) ficou conhecido por sua observação de que documentos poderiam ser tridimensionais o que possibilitou a inclusão da escultura. Desde 1928, objetos de museus foram incluídos pelos documentalistas dentro da definição de “documento”.” (BUCKLAND, 1997, página web)

Uma década e meia antes, Raul Contreras, também investigando o conceito de documento, chama a atenção para a estrutura de um documento:

“Resulta que um documento possui uma estrutura constituída pela matéria mediante a qual a representação se torna perceptível: pedra, pergaminho, papel, linho, película cinematográfica e etc.; pelo meio que se adota para fixar nessa matéria a representação: signos gráficos (gravados ou executados a mão), pintura a óleo e etc.; e pelo conteúdo mesmo do documento: um texto biográfico, laudatório, funerário, um testamento, um relato e etc.” (CONTRERAS, 1983, p. 32)

As duas recentes definições acima, de renomados pesquisadores em diferentes áreas ligadas ao documento, nos permitem concluir que o conceito de documento na atualidade é algo bastante amplo e inclui todos os suportes e formas de representação humana. Fica evidente a necessidade de definir conceitos específicos para os interesses de áreas definidas de pesquisa, no nosso caso, os termos que giram em torno de “documento eletrônico”.

A “coisa” que inicialmente vamos chamar genericamente de documento eletrônico começa a surgir na humanidade no final do século XIX:

No final do século XIX, vários inventos para registro do som culminaram no início do século XX com os discos de áudio, e logo depois, o uso também de fitas magnéticas. Estas últimas, após um período de evolução, também passaram a ser utilizadas para gravação de vídeo (os primeiros programas televisivos gravados). No final do século XX surgiram os discos do tipo CD (*Compact Discs*), inicialmente para gravações de áudio (CD-ROM) e depois surgindo os modelos específicos para vídeo (DVD's). (BODÉ, 2006, p.5)

Esses novos documentos, além de utilizarem suportes físicos diferentes dos anteriores e novas maneiras de representar a informação, possuem características peculiares em relação a todos os outros documentos já utilizados pela humanidade, talvez a principal delas seja o fato de que são **‘legíveis por tecnologia’**, ou seja, necessitam de dispositivos e máquinas para que possam ser “lidos” pelos humanos, com a evolução ocorrida ao longo do último século, outra característica presente é a **‘independência de suporte e conteúdo’**, em contraste com um pergaminho da idade média, p. ex., hoje é possível em certos documentos, extrair, modificar ou copiar seu conteúdo sem nenhuma alteração fundamental no suporte físico, sendo possível, até mesmo, a substituição do suporte físico com manutenção de todas as características importantes do conteúdo original.

Se já não houvesse um problema terminológico, poderíamos utilizar ainda o termo **documentos tecnológicos**, para indicar a forte presença e necessidade de tecnologia para produção e uso desses documentos. É claro que o papel, a tinta, a pena e etc. são também tecnologias, ainda que hoje bastante simples e totalmente dominadas. Mas nos referimos, quando falamos em tecnologia, ao uso da eletrônica, principalmente, e hoje, cada vez mais, ao uso da microeletrônica aplicada à informática, o que resulta não apenas em tecnologias físicas (hardware) mas igualmente às tecnologias de sistemas de computadores (software): sistemas operacionais, aplicativos e etc.

A forte presença da informática na atualidade talvez ofusque a existência de documentos que vem sendo produzidos ao longo do último século sem o uso do computador: discos de áudio (em sua última versão, produzidos no material vinil), fitas magnéticas, tanto para áudio como para vídeo. Principalmente do ponto de vista da preservação de documentos, não se pode ignorar esses outros espécimes¹.

Não é nosso objetivo aqui desenvolver uma difícil discussão sobre essa questão terminológica, mas é importante notar que o termo 'Documentos Eletrônicos' não pode ser aplicado indistintamente, há documentos sonoros, por exemplo, que não utilizam nenhum tipo de eletrônica, como os primeiros gramofones totalmente mecânicos. E, principalmente, o uso do termo 'Documento Digital' possui aplicabilidade restrita dentro do conjunto maior desses novos documentos. É evidente que nem todo documento eletrônico é digital, no entanto, não é tão evidente que todo documento digital não precisa ser eletrônico. Vejamos mais sobre isso.

A designação de um documento tem estado relacionada ao suporte físico utilizado ou ao tipo de conteúdo registrado. Assim, temos documentos em papel, pergaminho ou papiro; e também documentos sonoros ou filmicos, p. ex. Um mesmo exemplar pode ser chamado 'documento em papel' e 'documento textual'. Esses termos estão relacionados ao estudo da **matéria, meio e conteúdo** dos documentos, como por exemplo Raul Contreras observou acima nesse texto. Nesse sentido, os termos **Documento Eletrônico** e **Documento Digital** deveriam ter como suporte físico algo eletrônico ou seu conteúdo em códigos digitais. Porém, na verdade, o que vem sendo chamado de documento eletrônico ou documento digital, efetivamente, não está gravado em algo 'eletrônico', CD's e DVD's são constituídos essencialmente de plástico policarbonato, discos HD (inclusive os presentes em grandes servidores de arquivo) são constituídos por placas metálicas e etc., a eletrônica está presente sim, mas nos equipamentos de gravação e leitura desses suportes físicos. Apesar de se fazer notar que mais recentemente, suportes hoje conhecidos como *Pen Drives* são constituídos essencialmente de memórias microeletrônicas. Com relação à codificação digital é preciso esclarecer que qualquer pessoa pode, numa folha em branco de papel, escrever uma mensagem em dígitos binários, compondo códigos² correspondentes a caracteres (letras, espaços, pontuação, acentuação e números). Ressalte-se ainda que essa folha de papel pode ser lida, pelo menos para mensagens curtas, diretamente por seu produtor humano, sem o uso de tecnologia eletrônica qualquer. Sobre isso ainda é preciso dizer que

¹ Veja no sítio da Biblioteca Nacional (<http://www.bn.br/fbn/musica/acervo.htm>) acervo de músicas originais disponibilizadas em formato digital.

² Ver tabela ASCII e UNICODE para uma padronização do uso da linguagem binária.

nos primórdios da computação foram utilizados cartões perfurados (em papel) em codificação binária para entrada de dados em computadores³.

Michael Buckland, investigando a noção de documento digital, a partir de vários pesquisadores documentalistas, observa que:

“A evolução da noção de ‘documento’ entre Otlet, Briet, Schuermeyer e os outros documentalistas enfatizaram fortemente o que funcionava como um documento ao invés das tradicionais formas físicas de documentos. A mudança para a tecnologia digital parece dar ainda mais importância a essa distinção. Uma análise sobre o pensamento de Levy tem mostrado que uma ênfase na tecnologia dos documentos digitais impediu nossa compreensão dos documentos digitais como documentos.” (BUCKLAND, 1998, página web).

Concluindo, a despeito das possíveis controvérsias levantadas, parece indiscutível que existiram (e muitos ainda estão preservados) inicialmente documentos que se utilizaram de tecnologia para produção e visualização, depois documentos que se utilizaram de tecnologia eletrônica e informática para sua produção e visualização e por último, os documentos codificados digitalmente que se utilizam de microeletrônica e software para produção e visualização. Hoje, esses últimos são aqueles conhecidos como **Documentos Digitais**.

2.2 Formatos de Arquivo

O conceito de formatos de arquivo é de suma importância para a compreensão do que são documentos digitais e das estratégias que podem ser aplicadas para sua preservação, principalmente para longos períodos de tempo.

Uma definição bastante completa sobre formatos de arquivo foi elaborada em relatório de pesquisa da Universidade de Leeds no Reino Unido:

“No nível mais básico, objetos digitais são seqüências de zeros e uns que representam dados codificados. Diferentes **Formatos de Arquivo** especificam como estes códigos representam o conteúdo intelectual criado por um autor do objeto digital. Um exemplo disto é o formato *Microsoft Word*. Este formato é uma especificação para armazenamento de dados textuais, bem como informações de formatação. Muitos Formatos de Arquivo são incrivelmente complexos, de maneira que os códigos podem ficar ininteligíveis para um observador humano. Para que este objeto digital tenha sentido, um software será necessário para interpretar e exibir [ou renderizar] os dados para o usuário.” (UNIVERSIDADE DE LEEDS, 2003, grifo nosso)

Grosso modo, um formato de arquivo qualquer nos diz como a organização dos dígitos binários de um documento com qualquer conteúdo (som, texto, imagem e etc.) está constituída, ou seja, onde começa o conteúdo efetivamente, tipo de fontes gráficas e cor (para texto) ou altura e timbre (para som) ou quantos bits especificam as cores (para imagens) e uma infinidade de outras informações técnicas. Essa ‘organização’ da informação binária permite a produção de arquivos de computador em um certo formato e o desenvolvimento de softwares que conseguem ‘ler’ esses arquivos, da maneira correta.

O conhecimento dessa informação técnica, a especificação do formato de arquivo, é um elemento fundamental para sua preservação ao longo do tempo. Pois permitirá interpretar corretamente os zeros e uns dos arquivos de computador. Infelizmente, mesmo alguns desenvolvedores de software que criaram alguns

³ Esses cartões, conhecidos como *Punch Cards* podem ser pesquisados em instituições como o *National Museum of American History* (<http://americanhistory.si.edu/collections/>).

formatos de arquivo, não possuem essa informação da maneira adequada, como nos relata Murray e VanRyper *“Nem todos os formatos são documentados porém, e alguns documentos são tão vagos, pobremente redigidos ou desatualizados que se tornam essencialmente inúteis.”* (MURRAY & VanRYPER, 1994, p. xv)

Para piorar ainda mais o quadro, é importante lembrar que cada formato de arquivo, possui normalmente várias versões de especificações técnicas, algumas delas consideravelmente diferentes das anteriores, por exemplo o formato pdf (*portable document format*) está atualmente em sua versão 1.7 e iniciou na versão 1.0.

Um outro aspecto preocupante, do ponto de vista da preservação de documentos, relativo a formatos de arquivo é a existência de restrições para acesso às informações técnicas dos formatos. Há especificações técnicas proprietárias que, por motivos legais, não podem ser utilizadas para desenvolver software necessário para ler os documentos digitais gerados a partir dessas mesmas especificações. Daí a existência da recomendação, cada vez mais freqüente, do uso de especificações não proprietárias, ou seja, abertas ao acesso público. Nessa linha, a especificação PDF/A (*Archive*), recentemente transformada em norma ISO, é uma excelente opção fins de preservação, pelo menos para documentos textuais⁴.

3.0 Preservação dos suportes físicos

A preocupação com a preservação de documentos digitais não deve deixar de incluir ações sobre a preservação dos suportes físicos onde esses documentos estão gravados. Não existe qualquer documento digital sem um suporte físico para armazená-lo, por mais virtuais que alguns documentos possam parecer, principalmente em ambientes de rede como a internet.

No entanto, a preocupação com o suporte físico dos documentos digitais não parece ser a coisa mais relevante diante de tantos desafios existentes, principalmente relativos à obsolescência de software e hardware, ações de migração, manutenção da autenticidade e etc. Além disso, medidas relativamente simples como a duplicação dos documentos, refrescamento⁵ de suportes ou criação de *backups*⁶, mitigam bastante os problemas relativos à preservação dos suportes físicos.

Nos parágrafos seguintes, reproduzimos um texto⁷ com modificações sobre os cuidados com a preservação de suportes físicos do tipo fitas magnéticas e discos CD.

3.1 Fitas Magnéticas

A fig. 1 mostra a estrutura básica de uma fita magnética típica.

⁴ Para mais informações consulte: **LeFURGY**, William G. PDF/A: Developing a file format for long-term preservation. **RLG News**, NY, v. 7, n. 6, 2003. Disponível em: <http://www.rlg.org>. Acesso em: 10 novembro 2005.

⁵ Será visto mais adiante nesse texto.

⁶ Principalmente através da cópia em fitas magnéticas.

⁷ Texto original apresentado por nós no III Simpósio de Técnicas Avançadas em Conservação de Bens Culturais, Olinda, 2006. Disponível para acesso em <http://www.cipedya.com/doc/101657>

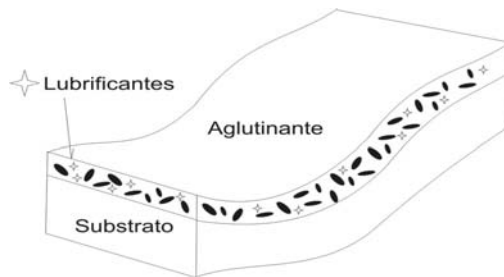


Fig. 1 – Estrutura básica de uma fita magnética (adaptada de Van Bogart, 2001)

A sustentação física da fita magnética é obtida pela camada de substrato. As partículas magnéticas estão inseridas no aglutinante.

É preciso ressaltar que, como Van Bogart observou, vários fatores estão relacionados com o tempo de vida de fitas magnéticas, entre eles "*número de vezes que a fita é utilizada durante sua vida*" ou "*disponibilidade futura de tecnologia de sistema para se reproduzir a fita*" (BOGART, 2006, p. 21). Nosso foco aqui, relaciona-se às condições de armazenamento, manuseio e transporte e sua relação com a estrutura física do suporte.

Hidrólise do aglutinante – A quantidade de umidade relativa no ambiente afetará o aglutinante e os pigmentos magnéticos da fita. Em temperaturas menores este processo torna-se mais lento.

Armazenagem em temperaturas elevadas (> 23°) – Em temperaturas elevadas, o aglutinante da fita tende a incorporar partículas, deformando as informações gravadas. Com índices de umidade elevados (> 70 UR) aumenta a degradação do aglutinante e com mais umidade o suporte da fita também tende a ceder.

Fungos – O armazenamento com altas taxas de umidade e temperatura propiciará o aparecimento de fungos junto ao aglutinante e outros aditivos presentes.

Variações de condições de armazenamento – O substrato das fitas magnéticas tenderá a contrair-se ou se expandir a partir de variações de temperatura e umidade o que pode provocar deformações permanentes no substrato.

Recomendação: Manter as condições de temperatura e umidade constantes.

Resumindo, em se tratando de fitas magnéticas a estabilidade do aglutinante é o ponto central para manter a integridade desse suporte. Alterações de temperatura, umidade e presença de partículas podem afetar esta camada da estrutura definitivamente. Cabe notar também que estas condições independem da fita magnética conter dados gravados ou não, em outras palavras é preciso considerar o tempo de armazenamento da fita, desde sua fabricação e não a partir de seu uso.

Por último, mas não menos importante, é preciso fazer notar que a vida útil de uma fita magnética, aliás este princípio aplica-se para qualquer suporte documental complexo como ela, será estabelecida a partir da vida útil de seu elemento mais frágil. Em outras palavras, identificar a vida útil significa identificar qual parte da estrutura do suporte é menos perene.

3.2 Discos CD (Compact Discs)

Antes de mais nada, cabe esclarecer que não existe uma mídia chamada CD. Este termo designa um grupo de discos com características comuns mas com especificidades próprias. Desta maneira, encontraremos os CD-ROM, CD-R, CD-RW, DVD-ROM, DVD-R, DVD-RW, para abordar os mais comuns do mercado.

Considerando então esta diversidade, faremos nossa análise a partir dos três grupos básicos, ROM (Read Only Memory), R (Recordable) e RW (ReWritable), aplicável tanto para CD's como para DVD's.

A fig. 2 mostra a estrutura básica da família de discos CD (ROM, R e RW). As recomendações a seguir são um resumo do pode ser encontrado em (BYERS, 2003) e (Understanding CD-R and CD-RW, 2003):

ROM – Este discos tem como ponto fraco a camada de alumínio de sua estrutura. Este material, quando exposto ao contato com oxigênio (via umidade do ar, ranhuras ou falhas em sua construção) entrará em processo de oxidação, perdendo parte de sua capacidade reflexiva. Em condições ambientais com temperaturas mais elevadas o efeito de oxidação será mais proeminente. O alumínio também pode ser afetado pela presença de solventes ou poluentes no ar ou por contato direto (etiquetas, marcas com canetas, etc.).

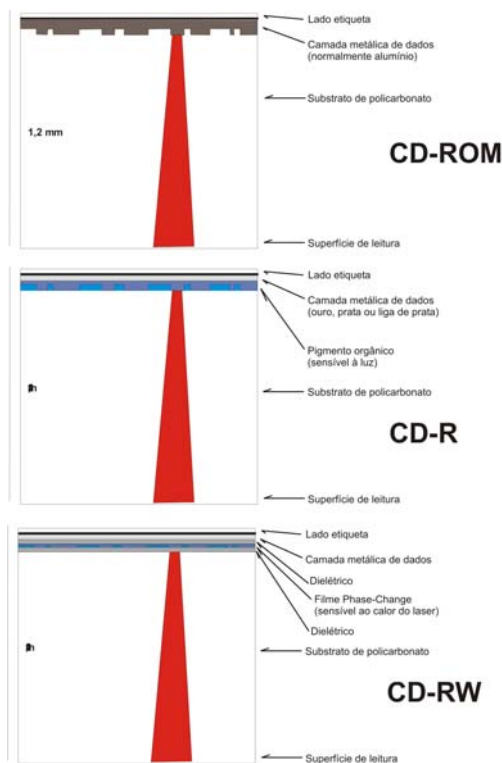


Fig. 2 – Estrutura dos discos do tipo CD (adaptado de Byers, 2003)

R – Neste tipo de disco, a camada reflexiva metálica não é composta de alumínio, mas de ouro, prata ou uma liga de prata (dependendo do modelo e fabricante). Discos construídos com ouro são mais duráveis em relação à perda de reflexividade, já que o ouro praticamente não oxida, no entanto, o custo desses discos é bastante alto. Já nos discos construídos com camada de prata há o perigo de corrosão pelo dióxido sulfúrico presente em poluentes. Algumas ligas de prata utilizadas são mais resistentes à corrosão. De qualquer forma o perigo de corrosão

por poluentes é consideravelmente menor que pela presença de umidade (caso do uso de alumínio em discos do tipo ROM).

Os discos do tipo R possuem também, além da camada metálica reflexiva, uma camada com pigmentos orgânicos, este material será alterado quimicamente pelo feixe de laser que marcará os bits no suporte. Infelizmente, estes pigmentos se degradam naturalmente com o tempo. São também sensíveis aos efeitos de luz direta (Ultra-violeta, Infra-vermelho e fluorescente).

RW – O ponto fraco deste grupo de discos é a camada de *phase-changing metal alloy film*. Este material degrada mais facilmente que a camada de pigmento orgânico em discos do tipo R. O princípio de gravação e regravação nestes discos funciona a partir do calor gerado pelo feixe de laser na camada de filme. Portanto, alterações de temperatura no ambiente podem afetar os dados gravados neste tipo de disco. Além disto, a luz direta pode afetar a temperatura dos discos e indiretamente causando alterações nos dados gravados.

O número de vezes que os dados são gravados e regravados afeta também a durabilidade da camada de filme.

4.0 Estratégias de Preservação

A prática e a teoria sobre a preservação de documentos digitais determinaram a criação de algumas estratégias a serem executadas para aumentar ao máximo a possibilidade de conservação dos documentos digitais em condições adequadas. Condições essas que devem possibilitar sua autenticidade (em sentido abrangente), sua integridade física, condições de acesso e leitura de seu conteúdo, além de outros itens. Os cuidados com a preservação dos suportes físicos, que vimos no item 4.0 anteriormente, também podem ser considerados como uma estratégia de preservação ou pelo menos uma etapa necessária, mas por sua própria especificidade foram vistos em capítulo separado.

As principais estratégias de preservação de documentos digitais estão relacionadas a algum tipo de mudança do software (formatos de arquivo) ou hardware (suporte físico e aparatos para leitura adequada), há propostas para emulação por software também e o uso de especificações técnicas bem documentadas e abertas ao conhecimento de todos os interessados nas atividades de preservação.

Veremos a seguir as definições básicas das principais estratégias que identificamos na literatura disponível.

4.1 – Emulação

Emulação no nosso contexto significa *“Comportamento de um computador ou impressora que é exatamente igual ao de equipamentos similares, permitindo que os programas e dados que rodam num deles, rodem nos outros”*⁸. O que se pretende é executar programas, incluindo sistema operacional, aplicativos e etc., já ultrapassados ou criados em outras plataformas⁹ tecnológicas. Em termos de

⁸ Dicionário HOUAISS eletrônico, verbete **emulação**.

⁹ Uma plataforma refere-se a todo o aparato tecnológico de hardware e software disponível, p. ex. a plataforma *Intel* inclui os processadores dessa marca e sistemas operacionais desenvolvidos para serem executados nesse hardware específico.

preservação de documentos digitais, isso possibilitaria que no futuro os documentos atuais pudessem ser interpretados exatamente como o são hoje.

Extraímos uma definição de **emulação** encontrada em um relatório do *Arquivo Nacional dos Países Baixos* “a recriação em hardware atual do ambiente tecnológico necessário para visualizar e usar objetos digitais de tempos passados” (HOEVEN, 2005, p. 7). Nesse mesmo relatório encontramos também a definição de um **emulador**: “é um programa executável em um computador e dessa forma virtualmente recria um computador diferente” (HOEVEN, 2005, p. 7).

Não é difícil compreender que o desenvolvimento de emuladores é uma tarefa complexa, os técnicos precisam conhecer (e muito bem) os detalhes de hardware e software originais e isso é um problema que cresce à medida que o tempo passa. Conseqüentemente, trata-se de uma estratégia com alto custo financeiro. Por outro lado, em condições ideais de funcionamento, os emuladores permitiriam o acesso a documentos digitais exatamente como foram produzidos originalmente.

Miguel Ferreira, num resumo sobre o tema a partir de vários autores que pesquisaram o assunto, alerta para dois possíveis grandes problemas relativos ao uso da emulação. Primeiro, o emulador é um sistema de software e como tal também sofrerá o processo de obsolescência tecnológica, ou seja, não se trata de uma ferramenta definitiva como solução para a preservação de documentos digitais. Segundo, é importante salientar a necessidade de recursos humanos que saibam operar o software original já ultrapassado quando esse estiver disponível no emulador desenvolvido, um problema de recursos humanos, “Por exemplo, num futuro próximo será difícil conceber que os utilizadores estejam aptos a enfrentar as particularidades do sistema operativo MS-DOS” (FERREIRA, 2006, p. 35).

4.2 – Migração

As operações de migração, dentro do escopo da preservação de documentos digitais, referem-se à mudança de algum componente do documento digital original. Assim, pode-se fazer a migração de um formato de arquivo para um outro formato de arquivo, ou de uma versão de formato de arquivo para uma outra versão, normalmente mais atual. Pode-se também, migrar o suporte físico onde o documento está registrado e é claro que conseqüentemente surge a necessidade de um novo tipo de equipamento para leitura desse novo suporte.

Uma espécie de migração que vem sendo chamada de *refresh* ou **refrescamento**, é a substituição do suporte atual (uma mídia CD-R, p. ex.) por uma nova do mesmo tipo, “A frequente verificação da integridade dos suportes físicos, assim como o seu refrescamento periódico, são consideradas actividades vitais num contexto de preservação digital.” (FERREIRA, 2006, p. 33)

Podemos agrupar, grosso modo, as atividades de migração entre as que se referem a hardware, software ou ambas simultaneamente.

No grupo das migrações de hardware, podemos citar a regravação de arquivos de imagens originalmente em disquetes (5^{1/4}, 3^{1/2} e etc.) para mídias modernas como o DVD ou regravar arquivos em HD (*hard disk*) em fitas magnéticas de *backup*. Em ambos os exemplos substituí-se as mídias e os dispositivos necessários para a leitura dos novos tipos de mídias.

As migrações de software referem-se principalmente à mudança de formatos de arquivo. Pode-se, p. exemplo, substituir documentos no formato pdf versão 1.6

para a versão pdf/a (substituição de versão do mesmo formato básico) ou substituir documentos criados no formato doc (*Microsoft*) para o formato pdf/a (substituição do formato original). Claro que conseqüentemente surge também a necessidade de novos aplicativos disponíveis para leitura dos documentos nos novos formatos.

Por último, as migrações de hardware e software referem-se a casos como a troca de filmes em fitas magnéticas no formato VHS para mídias DVD, havendo aí a troca de mídia e formato dos arquivos e dispositivos de leitura dos novos documentos.

4.2.1 – Formatos de Arquivo

Com relação às migrações de formatos de arquivo, basicamente, trata-se de desenvolver meios para *passar* de um formato de arquivo para outro. Assim, sucessivamente, ao longo do tempo, os documentos em arquivos seriam mantidos atualizados tecnologicamente.

No entanto, também há problemas nesta abordagem; num relatório sobre o tema, registra-se *“No presente estágio, migração como uma estratégia de preservação digital pode ser caracterizada como um processo incerto gerando resultados incertos”* (LAWRENCE, 1999, p. 5).

O mesmo relatório citado acima agrupa três grandes categorias de riscos associados com uma estratégia de migração de arquivos.

O primeiro deles, *Riscos associados com a coleção em geral*, refere-se aos riscos relacionados *“à presença ou ausência de apoio institucional, orçamento, sistemas de hardware e software, e uma equipe para gerenciar o arquivo”* (LAWRENCE, 1999, p. 6).

O segundo, *Riscos associados com os dados do formato de arquivo*, inclui *“os elementos da estrutura interna do formato de arquivo sujeitos a modificação”* (LAWRENCE, 1999, p. 6).

E, por último, *Riscos associados com o processo de conversão do formato de arquivo*, ou seja, *“o software de conversão pode ou não produzir os resultados pretendidos; erros de conversão podem ser fortes ou sutis”* (LAWRENCE, 1999, p. 6)

É preciso observar também que:

Tudo no universo digital dos computadores não está perfeitamente encaixado em simples formatos de arquivo, mas muitas vezes também em grupos de arquivos e aplicativos. É o caso dos conteúdos publicados em páginas *web*; como é sabido, neste início de século XXI, praticamente tudo pode ser encontrado numa página da internet, de fofocas a noticiários altamente confiáveis. (BODÉ(2), 2006, p. 35)

As migrações de formatos de arquivo implicam, portanto, possíveis alterações no documento original e as sucessivas migrações de formatos potencializam esse risco. Para mitigar esse problema, Miguel Ferreira cita uma proposta apresentada no *6th European Conference on Research and Advanced Technology for Digital Libraries*, que traduziu como 'Migração a-pedido' (*migration on request*), nessa proposta:

Neste tipo de migração, ao invés de as conversões serem aplicadas ao objecto mais actual, estas são sempre aplicadas ao objecto original. Deste modo, se de uma dada conversão resultar um objecto substancialmente diferente do original, numa futura conversão, o problema poderá ser resolvido recorrendo a um conversor de melhor qualidade ou a um formato de destino mais adequado. (FERREIRA, 2006, p. 40).

O que abordamos no item 2.0, sobre formatos de arquivo, aplica-se na estratégia de migração de formatos de arquivo, ou seja, é altamente recomendável manter os documentos digitais a serem preservados em formatos de arquivo com especificações abertas ao público e adequadamente documentadas. Somente dessa maneira será possível possibilitar a efetivação de migração de formatos de arquivo no futuro.

5.0 Conclusões

A primeira conclusão necessária e importante de ser feita, diz respeito à complexidade das atividades de preservação de documentos digitais. Trata-se de problemas inter e multidisciplinares, problemas que requerem equipes formadas por técnicos em TI e disciplinas próximas ao documento, como arquivística, biblioteconomia, museologia e etc. Note-se também que o que foi abordado nesse texto é apenas uma pequena amostra de um universo muito maior de propostas científicas.

Concluimos também que não existe uma única abordagem para a preservação de documentos digitais. É necessário avaliar cada acervo e características específicas dos documentos e assim determinar qual a melhor ou melhores soluções aplicáveis. Ou seja, estabelecer a política de preservação de documentos digitais.

6.0 Bibliografia

- BODÊ(2), Ernesto. Preservação de Documentos Eletrônicos: O Formato de Arquivo com possível solução. Monografia de final de curso de biblioteconomia apresentada no Departamento de Ciência da Informação e Documentação da Universidade de Brasília em 2006.
- BODÊ, Ernesto. **Formatos de Arquivo e a Preservação de Documentos Digitais**. Trabalho apresentado no II Congresso Nacional de Arquivologia. Porto Alegre, 2006. Anais disponíveis em <http://www.aargs.com.br/cna/index.php?id=27>
- BUCKLAND, Michael K. What is a "document" ?. preprint 1997. Disponível no sítio do autor: <http://people.ischool.berkeley.edu/%7Ebuckland/index.html>. Acesso em 28 de novembro de 2007.
- BUCKLAND, Michael K. What is a "digital document" ?. preprint 1998. Disponível no sítio do autor: <http://people.ischool.berkeley.edu/%7Ebuckland/index.html>. Acesso em 28 de novembro de 2007.
- BYERS, Fred R. **Care and handling of CDs and DVDs**. Washington: Council on Library and Information Resources, 2003.
- CONTRERAS, Luis Núñez. Concepto de Documento. In: Archivística 2ª ed. Sevilla: Diputación. 1983.
- FERREIRA, Miguel. Introdução à preservação digital – Conceitos, estratégias e actuais consensos. Guimarães, Portugal: Escola de Engenharia da Universidade do Minho, 2006.
- HOEVEN, Jeffrey van der, WIJNGAARDEN, Hilde van, VERDEGEM, Remco, SLATS, Jacqueline. Emulation – a viable preservation strategy. The National Library of the Netherlands/koninklijke Bibliotheek. 2005. Disponível em http://dioscuri.sourceforge.net/docs/Emulation_report_KBNA_2005_en.pdf, acesso em 28 de novembro de 2007.
- LAWRENCE, Gregory W. et al. **Risk management of digital information: a file format investigation**. Washington: Council on Library and Information Resources, 1999. Disponível em <<http://www.clir.org/pubs/reports>>. Acesso em 10 de abril de 2006.
- MURRAY, James D. VanRYPEN, William. Encyclopedia of Graphics File Formats. Califórnia: O'Reilly & Associates, 1994.
- ROTHEMBERG, Jeff. Preserving Authentic Digital Information. In: Authenticity in a Digital Environment. Washington: Council on Library and Information Resources, 2000. Disponível em <<http://www.clir.org/pubs/reports>>. Acesso em 28 de novembro de 2007.

UNIVERSITY OF LEEDS. Survey and assesement of sources of information on file formats and software documentation. Final Report.

Understanding CD-R and CD-RW. California: Optical Storage Technology Association, 2003. Disponível em http://www.osta.org/technology/pdf/cdr_cdrw.pdf. Acesso em 28 de novembro de 2007.

VAN BOGART, John W.C. Armazenamento e manuseio de fitas magnéticas : um guia para bibliotecas e arquivos. Rio de Janeiro: **Projeto Conservação Preventiva em Bibliotecas e Arquivos**: Arquivo Nacional, 2001.