

PRESERVAÇÃO DE ACERVOS DOCUMENTAIS ELETRÔNICOS: DESAFIOS ALÉM DA CLIMATOLOGIA E DO ACONDICIONAMENTO

Ernesto Carlos Bodê (*)
 (*) Universidade de Brasília - Unb

Introdução

Hoje, além dos Acervos Documentais tradicionais já formados e ainda em formação, que utilizam suportes documentais como pergaminho, papiro e diversos tipos de papel (inclusive o fotográfico), encontramos novos suportes como fitas magnéticas (VHS, de rolo, cassete e outras), discos de vinil (e outros materiais), Discos do tipo HD, Disquetes de diversos formatos, Compact Discs (CD's) de diversos tipos como o CD-ROM, CD-R, CD-RW, DVD's também de diferentes tipos, além de outros.

De um paradigma ainda reinante e perfeitamente aplicável a determinados Acervos Documentais - tanto aqueles formados por documentos em suportes ditos tradicionais (papel, pergaminho, etc.) como parte daqueles em suportes de última geração - que compreende técnicas de controle ambiental (umidade, temperatura e etc.), acondicionamento e proteção contra pragas (insetos, etc.). Os profissionais da área de Conservação, Restauo e Preservação se vêem frente a problemas como obsolescência tecnológica de hardware e software (com suas várias nuances), o uso de formatos eletrônicos de arquivo em padrões da indústria ou proprietários, a separação entre conteúdo e suporte documental, além de questões como a necessidade de trabalho inter e multi disciplinar.

Em suma, o panorama atual é rico em desafios. Objetivamos neste trabalho, além de apresentar um breve resumo dos novos problemas (4), focar nossa atenção naquele que diz respeito à utilização de formatos eletrônicos padronizados e abertos de arquivo como alternativa para a preservação de grande parte dos Acervos Documentais que surgiram no último século e principalmente na segunda metade do século XX.

Materiais e Métodos

Com relação aos primeiros documentos, num trabalho sobre arquivística que aborda práticas de civilizações pré-clássicas, verificamos:

“O vale do Nilo e a Mesopotâmia permitiram a fixação de populações laboriosas que, desde o 4º milênio A.C., irão formar cidades e Estados organizados, onde a escrita desempenha já um papel primordial.” (7)

Mais adiante o mesmo texto menciona tratar-se, no caso, de *placas de argila com escrita cuneiforme.*

A respeito do papel, *“A invenção do papel representou uma inovação técnica tão importante como a da imprensa. Ambas são invenções da antiga China. O papel foi inventado em 105 d.C., por Cai Lun, alto funcionário da corte imperial, na dinastia Han (206 a.C.-220 d.C.), numa época que correspondia, no Ocidente, ao reinado do imperador Trajano. A vitória dos árabes contra os chineses, na batalha de*

Talas, perto do lago Balkhach, em 751 d.C., permitiu a captura de artesãos chineses conhecedores da técnica de fabricação do papel, o que levou os árabes a produzi-lo em Bagdá por volta de 793 d.C., A Europa só veio a fabricar papel no século XIV” (8)

E sobre outros suportes da antiguidade, *“Pedra e bronze foram os recipientes duros para a gravação das grandes mensagens da Antigüidade, da lei mosaica à Lei das XII tábuas. O papiro e o pergaminho serviram de modestos substitutos do papel, antes da invenção deste”* (8)

Na verdade, o pergaminho em seus diferentes tipos e o papiro parecem ter representando um papel bastante importante e não tão modesto assim na antigüidade. Vale lembrar também o a fabricação e uso do papel como o conhecemos hoje em escala industrial, foi um fenômeno que veio somente a acontecer no final do século XIX.

De qualquer forma, até final do século XIX, existiam basicamente documentos textuais, apesar de, desde a idade média, já encontrarmos ilustrações e iluminuras, além de outros adornos, inseridos em documentos textuais, principalmente em livros. Com a possível exceção da existência de mapas, ainda raros no período.

A partir do final do século XIX e início do século XX, começam a surgir as primeiras tecnologias para fabricação de fotografias (imagens fixas) e registros sonoros, evoluindo logo para os registros de imagens em movimento. As melhorias nas tecnologias de impressão permitirão também a elaboração de mapas dos mais variados tipos e oferecidos em escala ao mercado de consumo.

Estas primeiras tecnologias resultariam na formação de Acervos Documentais, em alguns casos conhecidos como de Multi Meios ou Audiovisuais, ao longo do século XX.

Pode-se falar então em documentação sonora, fotográfica, fílmica ou textual, além de microfilmes e microformas. De modo geral, estas são as possibilidades para armazenar informações (texto, imagem, som, cor e outras) em suportes físicos, ou seja, trata-se da definição comum de documento, *“Documento é o termo genérico que designa os objetos portadores de informação”* (6).

Ao longo do século XX e de maneira independente, ocorre em paralelo, duas grandes evoluções, a da eletrônica e a da computação. Ambas acabam por se unir em proveito próprio e hoje convivem simbioticamente, no que se chama hoje, de maneira geral, informática.

Note-se que o termo computar, do latim *computare*, estava inicialmente ligado à operação de contar ou seja, operações ligadas à aritmética e matemática. É natural

que o termo tenha sido aproveitado pelos computadores, pois estas eram as primeiras aplicações destes, principalmente com relação a atividades de contabilidade. Logo, os computadores passaram a lidar com outras informações além de números e somas.

De meados do século XX até seu final, observamos a utilização de documentos que contém informações muito diferentes e além do tipo textual. Cunhou-se o termo Arquivo de Computador e com a utilização de suportes como as fitas magnéticas e os discos rígidos e por último as tecnologias óptico magnéticas como os CD's e DVD's, além da evolução do poder de processamento e qualidade dos softwares disponíveis surgem os arquivos textuais, sonoros, de imagens fixas, de imagens em movimento. Surgem também arquivos complexos que contém diferentes tipos de informação, como texto, imagens em movimento ou fixas, som e etc.

Evidente é, que a evolução tecnológica possibilitou o surgimento de suportes documentais utilizados fora do mundo da informática, como as fitas de vídeo para gravações de programas de TV, as fitas domésticas do tipo VHS, além de várias outras para gravação amadora e profissional. Nosso foco neste trabalho, no entanto, está voltado para os Arquivos Eletrônicos e Suportes Documentais utilizados no âmbito da computação.

O resultado da vertiginosa evolução tecnológica que ocorreu durante todo o século XX e ainda continua ocorrendo neste início de século XXI, no âmbito da tecnologias de informação e comunicações, é o surgimento de todo um conjunto de tipos documentais em formato eletrônico. Independentemente do conteúdo, ou seja, da informações registradas, qualquer documento que existia em suportes documentais tradicionais, como o papel, pode assumir um determinado Formato Eletrônico.

Alguns autores propõem inclusive, a digitalização de documentos em suportes tradicionais como o papel para a preservação por longos períodos destes documentos, *“a tecnologia da imagem digital pode ser usada para gerar uma cópia de alta qualidade do item original. Limitando o acesso direto a documentos valiosos, a formação de imagem digital torna-se um ‘elemento de preservação’, distinto de um ‘elemento de acesso’”* (2)

Cabe aqui também lembrar que hoje os Arquivos Eletrônicos que contém documentos, principalmente Arquivos Sonoros, também estão disponíveis para uso em outros equipamentos além de computadores, é o caso de dispositivos portáteis como o iPod, MP3 players, Pen Drives e outros similares.

Aliás, estes dispositivos portáteis são uma excelente prova do fenômeno relacionado aos Formatos Eletrônicos de Arquivo, que chamo *Divórcio entre Suporte e Conteúdo*.

Nos documentos que utilizam suportes tradicionais como o papel, papiro, filmes e etc., havia uma necessária união entre o suporte e aquela informação gravada, o conteúdo informacional não podia ser retirado do suporte original, sem prejuízos ao documento. Note-se que não estamos falando de cópias dos originais.

As exceções relevantes no mundo todo começaram a surgir a partir do uso das primeira fitas magnéticas, quer seja para gravação de áudio, vídeo ou ambos. Com o uso destes suportes documentais o conteúdo de uma fita magnética pode integralmente ser transferido para outra fita magnética.

Os discos de computadores como os HD (hard-disc), os CD's e DVD's também possibilitam este recurso.

Os novos Documentos (Arquivos Eletrônicos) podem migrar de um suporte inicialmente utilizado para um novo, não necessariamente igual ao anterior. E mesmo assim mantendo suas características originais de autenticidade e integridade. Claro que, contanto que observados os procedimentos adequados.

Resultados

Mas o que é um Formato de Arquivo, Brown num trabalho sobre o tema registra *“Formatos Eletrônicos codificam informações numa forma que somente pode ser processada e tornada compreensível através de uma combinação específica de hardware e software”* (1).

Em outras palavras e relacionando com Documentos, este enunciado nos diz que o conteúdo de um documento (texto, som, imagem, a formatação e suas combinações), através do uso de um *determinado* Formato de Arquivo é transformado para uma outra linguagem específica deste Formato, linguagem que precisa de certo hardware e software para ser novamente inteligível. Isto vale tanto para documentos que nascem no computador, como um texto digitado e formatado num editor de texto comum ou para um documento manuscrito em pergaminho e digitalizado para o computador.

Note-se que utilizamos o termo *“através do uso de um determinado Formato”*, com isto, queremos enfatizar que existem vários Formatos de Arquivo. Isto ocorre por várias razões, as principais estão relacionadas à concorrência entre produtores de software e tipo de conteúdo. Quando uma empresa de software cria um programa para digitar e formatar um texto na tela do computador, ela guarda esta tecnologia contra a concorrência, entre várias informações, como a linguagem utilizada e o próprio código fonte do software, está a estrutura de codificação para gerar o Formato de Arquivo, que será utilizado para gravar os textos gerados nos discos (ou similares) do computador. Desta maneira, outra empresa que também fabrique um software com exatamente a mesma finalidade, ou seja, digitar e formatar textos, desenvolverá outra codificação e outro Formato de Arquivo. Assim temos o formato *.doc*, gerado pelo aplicativo *word* da *Microsoft*, o *.pm6* gerado pelo aplicativo *PageMaker* da *Adobe*.

Por outro lado, o tipo de conteúdo também obriga a estruturar Formatos de Arquivo de maneiras diferentes, são situações bem diferentes, codificar caracteres digitados na tela, pontos de uma imagem digitalizada e som. As estruturas internas de codificação destes Formatos é bastante diferente. Teríamos formatos como o *.doc*, o *.jpg* ou o *.mp3*, respectivamente.

Por quê falar em Formatos de Arquivo como alternativa para Preservar documentos, quer sejam em formatos tradicionais digitalizados para uso no computador ou nascidos no computador ? Afinal de contas, não se trata bem de uma alternativa, para uso no computador um documento terá que estar, obrigatoriamente, codificado dentro de um determinado Formato de Arquivo.

A resposta a esta pergunta não está em *Por quê* usar o Formato de Arquivo como alternativa para preservação mas em *Como* utilizá-lo. Já que certos cuidados podem ser tomados na escolha do Formato de Arquivo mais adequado para efeitos de preservação por longos períodos, em séculos ou até *ad eterno*.

Bem, quais seriam os problemas que impedem a efetiva preservação de um documento, por longos períodos, com o uso de um determinado Formato de Arquivo ?

Primeiro, vamos analisar o que pode sair errado em condições normais de uso de arquivos no computador, tem termos de preservar os documentos por longos períodos. Algumas possibilidades que tem sido comumente apontadas na literatura são:

Orfandade – Não é raro que uma nova versão de software tenha algum tipo de problema para abrir uma versão antiga gerada pelo mesmo aplicativo. Com o passar dos anos, há uma possibilidade real de perda dos documentos contidos nos arquivos mais antigos da corporação.

Obsolescência – Em função dos avanços nas tecnologias da informação e comunicações, novos aplicativos surgem e outros desaparecem do mercado, sem deixar a possibilidade de aquisição para todos aqueles usuários que o utilizaram e geram centenas ou milhares de arquivos no computador. A longo prazo, nada garante que um aplicativo hoje comum e largamente utilizado estará disponível em 50 ou 60 anos e, se estiver, se funcionará com os sistemas operacionais e hardwares do futuro.

Identificamos que as pesquisas realizadas e a literatura sobre o tema enfoca três possíveis soluções para estes problemas.

A primeira delas é a adoção de padrões de indústria para o uso de Formatos de Arquivo. Um bom exemplo neste sentido foi a recente criação de uma norma ISO que transformou o formato de arquivo .pdf (Portable Document Format) em padrão internacional e aberto à indústria. As vantagens desta abordagem estão relacionadas a eliminar os problemas legais de propriedade para desenvolver aplicativos que possam ler documentos em arquivos de computador gerados em formatos de arquivo proprietários, o que pode ser necessário no futuro para resgatar documentos antigos. Ainda como vantagem, por ser um padrão internacional e aberto, toda a informação necessária para desenvolver um aplicativo compatível com determinados arquivos de computador, estará facilmente disponível no futuro.

Há limites nesta abordagem, no entanto. Primeiro, não existem muitas normas padrão no mercado. Além do formato .pdf, existem alguns formatos gráficos de imagem, basicamente. Note-se que o formato .pdf, por exemplo, é um padrão aberto para determinados tipos

de Formato de Arquivo (principalmente documentos textuais), não foi projetado para conteúdo de som e imagem em movimento, por exemplo, apesar de ser possível sua utilização com incorporação de outros arquivos. A grande maioria dos formatos de arquivo utilizados no mercado são proprietários e não podem hoje e não há garantias de que poderão no futuro, ser consultados para desenvolver aplicativos que leiam os formatos em uso no presente. No entanto, o tema vem ganhando maior importância e, parece ser razoável esperar que, cada vez mais, surjam novos Formatos de Arquivo padrão no mercado.

Existe inclusive uma proposta de um formato universal de arquivo, esta proposta pretende resolver a preservação do Formato e a diversidade de conteúdos encontrados em arquivos de computador, *“O UPF [Formato Padrão de Preservação] foi projetado para ser independente das aplicações de computador utilizadas para criar conteúdos, independente do sistema operacional de onde estas aplicações foram originadas e independente da mídia física onde estes conteúdos foram armazenados.”* (9)

Uma segunda abordagem refere-se a procedimentos de Emulação de software e hardware, basicamente, a proposta é criar programas capazes de, no futuro, imitar ou emular aplicativos antigos, inclusive os próprios sistemas operacionais utilizados hoje.

Esta proposta, que em princípio, parece bastante razoável, na verdade, traz problemas sérios de implementação. Na prática, não é fácil desenvolver estes softwares emuladores, além do mais, existe uma variedade imensa de aplicativos e versões de sistema operacional. Para complicar, alguns formatos de arquivos são proprietários e não se tem acesso à estrutura do formato, impossibilitando assim o desenvolvimento de um bom emulador.

Por último, uma outra linha de atuação para a preservação dos documentos em Formatos de Arquivo é a migração de formatos. 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 em termos de software e hardware.

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 que gerando resultados incertos”* (3).

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*, e 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”*;

O segundo, *Riscos associados com os dados do formato de arquivo*, incluem *“os elementos da estrutura interna do formato de arquivo sujeitos a modificação”*;

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 súbitos*”.

Finalmente é preciso notar 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.

Estes conteúdos também podem ser relevantes para um futuro historiador preocupado em entender nossa cultura presente, e portanto, preocupações com preservação são pertinentes. No entanto, trata-se de uma questão complexa, pois além de simples arquivos de computador, estas páginas estão associadas a softwares para leitura (browsers) e para execução (plugins), além dos equipamentos e softwares servidores de rede. E portanto, é um assunto que merece um estudo à parte.

Conclusões

Um estudo da evolução tecnológica dos documentos e seus suportes, ao longo da história da humanidade, mostra uma evolução constante. Os documentos sempre foram parte essencial, e cada vez mais o são, para a humanidade.

No processo de evolução de documentos e seus suportes, várias características marcantes podem ser destacadas, primeiro a diversidade de conteúdos registrados, de uma era basicamente formada por documentos textuais, evoluímos para conteúdos de som, imagem, e diversos outros. Além disto hoje, com a utilização de documentos eletrônicos-digitais, dentro de arquivos de computador, estruturados em formatos de arquivo específicos, ocorre uma separação entre os suportes físicos e seus conteúdos. Os documentos em arquivos de computador, ainda precisam ser gravados em suportes físicos, no entanto, podem ser gravados em suportes diferentes em termos tecnológicos e inclusive é possível gerar várias cópias idênticas dos mesmos documentos em locais diferentes.

Na base desta evolução, está a figura do Formato de Arquivo. Dependendo do fabricante do aplicativo, sistema operacional e conteúdo (som, texto, imagem e etc.) haverá um Formato de Arquivo específico no mercado.

Se por um lado, a figura do documento tradicional transposta para arquivos estruturados em Formatos de Arquivo, pode parecer uma solução em termos de preservação de documentos para gerações futuras, esta é uma impressão errada. Diversos problemas novos surgem com o uso de Formatos de Arquivo em documentos.

Além dos documentos atuais, nascidos no computador e que nunca tiveram um suporte físico tradicional, como o papel. A digitalização de documentos antigos em suportes como papel, pergaminho e etc, vem sendo apontada como uma

solução para a preservação destes documentos para as futuras gerações.

No entanto, a digitalização e migração de suporte sobre um objeto de conservação, digamos um manuscrito da idade média, resolve vários problemas de conservação e até outros como a questão do acesso a estes documentos preciosos.

Porém, o arquivo gerado (dentro de um determinado Formato de Arquivo específico) passa a ser um novo objeto de conservação.

Manter este arquivo operacional para as gerações futuras não é tarefa fácil e deve ser contemplada por uma Política de Preservação de Arquivos no mundo digital.

Vários caminhos podem ser trilhados neste sentido, como a adoção de Formatos de Arquivo padrão (normas ISO), procedimentos de Emulação de software e hardware e Migração de formatos.

Não parece existir uma solução única que funcione como uma panacéia para o problema. Nossa opinião é de que é preciso um estudo de cada Acervo Documental e um Planejamento que leve em consideração as soluções possíveis e aplicáveis.

Referências

- (1) Brown, Adrian. Selecting File Format for Long Term Preservation. UK: The National Archives, 2003.
- (2) Conway, Paul. *Preservação no Universo Digital*. Rio de Janeiro: Projeto Conservação Preventiva em Bibliotecas e Arquivos: Arquivo Nacional, 2001.
- (3) Lawrence, Gregory W. et al. Risk management of digital information: a file format investigation. Washington: Council on Library and Information Resources, 1999.
- (4) Munoz-Vinas, Salvador; Contemporary theory of conservation. Oxford: Butterworth-Heinemann, 2004.
- (5) 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.
- (6) Le Coadic, Yves-François. *A ciência da informação*. Brasília: Briquet de Lemos, 2004.
- (7) Silva e Ribeiro. *Arquivística*. Lisboa: Afrontamento, 2000.
- (8) Doctors, Márcio. *A cultura do papel*. Rio de Janeiro: Casa da Palavra, 1999.
- (9) Shepard, Thom. *The universal preservation format*. Boston: WGBH [sd]

E-Mail do Autor

bode@stj.gov.br