

# **PRESERVAÇÃO DE COLEÇÕES DE DOCUMENTOS DIGITAIS**

**Artigo apresentado no  
Seminário Internacional de Bibliotecas Digitais  
São Paulo/2007**

**Título:**

**PRESERVAÇÃO DE COLEÇÕES DE DOCUMENTOS DIGITAIS**

**Resumo:**

O trabalho apresenta conceitos e estratégias para a consecução de cuidados para a preservação de coleções de documentos digitais. Inicialmente, apresenta-se a importância das atividades de preservação de coleções digitais. Em seguida, são apresentados conceitos relevantes relacionados ao tema, como o próprio conceito de documento digital, questões terminológicas são abordadas também. Finalmente, são apresentados elementos para compreensão e implementação de estratégias de preservação de coleções de documentos digitais. Conclui-se pela importância do tema e necessidade de trabalho conjunto entre os responsáveis pela preservação de coleções e especialistas em áreas ligadas à tecnologia da informação.

**Palavras-chave:**

Documentos eletrônicos, preservação, coleções, fidedignidade, autenticidade, formatos de arquivo, obsolescência tecnológica.

**Abstract:**

This article presents concepts and plans for preservation of digital collections. On the beginning it presents the importance of preservation for digital collections. Following it it introduces important concepts on preservation, including the digital object concept itself, terminological points are reached too. Finally it presents elements to a better understanding and deploying of digital objects preservation. Concluding, it reinforces the importance of the subject and the necessity of a common work between the responsables for digital objects collection preservation and TI experts.

**Key-words:**

Electronic documents, preservation, collections, reliability, authenticity, file formats, technological obsolescence.

**1.0 Introdução**

Este trabalho apresenta os resultados parciais de um projeto de pesquisa de mestrado acadêmico relacionado ao tema preservação de documentos eletrônicos e digitais. No presente artigo, apresentamos os resultados já obtidos no que se refere a conceitos, terminologia e, principalmente, políticas e recomendações para a adequada preservação a longo prazo de documentos eletrônicos digitais. Embora nossa pesquisa tenha como escopo documentos eletrônicos digitais de diversas naturezas, neste trabalho, procuramos enfocar aspectos referentes à preservação de documentos eletrônicos digitais de caráter biblioteconômico e, conseqüentemente, acervos de coleções em bibliotecas digitais.

Utilizamos diversas fontes de informação para subsidiar nosso referencial teórico. Entre elas, é importante ressaltar o material produzido pelo *RGL* e *OCLC*<sup>1</sup> e a *Library of Congress*<sup>2</sup>, todas nos EE.UU. No Canadá, o projeto *InterPares*<sup>3</sup>, hoje na fase InterPares

---

<sup>1</sup> O *RLG* (Reserch Library Group, <http://www.rlg.org>) e o *OCLC* (*Ohio College Library Center*, <http://www.oclc.org>) são redes que incluem bibliotecas, museus, arquivos e outras instituições culturais e fomentam diversas iniciativas de pesquisa nos Estados Unidos e em outros países.

<sup>2</sup> A *Library of Congress* (Biblioteca do Congresso estadunidense) possui uma divisão específica para a preservação de documentos digitais: <http://www.digitalpreservation.gov>

2. Temos também o Arquivo Nacional<sup>4</sup> do Reino Unido, com um programa específico para a preservação digital. Todos estes grandes centros congregam pesquisadores e especialistas em diversas áreas que têm produzido artigos e trabalhos sobre o tema preservação digital. No Brasil, já encontramos trabalhos acadêmicos que abordam o tema da preservação digital. Destacamos a tese de *Kátia de Paula Thomaz* (2004) no Programa de Pós-Graduação em Ciência da Informação da UFMG e a dissertação de mestrado de *Humberto Celeste Innaelli* (2006) na Universidade Estadual de Campinas. Há outros centros de pesquisa e autores, tanto nacional como internacionalmente, e serão citados no transcorrer de nosso texto.

## 2.0 – O Documento

Neste trabalho nos interessa o conceito de documento utilizado no seguinte campo semântico: “*Qualquer base de conhecimento, fixada materialmente e disposta de maneira que se possa utilizar para consulta, estudo, prova, etc.*” (FERREIRA, 1986, p. 605). Este conceito amplo é fundamental na história da cultura da humanidade e, mais acendadamente ainda, em nossa pós-modernidade. Ganha contornos específicos dependendo dos diferentes pontos de vista. Assim, nas ciências ligadas à administração de organizações, por exemplo, ressalta-se seu valor na tomada de decisões e como fator de comunicação mais ou menos eficaz. Aqui, nos interessa o viés ligado à ciência da história e às ciências da documentais que se preocupam com sua organização e tratamento. Mais especificamente, nos interessa a característica que documentos podem assumir de se tornarem instrumentos para a preservação da memória de indivíduos, organizações e em última análise até mesmo da humanidade.

Uma definição mais adequada para este trabalho é a seguinte:

*Documento em um sentido bem amplo e genérico é todo o registro de informação, independentemente de seu suporte físico. Abarca tudo que pode transmitir o conhecimento humano: livros, revistas, fotografias, filmes, microfilmes, microfichas, folhas, transparências, desenhos, mapas, informes, normas técnicas, patentes, fitas gravadas, discos partituras, cartões perfurados, manuscritos, selos, medalhas, quadros, modelos, facsímiles e, de maneira geral, tudo que tenha um caráter representativo nas três dimensões e esteja submetido à intervenção de uma inteligência ordenadora.* (HEREDIA HERRERA, 1991, p. 122)

Por outro lado, uma definição ainda mais precisa e estruturada obtida a partir de diversos autores e instituições:

*Pode-se definir documento genérico como qualquer informação registrada independentemente do suporte utilizado, a qual pode ser tratada como unidade. No primeiro nível de desdobramento, é possível distinguir-lhe dois elementos constituintes, a saber: o suporte, o meio físico sobre o qual a informação é fixada; e a mensagem ou notícia veiculada. No segundo nível, a mensagem pode ser decomposta em outros três elementos, quais sejam: a estrutura sobre a qual a informação foi registrada, envolvendo cabeçalhos e outros dispositivos para identificar e rotular partes do documento, negrito, itálico etc.; o conteúdo, propriamente dito; e o meio de fixação desse conteúdo com possibilidades para o texto, o gráfico, a figura, a tabela etc.* (THOMAZ, 2004, p. 77)

Em todas as definições acima e em várias outras que podem ser encontradas em diversas áreas da literatura científica, dois aspectos merecem destaque.

O primeiro, com relação à explicitação da presença e necessidade de um suporte físico para a existência de um documento qualquer. De fato, este é um ponto fundamental, principalmente quando se discorre sobre documentos digitais, área onde tem surgido algumas confusões terminológicas. Na atualidade, com o uso de tecnologias de rede e

---

<sup>3</sup> O projeto InterPares (*The International Research on Permanent Authentic Records in Electronic Systems*) é promovido por instituições do Canadá, EE.UU e a UNESCO e divide-se atualmente em duas fases InterPares 1 e InterPares 2, <http://www.interpares.org>.

<sup>4</sup> *The National Archives of United Kingdom*, <http://www.nationalarchives.gov.uk/>.

acesso a documentos em nossas telas de computador é possível que se tenha a sensação de que alguns documentos são como que etéreos, estão em algum tipo de espaço imaterial. De fato, não é difícil encontrar termos como *documento virtual*. É preciso que se esclareça, desde já, que não existe nenhum documento - eletrônico, digital ou de qualquer outra designação – que não esteja fixado em algum tipo de suporte físico, mesmo que em algum lugar numa rede ou em algum tipo de memórias internas de computadores.

E o segundo aspecto, documentos contêm informações e conhecimento. E estes conteúdos representam, de diferentes maneiras e formas, a memória que precisa ser mantida por mais ou menos tempo para cumprir diversas finalidades, dependendo de seu valor administrativo, cultural ou histórico. E mais importante, esta memória não está apenas em um tipo específico de documento, como o documento arquivístico histórico ou manuscritos escritos em pergaminhos da idade média. Esta memória está também inscrita em prosaicos livros modernos, gravações de áudio, fotografias e etc<sup>5</sup>. No romance de H. G. Wells, a *Máquina do Tempo*, recentemente produzido em versão cinematográfica<sup>6</sup>, há uma belíssima demonstração de como prosaicos livros de uma biblioteca tem seu valor para a humanidade. O personagem viajante no tempo alcança 800 mil anos após o final do séc. XIX e, nas cenas finais da versão do cinema, podemos ver o que resta de uma humanidade voltando a receber conhecimento de uma biblioteca do passado, diga-se de passagem em versão com acervo digital.

No entanto, trata-se de ficção científica e como veremos mais adiante os documentos digitais, pelo menos dentro da tecnologia que dispomos hoje e com os cuidados e políticas que têm sido implementados, não são tão duráveis assim e, na verdade, estão em sério risco de perda irremediável.

## 2.1 - O Documento Digital

Até que a humanidade obtivesse sistemas de escrita completos como os atuais, ela fez uso de símbolos gráficos e mnemônicos de vários tipos para armazenar informações, sobre um dos mais antigos artefatos encontrados:

*Artefatos desenterrados em Bilzingsleben, Alemanha, datados de pelo menos, 412.000 anos atrás [...] foram interpretados por seus descobridores como entalhes intencionais (algum tipo de símbolos gráficos). É evidente que os entalhes são marcas; o que significam e se significam algo, não está claro. (FISCHER, 2003, p.16)*

Ainda segundo este mesmo autor, a humanidade utilizou então sistemas pictográficos (como as representações em cavernas). Em um segundo momento, passou a utilizar símbolos gráficos para representar coisas reais como bens e animais, até o grande salto da fonetização, quando um símbolo gráfico representa um som correspondente na linguagem local. Tal invenção surgiu na Mesopotâmia entre 6.000 e 5.700 anos atrás, aproximadamente.

Como suportes para registros de informações a humanidade utilizou os mais diversos materiais, segundo HUNTER (1978), foram utilizados a madeira, metais, pedras, troncos, tecidos, o papiro (*Cyperus papyrus*), pergaminho e finalmente o papel.

O uso de argila em tábuas é particularmente importante, pois ao que parece o primeiro sistema completo de escrita (por volta de 2.500 AC) utilizou este material como suporte. No Egito, o uso do papiro (o mais antigo conhecido tem 3.700 anos) rivalizou com o uso de tábuas de argila. Na verdade os egípcios desenvolveram diferentes sistemas de escrita para diferentes aplicações, rituais, contabilidade e etc., para cada aplicação havia um sistema de escrita e suportes específicos como paredes, metais preciosos e etc.

<sup>5</sup> Inserir referência aos textos Fontes Históricas.

<sup>6</sup> The Time Machine. Direção Simon Wells. DreamWorks SKG / Warner Bros. 2002.

O uso do pergaminho também foi um fato importante para o registro de informações, “*O rei de Pérgamo (197-159 AC) normalmente recebe os créditos pela invenção e acredita-se que esteja relacionada com o desejo de produzir um material de escrita que rivalizasse com o papiro egípcio*” (HUNTER, 1978, p.12).

Finalmente, surge o papel que possibilitou um grande salto na produção de documentos, já que tratava-se de um material de fácil fabricação e menor custo, além da qualidade em relação a outros suportes. A data normalmente atribuída à invenção do papel é a de 105 DC, na China (HUNTER, 1978, p.50).

Vários outros suportes foram utilizados para o registro de documentos e em determinados períodos históricos alguns competiram entre si, como o papel e o pergaminho. O tipo de papel próximo do que é utilizado hoje só existiu a partir do século XIX (DOCTORS, 1999).

Temos então, até o século XIX, uma produção documental, registrada basicamente em papiro, pergaminho e papel, documentos com conteúdo textual, de diferentes naturezas, de inventários de bens até a literatura e a filosofia. Em meados do século XIX surge uma invenção que acrescenta uma nova diversidade aos acervos documentais, trata-se da fotografia, “*A invenção da fotografia foi anunciada oficialmente em 19 de agosto de 1839, pelo francês Louis Jacques M. Daguerre (1787-1851), sob a forma do daguerreótipo*” (SMIT & GONÇALVES, 2005, p.9). Esta invenção passaria por um processo de evolução tecnológica que culminaria, no final do século XX com o advento da fotografia digital, a qual, por si só, tornou-se uma nova revolução. Também com tecnologia bastante próxima dos registros fotográficos, apesar da aplicação diferente, encontramos também o microfilme como meio para registro documental, ainda hoje bastante utilizado.

No final do século XIX, vários inventos para registro do som culminaram no início do século XX com os discos com gravações sonoras, e logo depois, o uso também de fitas magnéticas. Estas últimas, após um período de evolução, também 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 e depois surgindo os modelos específicos para vídeo (DVD's).

O próximo grande passo seria dado pelo uso de computadores pela humanidade. Os primeiros computadores modernos apareceram na década de 40, embora haja muitas contribuições individuais para o avanço da tecnologia, esta cresceu e se desenvolveu, na América do Norte, especialmente “*graças à associação entre militares, universidades e firmas*” (KIDDER, 1981, p. 13). O uso, cada vez maior de computadores, inicialmente pelas grandes corporações, mas a partir da década de 80 do século XX, também pelo cidadão comum, representou um grande salto para o registro, armazenamento e recuperação de documentos. Estas máquinas, em função da exigência de cada vez mais espaço para registro de seus *bits* (codificação digital), passaram a utilizar diferentes tecnologias, desde as fitas magnéticas, passando por discos magnéticos, ópticos e diversos outros. Hoje, no início do século XXI, as novidades incorporadas ao conjunto de mídias são os *tocadores de áudio*, *pen-drives* e outros.

Apresentamos a seguir um quadro comparativo entre as características dos diferentes documentos apresentados no breve histórico acima.

Documentos		
Período	Suporte utilizado	Mensagem
Antes 6.000 AC	Material disponível na natureza: ossos, cascas de animais, madeira.	Sinais, desenhos e marcas mnemônicas.
6.000 AC e Final séc. XIX	Material elaborado para uso específico: Argila, pergaminho, papiro e o papel.	O conteúdo da mensagem se apresenta estruturado mas apresentado sob a forma de texto ou ilustrações e pinturas.
Desde o final séc. XIX até Hoje	Material elaborado para uso específico com maior grau de sofisticação: papel moderno, películas, mídias magnéticas e ópticas.	Além do conteúdo estruturado, há uma miríade enorme de formas de apresentação além do textual: imagens fixas e em movimento, bancos de dados, planilhas e etc.

Fig. 1 – Quadro comparativo das fases de evolução do documento.

A partir da análise das informações do breve histórico acima exposto, podemos tecer algumas considerações. Desde o final do séc. XIX começam a surgir os primeiros documentos, que podemos chamar hoje, ainda impropriamente, de eletrônicos. Na verdade, estas primeiras tentativas de armazenar conteúdo informacional além do texto<sup>7</sup>, como o som e imagens em movimento merecem a designação, neste período, de documentos legíveis por máquina<sup>8</sup> (fonógrafos e projetores de cinema, por exemplo), já que a eletrônica propriamente dita ainda nem mesmo existia, vindo a se desenvolver ao longo do séc. XX. De fato, a característica de se **necessitar de máquinas** para se ter acesso ao conteúdo destes documentos é uma transformação tecnológica importante e que traz conseqüências para a preservação destes documentos, sendo a mais óbvia a necessidade de manutenção destas máquinas juntamente com seus documentos. Mesmo com o desenvolvimento de suportes mais sofisticados para registro de informações, como os discos vinil e o desenvolvimento da eletrônica propriamente dita, numa primeira fase, estes documentos ainda têm uma característica em comum com todos os outros até então produzidos: o conteúdo da mensagem não pode ser dissociado do suporte físico utilizado, não sem danificar o documento. Somente a partir do desenvolvimento das fitas magnéticas (para áudio ou vídeo) começa a surgir o fenômeno da **independência** entre o suporte e o conteúdo do documento, característica esta que trará conseqüências importantes do ponto de vista das ações de preservação para estes documentos, como indicaremos mais adiante<sup>9</sup>. Através do uso dos computadores, desde meados do séc. XX, desenvolvem-se as tecnologias digitais aplicáveis aos documentos. Neste ponto, passamos a ter o uso de

<sup>7</sup> A fotografia foi uma bem sucedida de registro de imagens reais, com característica bem diferenciadora das pinturas da época, por mais realistas que fossem. Apesar de não necessitar de equipamentos tecnológicos para sua “leitura” e somente para sua produção hoje, com a fotografia digital, este quadro está mudando. A fotografia digital requer computadores e máquinas para sua visualização adequada. Também é importante notar que o microfilme (basicamente um processo fotográfico) só é legível através de equipamento específico para sua leitura.

<sup>8</sup> Observamos que na atualidade, a característica *Legível para Máquinas*, por si só, não define que um documento seja eletrônico. O acesso a manuscritos antigos (digitalizados) no outro lado do planeta via rede de dados e computadores (legível neste lado do planeta através destas máquinas), não transforma o manuscrito antigo em documento eletrônico (apesar da existência de uma cópia deste codificada digitalmente).

<sup>9</sup> É curioso observar que alguns tipos de bases de dados (no todo e dentro das definições propostas podem ser consideradas um único documento) podem estar armazenadas em suportes físicos diferentes, estar divididas em diferentes discos em diferentes computadores, ou parte dos dados em fitas magnéticas ou discos ópticos off-line por exemplo (em prateleiras onde são inseridos no equipamento conforme a demanda de informações).

eletrônica digital e armazenamento de conteúdos sob a forma de codificação digital<sup>10</sup>, na atualidade qualquer tipo de conteúdo além do texto, como o som e imagens. Surge então o documento **eletrônico digital**. Nós defendemos que a terminologia mais adequada seja documento eletrônico e digital, já que há documentos eletrônicos que não utilizam tecnologia digital (como as fitas magnéticas com registros de história oral, em gravações eletrônicas analógicas) e há inclusive documentos com codificação digital (legíveis por máquinas eletrônicas) mas gravados em papel, é o caso dos cartões perfurados ou *punch cards* para entrada de dados em computadores *mainframes*, já ultrapassados.

Independentemente dos rótulos terminológicos utilizados, nos parece que o mais importante é definir as características essenciais destes documentos eletrônicos e digitais. Aqui, esclarecemos que este texto tem como escopo este grupo específico de documentos, ou seja, aqueles documentos que surgiram no final do séc. XX e tiveram seu grande avanço qualitativo e quantitativo com o advento da microinformática. As referidas características essenciais são: *Legibilidade por Máquinas, Independência entre Suporte Físico e sua correspondente Mensagem* e, finalmente, o fato de serem *Codificados em linguagem Binária Digital*.

Em se tratando da correta caracterização de documentos eletrônicos e digitais, não poderíamos também de deixar de mencionar uma outra informação importantíssima: a quantidade existente destes documentos em relação aos demais. Segundo levantamentos citados em artigo no *Information Management Journal*, atualmente produzimos cerca de *161 exabytes* de informações digitais, para ilustrar o que significa esta quantidade de informações ela equivale a “*três vezes a informação contida em todos os livros já escritos*” ou “*12 pilhas de livros que alcançariam da terra, cada uma, o sol*” (IM, 2007, p. 8). Portanto, a presença dos documentos eletrônicos e digitais como representantes de nossa cultura e modo de vida hoje é muito relevante e estes dados apresentados enfatizam a importância da preocupação com políticas de preservação de, pelo menos, uma parte deste imenso patrimônio da humanidade.

## 2.2 – A Preservação no Mundo Digital

Os termos preservação e preservação digital trazem alguma confusão terminológica, de fato, o próprio termo preservação, por si só, traz alguns problemas.

Segundo (MUÑOZ VIÑAS, 2005, p. 7) o termo conservação pode se referir a dois sentidos, o primeiro deles em oposição a *restauração*, um sentido mais restrito de atividades e o segundo como a soma das atividades do primeiro sentido mais *restauração* e outras atividades correlatas, tais como pesquisa histórica e apoio administrativo. Ainda segundo Muñoz, os problemas de tradução entre línguas latinas e outros ramos linguísticos são responsáveis pelas imprecisões como o uso do termo preservação como conservação e vice-versa. Segundo Conway, “*Preservação [preservation] é a aquisição, organização e distribuição de recursos a fim de que venham a impedir posterior deterioração ou renovar a possibilidade de utilização de um seletivo grupo de materiais*” (CONWAY, 2001, p. 14), ou seja, conservação num sentido amplo.

Tradicionalmente, os termos preservação, conservação e restauração estão associados aos documentos em suportes tradicionais, através dos quais estas disciplinas se

---

<sup>10</sup> Inicialmente dados computacionais, números e textos e mais recentemente (final séc. XX) com o advento do que se convencionou chamar multimídia: som, imagens fixas e em movimento e a combinação de todos estes elementos.

desenvolveram<sup>11</sup>. Quando se fala em atividades de preservação para documentos em suportes tradicionais serão inevitáveis considerações sobre as características dos suportes físicos, como a climatização do ambiente (temperatura e umidade relativa do ar), controle de pragas biológicas, luminosidade e outros fatores. Talvez devido ao desenvolvimento da disciplina preservação que ocorreu com documentos em suportes tradicionais (os documentos digitais são um fenômeno recente) e toda a pesquisa desenvolvida e publicada, quando se fala em preservação de documentos eletrônicos e digitais a associação com suportes físicos é muito comum.

O que em parte está correto. Primeiro, porque como já defendemos antes, não existem documentos (mesmo os eletrônicos digitais) sem um suporte físico onde em algum lugar estão gravados. Portanto, existindo suportes físicos, as preocupações com seu estado de conservação são necessárias e importantes, quanto mais importante o documento mais importante será o cuidado com sua preservação. Segundo, porque ainda existem e espera-se que continuem a existir por um longo período, documentos eletrônicos indissociáveis de seu suporte físico, como os discos vinil e alguns tipos de fitas magnéticas. Nestes termos, é natural da ênfase às atividades de preservação que levam em consideração somente aspectos dos suportes físicos.

No entanto, infelizmente, no universo digital a extensão do problema preservação é muito mais ampla do que ocorre com documentos em suportes tradicionais. Com relação ao suporte físico, defendemos que este seja talvez o menor dos problemas. Como já demonstramos antes, os documentos eletrônicos e digitais têm como importante característica a dissociação entre suporte físico e conteúdo informacional, desta forma, é perfeitamente possível o descarte do suporte físico e manutenção do conteúdo informacional em um novo suporte (igual ao original ou de outra tecnologia). Para ilustrar, considere uma base de dados fornecida em CD-ROM com todas as edições e textos completos de um determinado periódico. Prevendo-se o desgaste deste disco CD-ROM com o tempo, poderíamos (re)gravar todo seu conteúdo em um novo disco CD-R ou em fitas magnéticas. E repetir este processo quando do desgaste destes discos. Sempre preservando-se o conteúdo original para consultas futuras. Estes procedimentos não impedem cuidados no armazenamento e climatização destes documentos até que se faça a transposição de seus conteúdos para outros suportes, já que defeitos nas mídias originais podem acarretar a perda do conteúdo informacional destes documentos. No universo digital, podemos dividir as atividades de preservação entre dois grandes grupos: o físico e o lógico.

Os **cuidados físicos** incluem cuidados com os suportes físicos, e incluem também a preocupação com a obsolescência tecnológica de hardware (que veremos mais adiante), ou seja, garantir que juntamente com estes documentos (somente legíveis por máquina) existam equipamentos apropriados para sua leitura (leitor de disquete, de CD, fita magnética e etc.), pelo menos até uma eventual transposição para outro tipo de suporte com tecnologia atualizada.

Já os **cuidados lógicos** são um pouco mais complexos. Como já expomos antes, os documentos eletrônicos e digitais têm seu conteúdo codificado em linguagem binária digital, essencialmente tudo está gravado através de marcas que correspondem aos bits 0 e 1, ou seja, os dados que efetivamente podem ser lidos. Caberá a vários programas de computador (*software's*) a interpretação adequada desta codificação. Por exemplo, para a leitura de uma planilha de cálculos é necessário o aplicativo compatível para sua leitura o qual necessita de um sistema operacional correspondente que por sua vez

---

<sup>11</sup> Claro é que estas disciplinas também trabalham com outros objetos que não os documentos em suportes tradicionais, como esculturas e até edificações.

necessita de toda uma gama de recursos (driver's, dll's<sup>12</sup> e etc.) para funcionar adequadamente. Desta forma, garantir o acesso a estes documentos envolverá também, de alguma maneira, especificar como a codificação binária de zeros e uns está estruturada (adiante falaremos sobre formatos de arquivo) e assim permitir sua interpretação (pelo aplicativo original que elaborou o documento, um editor de texto por exemplo) ou outro software que possa interpretar os mesmos dados.

Os cuidados físicos e lógicos precisam ser levados em consideração numa política de preservação no universo digital, o descuido com um deles implicará na inviabilidade de acesso aos documentos eletrônicos e digitais no futuro. Se é possível manter um acervo de livros do tipo 'Obras Raras' com cuidados relativos ao armazenamento e climatização e eventualmente atividades de restauração, os acervos de documentos eletrônicos e digitais precisam de cuidados muito mais sofisticados.

### 3.0 - Estratégias para a preservação de coleções

Nos parágrafos anteriores procuramos apresentar e expor conceitos importantes relacionados ao documento eletrônico e digital e sobre sua preservação. A divisão que expusemos - problemas físicos e lógicos - é extremamente simplificadora e nos serviu apenas como ponto de partida. Agora, abordaremos considerações um pouco mais aprofundadas sobre estratégias de preservação de coleções de documentos eletrônicos e digitais.

Primeiro, é preciso esclarecer um ponto relevante. A comunidade de pesquisadores tem abordado a preservação de documentos eletrônicos e digitais a partir de duas perspectivas "*considerações quando a digitalização é usada para a conversão de documentos para o meio digital*"<sup>13</sup> e "*considerações sobre a preservação de documentos nascidos digitais*" (OGDEN, 1998, p. 213).

Neste trabalho, nos interessam as considerações sobre preservação dos documentos já digitais, oriundos de um processo de digitalização ou nascidos digitais<sup>14</sup>. Neste texto, para efeitos de simplificação, utilizaremos doravante o termo **objetos digitais**, no lugar de documentos eletrônicos digitais.

Paul Conway definiu os objetivos da preservação de objetos digitais como sendo: "*A meta fundamental da preservação digital é preservar o acesso contínuo aos dados digitais por tanto tempo quanto estes dados tenham valor*" (CONWAY, 1998, p. 67). Para preservar este acesso ao longo do tempo, é preciso levar em consideração vários elementos. Nos tópicos a seguir apresentaremos aqueles que a nosso ver são os essenciais. Considerando o formato e limites do presente texto, optamos por fazer uma introdução sobre estes elementos e indicar algumas fontes onde é possível buscar mais aprofundamento teórico.

### 3.1 – Atualização tecnológica de hardware e software

Sabemos que não teremos acesso ao conteúdo dos objetos digitais senão através de máquinas (leitoras de mídias e computadores). Precisamos também de todo o software

<sup>12</sup> *Driver's* e *dll's* são partes menores de sistemas de computação.

<sup>13</sup> Os processos de conversão de documentos analógicos, como aqueles em suporte papel, para o meio digital têm sido aplicados para resolver ou melhorar condições de acesso aos documentos, além da preservação em si. E também para preservar os originais no sentido de diminuir o desgaste pelo uso direto destes.

<sup>14</sup> Para ter acesso a considerações de diversos profissionais sobre acesso e preservação, incluindo procedimentos de digitalização para coleções de documentos ver: *Going Digital: strategies for access, preservation, and conversion of collections to a digital format*. DeWITT, Donald L. (org.). London: Haworth Press, 1998.

necessário para que um computador possa funcionar adequadamente. Assim, podemos dizer que os objetos digitais têm uma forte dependência com todo este aparato. Manter os primeiros em condições de uso para acesso futuro implica em cuidados com os segundos.

Porém, nenhum sistema composto de hardware e software durará mais que algumas décadas (já se fala em anos). A obsolescência tecnológica que temos verificado implica na falta de peças de reposição e técnicos capazes de reparar estes equipamentos (veja o caso das máquinas de escrever de algumas décadas atrás ou os primeiros computadores fabricados).

A melhor solução parece ser estar atento para este processo e não ignorá-lo, as atualizações devem ocorrer permanentemente para mitigar o processo de obsolescência tecnológica. Sobre a obsolescência de software, falaremos mais sobre isto no item *integridade dos conteúdos*.

Pode-se ter acesso a mais informações e propostas de ação no cap. 2 do livro on-line disponibilizado pela *Digital Preservation Coalition* (DPC, 2007) ou nos textos disponibilizados pelo projeto *InterPares* (<http://www.interpares.org>).

### **3.2 - Deterioração dos Suportes**

Todo material físico passa por um processo de desgaste em função do tempo, possui uma vida útil. Em condições ideais de temperatura, umidade e iluminação é possível prolongar ao máximo esta vida, no entanto, todos sucumbirão. Na prática, é muito difícil manter documentos e seus suportes físicos em condições ideais de guarda, principalmente por longos períodos. Até porque, quando estes documentos estão no início de seu ciclo de vida, ainda em uso administrativo, não maioria das vezes, não há como impor condições nem mesmo adequadas quanto mais ideais de armazenamento. Para reforçar estes problemas é preciso lembrar que os objetos digitais são infinitamente mais sensíveis que os documentos em suportes tradicionais. Sabemos da existência de documentos com milênios de idade, em argila, papiro ou pergaminho e estes documentos, apesar de seu péssimo estado de conservação, através da aplicação de técnicas adequadas, ainda podem ser lidos pelo homem. Já no caso dos objetos digitais, por outro lado, mínimas falhas em seu conteúdo podem invalidar todos os arquivos. Isto se deve às particularidades da tecnologia utilizada para criar os arquivos e agrupar os bits; grosso modo, o conjunto dos bits de um arquivo compõe uma estrutura que precisa ser respeitada integralmente.

E como danos ao suporte físico podem<sup>15</sup> danificar o conteúdo dos bits dos objetos digitais, os cuidados com o suporte físico são importantes. Como já expusermos antes, os objetos digitais não guardam uma relação conteúdo e suporte físico indissociável, e portanto, podemos migrar o conteúdo para outros suportes físicos. Mas isto precisa ser feito antes de haver danos ao suporte físico, pois após o conteúdo ter sido danificado não há mais o que possa ser feito, perdeu-se o documento.

Visto que não há nada que possa ser feito para impedir a deterioração dos suportes físicos dos objetos digitais, duas alternativas nos restam para mitigá-la: **1** – Estabelecer condições ideais de armazenamento e climatização e **2** – Estabelecer uma política de migração periódica de suportes.

---

<sup>15</sup> Atente-se para o detalhe de que não necessariamente todo o espaço de armazenamento em uma mídia qualquer é utilizado. Assim, danos como arranhões em mídias podem não afetar o conteúdo dos documentos, desde que a área afetada não contenha dados gravados. Há também artifícios tecnológicos, que dentro de certos limites, podem recuperar uma parte do conteúdo (parte das seqüências de bits) perdido dos objetos digitais.

No primeiro caso, há que se partir do levantamento dos suportes físicos utilizados, discos magnéticos, compact discs (CD's)<sup>16</sup>, DVD's, fitas magnéticas e etc. Cada uma destas categorias - e dentro delas há também variações entre diferentes fabricantes e modelos - possui suas próprias especificações que devem ser seguidas. No segundo caso, também para cada tipo de suporte, há que se determinar sua vida útil média e claro, antes do fim da mesma, é preciso providenciar a troca deste suporte. Esta atividade deve ser feita conjuntamente com as preocupações com a atualização tecnológica do hardware responsável pela reprodução destes documentos, ou seja, procurar utilizar novas mídias com tecnologia atualizada, diminuindo os problemas com a falta de manutenção em equipamentos muito antigos.

Informações técnicas e detalhes específicos devem ser buscados junto aos fabricantes dos suportes físicos dos objetos digitais. Há também algumas publicações que podem ser consultadas. Sobre fitas magnéticas e material sonoro (que inclui CD's) pode-se consultar os trabalhos da coleção CPBA publicados no Brasil em 2001 mas originais do início da década de 90 (Van BOGART, 2001) e (LAURENT, 2001). Há também um trabalho bem mais atualizado especificamente para CD's e DVD's (BYERS, 2003). Nacionalmente, pode-se consultar uma dissertação de mestrado da UNICAMP que apresenta uma metodologia para testes de confiabilidade em mídias do tipo CD's (INNARELLI, 2006). Há também o trabalho A qualidade dos suportes no armazenamento de informações (MONTE, 2004).

### 3.3 - Integridade dos conteúdos

O conteúdo dos objetos digitais, sejam eles do gênero textual, sonoro, imagético ou qualquer outro, serão sempre gravados como seqüências de zeros e uns (bits). Após a transferência destes bits para a memória do computador (através de todo o *hardware* associado), será necessário software para interpretar e traduzir, num modo compreensível aos humanos, os conteúdos. Ocorre que os softwares também sofrem um processo de defasagem tecnológica e assim como o *hardware* estão em constante modernização. Esta "atualização" trás em seu bojo um problema: ler o conteúdo de um objeto digital muito antigo pode requerer o software antigo que foi utilizado e que pode já não estar mais disponível. Mais ainda, um determinado aplicativo, digamos um editor de texto, foi projetado para funcionar em um determinado sistema operacional, que por sua vez, foi projetado para funcionar em determinado tipo de hardware. Assim, a necessidade de uso de um software antigo requer todo um aparato de outros *softwares* e *hardwares* específicos. Esse quadro pode inviabilizar o acesso a objetos digitais muito antigos.

A solução mais evidente é estabelecer uma política de monitoramento e constante atualização dos objetos digitais garantindo que sempre possam ser lidos no futuro. Este é o processo de migração dos conteúdos de objetos digitais.

Sobre migração de software ver o capítulo 4 do livro on-line editado pelo *Digital Preservation Coalition* (DPC, 2007). Também pode-se encontrar informações no trabalho *Preservação no Universo Digital* (CONWAY, 2001).

### 3.4 - Fidedignidade de conteúdos

É claro que não basta apenas manter os objetos digitais intactos ao longo do tempo. É preciso lançar mão de estratégias para manter o acesso ao conteúdo dos mesmos. E desta forma, possibilitar a contínua leitura destes documentos, falamos anteriormente de cuidados para manter o suporte físico que mantém os dados íntegros e procedimentos de

---

<sup>16</sup> Observe-se que tanto os CD's como os DVD's possuem vários subtipos, como o CD-R ou CD-RW, por exemplo.

migração de dados que permitirão o contínuo acesso aos mesmos pelos softwares. Mesmo com todos estes cuidados e o sucesso destas estratégias, estes objetos digitais ainda poderão não ter seu conteúdo fidedigno. Há uma diferença sutil mas de grande importância entre manter a integridade funcional dos conteúdos dos objetos digitais e garantir que estes conteúdos sejam fidedignos, ou seja, representem realmente o que originalmente foi gravado nos mesmos.

Em coleções de documentos em suportes tradicionais este problema praticamente não se evidencia pois, por comparação, é fácil verificar se o conteúdo de um documento não foi alterado, basta por exemplo comparar o conteúdo de dois exemplares de um mesmo livro ou dois artigos do mesmo número de um periódico. Normalmente, eventuais alterações seriam facilmente detectáveis. No mundo digital o problema é mais delicado. Alterações em documentos digitais não podem ser facilmente detectáveis. Se não houver travas de segurança que impeçam estas alterações, a princípio não será possível verificar se houve alteração ou o que foi alterado.

Para garantir a fidedignidade de objetos digitais é necessário dispor de recursos de segurança, por *hardware* ou *software*, que impeçam alterações nos documentos. Por exemplo, documentos em um CD-R não podem fisicamente ser alterados, pelo menos não enquanto estiverem gravados neste CD-R. Parece que a única maneira de confirmar se houve alteração em um documento e o que foi alterado é a comparação deste documento com outro exemplar sabidamente fidedigno.

Para compreender melhor o conceito de fidedignidade pode-se estudar os textos produzidos pelo projeto InterPares, principalmente aqueles do projeto InterPares2 (<http://www.interpares.org>). Há também o livro *Preservation of the integrity of Electronic Records* (DURANTI, 2002), apesar de abordar os documentos de arquivo os conceitos chave aplicam-se a qualquer tipo de documento.

### 3.5 – Autenticidade do conteúdo

A característica da autenticidade de um objeto digital refere-se à comprovação de autoria daquele documento, ou seja, confirmar quem ou qual organização criou o documento. Sem mecanismos de confirmação da autoria de um objeto digital, sua credibilidade pode ser questionada também, notadamente quando se trata de um documento que comprove ações de indivíduos ou trabalhos de cunho literário, por exemplo.

É possível verificar a autenticidade de um objeto digital através de vários mecanismos, como o lay-out utilizado, tipos de fontes, vocabulário de época. E há ainda recursos de assinatura digital. Em geral, a análise de autenticidade de um documento qualquer, inclusive um objeto digital, não é simples e exige um considerável estudo e esforço intelectual.

Para se compreender melhor a extensão desta importantíssima característica de um objeto digital adequadamente preservado pode-se consultar o livro *Trusting Records* (MACNEIL, 2000), também com foco nos documentos de arquivo mas com conceitos aplicáveis a qualquer documento. Há também o artigo *Can Bits and Bytes be Authentic?* (HOFMAN, 2002). Sobre assinaturas digitais pode-se consultar o artigo *Assinaturas Digitais e a Arquivologia* (BODÊ, 2006).

### 3.6 – Formatos de Arquivo

Para cada Formato de Arquivo produzido por determinado software, existirá uma especificação técnica, na verdade, haverá também uma especificação para cada **versão** de um determinado formato, por exemplo a especificação TIFF 5.0 e a TIFF 6.0, cada uma, com seu detalhamento técnico. Dependendo do Formato de Arquivo, tal

especificação técnica pode ser extremamente diferente para cada versão de um mesmo formato.

As especificações de cada Formato de Arquivo são de caráter bastante técnico e estão no escopo de desenvolvedores de software em geral. Estas especificações técnicas explicam, detalhadamente, como as seqüências de bits no arquivo devem ser estruturadas e onde cada tipo de dado deve ser gravado. Para cada formato de arquivo haverá diferenças marcantes entre as especificações.

Um ponto crucial sobre Formatos de Arquivo e que está diretamente ligado aos problemas com sua preservação refere-se ao fato de se tratar de um formato proprietário ou não. Os formatos abertos de arquivo (aqueles em que o público tem acesso aos detalhes técnicos) são mais adequados para a preservação futura pois as possibilidades de compreender o significado de sua estrutura de bits é maior.

Existem várias propostas para tentar manter, no futuro, o acesso às informações de um documento gravado através de um determinado Formato de Arquivo, como a emulação e a migração. De qualquer forma, todas as propostas dependem do conhecimento sobre Formatos de Arquivo para que possam ser executadas com sucesso em maior ou menor grau.

Para se compreender melhor a questão dos formatos de arquivo pode-se consultar o artigo *The bits and bites of data formats* (ASCHENBRENNER, 2004) ou *Selecting file formats for long-term preservation* (BROWN, 2003).

#### 4.0 - Conclusões

Neste trabalho, procuramos inicialmente demonstrar o quanto documentos são importantes para a humanidade devido a toda a cultura que estes carregam para as gerações futuras. Em seguida, caracterizamos o que são documentos eletrônicos, digitais e eletrônicos digitais; procuramos fazer isto evidenciando as diferenças entre as características dos documentos em suportes tradicionais. Mais adiante, definimos o significado do termo preservação no universo dos objetos digitais. Para então ressaltar o quanto este conceito é amplo, ou seja, refere-se a muitas ações diferentes.

Por fim, apresentamos seis pontos chave para estratégias de preservação de acervos formados por coleções de objetos digitais: 1 – **Atualização tecnológica de hardware e software**, 2 – **Deterioração dos Suportes**, 3 – **Integridade dos conteúdos**, 4 – **Fidedignidade de conteúdos**, 5 – **Autenticidade do conteúdo** e 6 – **Formatos de Arquivo**.

Muito do que foi dito neste texto refere-se ao campo da tecnologias da informação e comunicações (TIC's). Será então que os profissionais da informação envolvidos com a preservação de seus acervos devem conhecer tanto assim sobre tecnologia ?

Defendemos que as ações tecnológicas (como a atualização de hardware) sejam executadas por especialistas desta área. No entanto, percebemos claramente que falta a estes mesmos especialistas a visão privilegiada de quem lida com documentos. E por isto tem uma consciência apurada sobre sua importância, valor e conseqüente necessidade de preservação para o futuro.

Os profissionais das áreas documentais, possuindo esta consciência com relação aos documentos, devem procurar trabalhar em conjunto com especialistas em tecnologia para levar adiante as ações necessárias. E podem levar pontos chave (como os seis introduzidos acima) para a mesa de discussões como referências importantes para o debate e assim assumindo uma postura pró-ativa diante da necessidade de providências para a preservação de coleções de documentos eletrônicos e digitais.

Esperamos que as informações aqui compiladas e apresentadas de maneira concisa possam ajudar os responsáveis pela preservação de acervos digitais a estabelecer

políticas que garantam a qualidade dos documentos e a memória presente nos mesmos ao longo do tempo para muitas outras gerações.

## Referências

- ASCHEBRENNER, Andreas. **The bits and bites of data formats: stainless design for digital endurance**. New York: **RLG Diginews**, v. 8, n. 1. Disponível em: <http://www.rlg.org/>. Acesso em: 20 fevereiro 2006.
- BODÊ, Ernesto C. Assinaturas digitais e arquivologia. **Arquivística.net**. v. 2, n. 1, 2006. Disponível em: <<http://www.arquivistica.net/ojs/viewarticle.php?id=51>>. Acesso em 10.06.07.
- BROWN, Adrian. **Selecting file format media for Long Term Preservation**. UK: The National Archives, 2003. Disponível em: <[http://www.nationalarchives.gov.uk/documents/selecting\\_file\\_formats.pdf](http://www.nationalarchives.gov.uk/documents/selecting_file_formats.pdf)>. Acesso em: 10.06.07.
- BYERS, Fred R. **Care and handling of CDs and DVDs**. Washington: Council on Library and Information Resources, 2003.
- CONWAY, P. **From analog to digital: extending the preservation tool kit**. In: DeWITT, D. L. Going digital: strategies for access, preservation, and conversion of collections to a digital format. London: Haworth Press. p. 65-79, 1998.
- CONWAY, P. **Preservação no universo digital**. 2ª ed.. Rio de Janeiro: Projeto CPBA: Arquivo Nacional, 2001.
- DOCTORS, M. (1999). **A cultura do papel**. Rio de Janeiro: Casa da Palavra.
- DIGITAL PRESERVATION COALITION - DPC. HandBook On-Line. Disponível em: <<http://www.dpconline.org>>. Acesso em: 10.06.07.
- DURANTI, L., Eastwood, Terry, MacNeil, Heather. **Preservation of the integrity of electronic records**. The Netherlands: Kluwer Academic Publishers, 2002.
- FERREIRA, A. d. (1986). **Novo dicionário da língua portuguesa**. 2ª ed. Rio de Janeiro: Nova Fronteira.
- FISCHER, S. R. **A History of writing**. London, Reino Unido: Reaktion Books, 2003.
- HEREDIA HERRERA, A. (1991). **Archivística general. teoría y práctica**. 5ª ed. Sevilla: Diputación provincial de sevilla.
- HOFMAN, Hans. **Can bits and bytes be authentic ?**: preservins the authenticity of digital objects. IFLA conference in Glasgow (revised paper), 2002.
- HUNTER, D. **Papermaking: the history and technique of an ancient craft**. New York: Dover Publications, 1978.
- IM. The digital data explosion. **Information Management Journal**. Lenexa: ARMA. v. 41, n. 3, 2007.
- KIDDER, T. **A alma da nova máquina**. São Paulo: Melhoramentos, 1981.
- LAURENT, Gilles. Guarda e manuseio de materiais de registro sonoro. **Projeto Conservação Preventiva em Bibliotecas e Arquivos**: Arquivo Nacional, 2001.
- MacNEIL, Heather. **Trusting records: legal, historical, and diplomatic perspectives**. The Netherlands: Kluwer Academic Publishers, 2000.
- MONTE, A. C. LOPES, Luis F. D. **A qualidade dos suportes no armazenamento de informações**. Florianópolis: VisualBooks, 2004.
- MUÑOZ VIÑAS, S. **Contemporary theory of conservation**. Great Britain: Elsevier, 2005.
- OGDEN, B. W. **The preservation perspective**. In: DeWITT, D. L. Going digital: strategies for access, preservation, and conversion of collections to a digital format. London: Haworth Press. p. 213-216. London: Haworth Press, 1998.
- SMIT, J., GONÇALVES, C. C. **Como organizar arquivos fotográficos: projeto como fazer**. São Paulo: AASP, 2005.

THOMAZ, K. P. **A preservação de documentos eletrônicos de caráter arquivístico: novos desafios, velhos problemas.** Belo Horizonte, 2004. Tese (Doutorado em Ciência da Informação) - Programa de Pós-graduação da Escola de Ciência da Informação da UFMG.

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.

Folha anexa:

**Título do trabalho:**

**PRESERVAÇÃO DE COLEÇÕES DE DOCUMENTOS DIGITAIS**

**Autor:**

**BODÊ, Ernesto C.  
Universidade de Brasília - Unb**