



O novo mundo dos arquivos

James Turner

Universidade de Montreal

Introdução

Para bem se entender a problemática atual dos arquivos, é preciso compreender o século XX sob o ponto de vista da extraordinária rapidez da evolução tecnológica. É suficiente lembrar que diversos atores, cada um tendo uma influência profunda sobre a sociedade humana, se instalaram no cenário tecnológico durante esse período: por exemplo, a eletricidade, o rádio, o telefone, o automóvel, o cinema, a máquina de escrever, para nomear somente alguns. A partir da Segunda Guerra Mundial, assiste-se à chegada da fotocopiadora, a eletrônica, a televisão, os satélites, e sobretudo os computadores. A partir da década de 1970, a telemática, ou seja, o computador conectado a outros computadores via linhas telefônicas, mudou profundamente as possibilidades de comunicação de documentos. Desde 1990, a Internet e a World Wide Web não cessam de nos espantar por causa do desenvolvimento quase cotidiano de novas possibilidades de interação no mundo da informação.

Depois de muitos anos, a disciplina de arquivística conheceu desenvolvimentos importantes no estabelecimento da teoria, nas técnicas de organização e nos métodos de trabalho. Consta-se, entretanto, que apesar de nossa disciplina ainda não estar estabilizada definitivamente, desde já é preciso rever seus fundamentos teóricos e estabelecer um novo paradigma para a disciplina em função das novas tecnologias da informação.

É útil observar nesse contexto que não há nada de novo. Pode-se constatar que são sempre as mudanças tecnológicas que determinam a maneira de se realizar nosso trabalho de organização da informação. O surgimento de novas e importantes

tecnologias no campo da informação, como nos casos do papel e da prensa de Gutenberg, causaram também mudanças fundamentais nos métodos de trabalho das pessoas que geravam a informação no momento desses desenvolvimentos e pelos séculos seguintes. Essas tecnologias também mudaram profundamente a sociedade em seu conjunto. Nós que vivemos sobre a Terra nesse momento somos testemunhas de desenvolvimentos que se desenrolam a uma velocidade impressionante.

Histórico recente

Durante os anos de 1960 assiste-se à implantação de computadores nos governos e corporações mais importantes. Muito caros, esses aparelhos são sensíveis à temperatura e precisam ser instalados nos locais talhados sob medida e com acesso controlado. Os computadores não são muito “inteligentes”, mas o que interessa é que podem calcular com muita rapidez. Somente hoje os computadores começam a ser capazes de tratar de atividades mais “inteligentes”.

Ao mesmo tempo, as organizações de menor tamanho buscam a máquina de escrever elétrica, que se espalha durante os anos de 1960 e 1970. Por volta do fim dos anos de 1970 assiste-se à chegada de aparelhos dedicados ao tratamento de textos. Ainda uma vez, os preços são tão elevados que somente as organizações bastante importantes têm condições de usar essas máquinas. Ao mesmo tempo, as máquinas de escrever eletrônicas chegam ao mercado, mas sua utilização não se torna muito difundida em razão da chegada quase simultânea dos microcomputadores.

O aparecimento dos microcomputadores em 1980 muda radicalmente o quadro tecnológico. O computador pessoal custa menos que um automóvel. Hoje um computador custa *muito* menos que um carro e é capaz de executar as importantes operações que os grandes computadores do tipo *mainframe* não realizavam nos anos de 1960 e 1970. O novo ambiente, que se instala rapidamente, cria um problema de escala para os aparelhos administrativos, que se vêem impossibilitados de seguir tantos desenvolvimentos.

Por exemplo, a política do NARA (National Archives and Records Administration, nos Estados Unidos) sobre os arquivos

ordinolinguês está tão mal estabelecida (aproximadamente 25 anos após o começo da informatização), que a chegada da microinformática nos obriga a interrogar sobre a pertinência dessa política. (Bergeron 1992,54)

Aliás, os exemplos de perdas de arquivos eletrônicos importantes se multiplicam: os dados do recenseamento americano de 1960, a primeira mensagem de correio eletrônico em 1964, os dados sobre as florestas do Brasil capturadas por satélite nos anos de 1970, os dados da NASA, e assim por diante. Os exemplos americanos são característicos da situação por toda parte do mundo.

A situação hoje

Atualmente a capacidade dos computadores muda de modo radical e muito velocemente, abalando assim os fundamentos teóricos do arquivismo. Nós transferimos para o ambiente informatizado as políticas desenvolvidas para os documentos sobre papel, mas a complexificação das tecnologias e a influência dessas últimas sobre nossos métodos de trabalho foram de tal ordem que essas políticas não são mais suficientes. O documento eletrônico tornou-se um conjunto de relações ou de trechos de informação, podendo residir em diferentes arquivos. (Bergeron 1992, 53). Por exemplo, o relatório anual de uma companhia pode consistir em arquivos de texto, cada um redigido por uma pessoa diferente, empregando um processador textual diferente num ambiente informático diverso. Pode-se encontrar na relação das fotos e outros gráficos criados com outros sistemas operacionais, assim como os quadros estatísticos criados com diferentes sistemas operacionais, e ainda gráficos gerados por outros sistemas, tudo reunido em um documento eletrônico colocado em página para a impressão sobre papel ainda por outro sistema operacional, e ainda com uma versão diversa para ser instalada no Web. O leitor recebe um simples documento em papel, mas o arquivista responsável pelo documento eletrônico deve pensar a organização para a armazenagem, a marcação e a preservação de todos esses arquivos, bem como a relação entre eles.

Outro problema de capacidade: não se pode mais conservar a informação apenas em formato linear. O hipertexto e as ligações hipertextuais e hipermediáticas, assim como as estruturas

relacionais das bases de dados, acrescentam uma outra dimensão e complexificam mais o problema. Por outro lado, a chegada dos arquivos multimídia torna mais complexos do que nunca os arquivos eletrônicos. (Bergeron 1992,53)

Outros fatores importantes que contribuem para as mudanças fundamentais nas teorias e nas práticas, quando se trabalha com os documentos eletrônicos, são a dependência diante da mídia e dos aparelhos, a impossibilidade de entrevistar os aparelhos, a volatilidade da informação, sua segurança e sua integridade, e a proliferação de formatos proprietários, de sistemas de exploração, de sistemas operacionais, de versões desses sistemas operacionais, bem como o preço do desenvolvimento de tudo isso.

Breve, no espaço de trinta anos, a natureza da matéria de que tratam os arquivistas terá mudado radicalmente. Deve-se já distinguir o conceito de suporte daquele de informação. Antes, como a informação estava sempre integrada ao suporte, tratavam-se os dois ao mesmo tempo e pensava-se nos dois como sendo uma coisa só: um documento. Para adaptar a expressão de Negroponte (1995), antes tratavam-se dos átomos, hoje tratam-se dos bits.

Questões atuais

Para os fins de nossa apresentação aqui, dividimos as questões em cinco categorias: os documentos e seus suportes, a interconectividade, a normalização, a conversão e a preservação.

Os documentos e seus suportes: A tendência para a numerização faz com que quase a totalidade dos arquivos seja já criada em formato informático. É claro, pode-se encontrar exceções; entretanto, essa tendência é clara. Ora, o antigo papel pode durar milhares de anos, mesmo em más condições. Pode-se maltratá-lo e mesmo assim ler facilmente o texto que está relatado sobre o papel. Ao contrário, os suportes eletrônicos são muito instáveis, mesmo nas melhores condições. A duração dos suportes eletrônicos é suficiente para muitas situações, é claro, mas pouca para a conservação a longo prazo dos arquivos. O problema é tributário do fato de que nossa tendência é adotar, para fins de gestão da informação, as tecnologias criadas para outros fins.

Para conjugar-se ao problema da longevidade dos suportes, tem-se recorrido ao *repiquage*. Periodicamente, copia-se o sinal

eletrônico sobre um suporte novo a fim de assegurar sua sobrevivência. Todavia, hoje, o desenvolvimento tecnológico está de tal forma rápido, que esta prática não é mais suficiente. Agora a mudança que precisa ser vista é a "migração", ou seja, a prática não somente de copiar um documento eletrônico antigo sobre um suporte novo, mas também de o converter a uma versão mais recente do sistema operacional empregado para concebê-lo, ou ainda em um outro sistema operacional mais normalizado e capaz de lê-lo, a fim de assegurar sua consultabilidade a longo prazo.

A interconectividade, representada atualmente pela Internet e pelo World Wide Web, acrescenta uma dimensão nova à problemática. Não somente pode-se conectar dois computadores via rede telefônica, mas encontra-se hoje redes inteiras de computadores interligadas em uma vasta super-rede em escala mundial. Existem nesse contexto do desenvolvimento da *Infovia*, numerosas vantagens para os depósitos dos arquivos: por exemplo, a visibilidade, a difusão ampla das fontes, a facilidade de consulta pelos usuários, a possibilidade do teletrabalho para os arquivistas. Mas até onde deveria ir esta presença? Dever-se-ia contentar com informações gerais num resumo das fontes, ou seria melhor colocar em linha os instrumentos de pesquisa, os planos de classificação, os calendários de conservação, e eventualmente o texto inteiro de documentos manuscritos? Dever-se-ia fornecer o acesso via as redes às nossas bases de dados, aos documentos eletrônicos, às imagens de documentos manuscritos?

Por outro lado, esse novo mundo nos apresenta um problema filosófico: sobre a *Infovia*, há uma verdadeira distinção entre arquivos numéricos e bibliotecas numéricas ("arquivos digitais" e "bibliotecas digitais")? Se todos os textos são conservados em formato eletrônico, em que a cópia original é estocada num computador para consulta através das redes, ou seja, se um documento de arquivo torna-se um fichário informático e se um livro torna-se também um fichário informático, podemos ainda distinguir as bibliotecas dos arquivos (*Preserving digital information 1996, 7*)? Esse problema demonstra a que ponto as mudanças tecnológicas são profundas.

Nesse complexo contexto, os metadados, essas camadas de dados adicionais que utilizamos para descrever e organizar os dados

contidos nos documentos eletrônicos, ganham muita importância. Há múltiplos tipos de metadados: para a apresentação do documento (por exemplo, os sinais de estilos, de caracteres itálicos), para exprimir suas relações com outros documentos (por exemplo, de linhas, de pontos), para exprimir a catalogação, a classificação, a indexação (os pontos de acesso para o tema), para gerir o fichário informático (por exemplo, as informações técnicas concernentes ao formato do fichário) etc.. Porém, é sobretudo a normalização dos metadados que é de uma importância capital nesse contexto. Se se deseja permitir o acesso a muita informação via redes, tem-se todo o interesse em normalizar práticas de descrição e de organização, senão o usuário será obrigado a aprender a linguagem de cada novo sistema com o qual deseja trabalhar.

Muitas iniciativas nesse sentido foram empreendidas, por exemplo, as Regras para a Descrição dos Documentos de Arquivos (RDDA, no Canadá), a *Encoded Archival Description* (EAD, nos Estados Unidos), a *Standart Generalized Markup Language* (SGML, norma ISO 8879), e a *Dublin Core*, a *Wrawick Framework* e seus sucessores (15 elementos de base para a comunicação de documentos em rede). Resta ainda muito trabalho a fazer, especialmente o aperfeiçoamento das normas e sua implantação universal de forma independente dos sistemas operacionais e do material informático. O que nos permite ser otimistas é que, a longo prazo, seremos os conservadores de documentos altamente estruturados e onde as informações concernentes à estrutura e à organização desses documentos "viajem" através das redes com os documentos como parte integrante de tudo isso, não importando onde estão os diversos destinatários eletrônicos pelo mundo afora. Com a sistematização das práticas, passa-se de um mundo tecnológico caótico a um mundo ordenado.

Para se chegar a um mundo no qual toda a informação está em formato eletrônico e acessível a quem possua um computador e uma ligação com as redes, é necessário considerar a conversão maciça dos fichários já existentes, senão não se poderia consultar as informações mais recentes. Como assinalava Clifford Lynch, este importante observador das atividades das grandes redes, se se confia nas informações disponíveis em linha, teremos a impressão de que a história da raça humana sobre a Terra

começou em 1970. Que fazer então com as informações acumuladas em nossos depósitos depois de séculos?

É preciso demonstrar se a conversão dos fichários existentes é desejável, se ela é necessária, se ela é possível. No momento, entretanto, há obstáculos importantes, especialmente as infra-estruturas atuais, o estado das tecnologias e os custos necessários. Tomemos por exemplo o estado dos *numériseurs* e os sistemas operacionais de reconhecimento ótico de caracteres (ROC). O alvo desses últimos é permitir a conversão de documentos impressos sobre papel e fichários tratáveis por computador e isso a custo abordável. Mas os melhores sistemas operacionais atuais não fazem prova de uma taxa de resultados além de 97% ou 98% (Linke 1997, 70). Em princípio, isso pode parecer muito elevado, mas quando se considera que aquilo se traduz por cerca de trinta erros por folha A4 datilografada em espaço duplo, compreende-se facilmente que a intervenção humana é necessária para efetuar a correção de cada página antes que a possamos considerar como consultável.

Por outro lado, seria necessário prever muitas vezes não apenas a conversão de textos em octetos, mas também uma reestruturação dos dados. Por exemplo, um fichário de informação estocado sobre fichas de cartão tomará sem dúvida a forma de uma base de dados. É preciso não somente prever os campos evidentes nas estruturas, mas também de outros para acomodar a informação analógica e aquela que pode ser acrescentada à mão sobre as fichas, senão há perda de informações.

Consideremos igualmente o caso da dimensão dos fichários de imagens de páginas, fichários onde o texto não é tratável por computadores, mas que se pode ler sobre uma tela. A uma resolução de 400 pontos por polegada (ppp), se conta em torno de 85Ko/página. Porém, quando se melhora a resolução para 600 ppp, ele nos custa em espaço de estocagem cerca de 500 Ko/página. Para atender a resolução do microfilme, seria preciso escanear a 1000 ppp. A título de exemplo desse problema à escala de um arquivo, nota-se que para contar o estado civil dos habitantes de Québec, em torno de 18 milhões de certidões, ele custará 650 Go de espaço de estocagem para registrar somente as imagens desses dados, que não estarão ainda em formato de fichários manipuláveis para

uso, sem falar nos trinta meses de trabalho para efetuar essa pesquisa (Lubkov 1997, 42).

Para disfarçar os problemas desses fichários de imagens que permitem ao usuário ver a colocação de um texto na página, mas que não lhe permite manipular os dados, desenvolvem-se atualmente linguagens de descrição de páginas. Isto acrescentou uma camada de metadados, permitindo afixar o texto com a sua colocação na página exigida, e que substituirão, pode-se esperar, esses sistemas operacionais intermitentes tal como o *Acrobat d'Adobe*, que oferece uma colocação em página que exige muita memória informática, mas que está sempre em forma de ficha não manipulável, como uma telecópia.

Os problemas associados à imagem fixa e em movimento são ainda mais importantes. A questão mais notável associada a esse gênero de documentos é a dimensão dos fichários quando esses documentos são informatizados. Para a imagem fixa, não há mais problema com as simples imagens em preto e branco, mas cada pixel que compõe a imagem tem necessidade de muito mais profundidade para exprimir as cores, e assim mais memória informática. Para uma imagem em torno de 20 cm por 25 cm, é preciso mais ou menos 1Mo de memória. Para a imagem em movimento, sem compressão, necessita-se 40 Mo/imagens. A taxa de *affichage* do filme é de 24 imagens/segundo, e do vídeo, 30 imagens/segundo. O custo em memória para estocar um filme de 90 minutos é então de 960 Mo por segundo de filme, e então de 59,6 Go por minuto e de 3,5 To/hora, ou seja, aproximadamente 5 To por 90 minutos de filme.

A título de exemplo do que estes Algarismos representam em um caso concreto, pode-se notar que o sistema "Cineon de Kodak", um dos poucos sistemas disponíveis para a numeração da imagem em movimento, necessitaria de 33 grossos cassetes para estocar este filme, ao custo de 13 mil dólares pela fita magnética somente! Além disto, o sistema necessitaria de 110 horas para converter a imagem em movimento do formato analógico ao formato numérico. Isto se traduz por mais de uma hora de tratamento por minuto de filme. Não falamos ainda de custos de tratamento. E com tudo isso, seria necessário transplantar cassetes em dez ou vinte anos para evitar a perda de todo esse trabalho!

Como os suportes numéricos não são confiáveis para a conservação a longo prazo, enaltece-se às vezes a impressão sobre papéis do código informático codificado em algarismos 1 e 0, em razão das propriedades de conservação a longo prazo do papel. Mais tarde um sistema operacional de reconhecimento ótico de caracteres lerá o código para reconstituir o fichário informático. Para a imagem em movimento, esta prática não será nada prática, pois um cálculo rápido nos dá os algarismos seguintes, baseados sobre um sistema que permite a resolução comandada de 320 milhões de pixels por imagem: a 24 imagens/segundo, serão necessários aproximadamente 8 bilhões de pixels/segundo de filme 35mm. No ritmo de 6000 bits/página (quando se datilografa com entrelinha simples, calcula-se 3000), contaremos 5600 páginas (uma pilha de aproximadamente 5m) por cada segundo de filme, e portanto 180 m³ por minuto de filme, vale dizer, 16 km de espaço para estocar nosso filme de 90 minutos! Decididamente, não se poderá considerar a numeração das coleções de imagem em movimento antes de se encontrar maneiras mais econômicas de estocar os fichários assim criados.

Considerando a preservação e a conservação dos arquivos eletrônicos, podemos nos voltar um pouco para as conclusões do grupo de trabalho sobre a preservação dos arquivos numéricos (Preserving digital information 1996, 37). Este grupo de trabalho conclui que a responsabilidade primeira para a informação numérica permanece com os criadores, os fornecedores e, eventualmente, os proprietários. Além disso, o grupo enaltece a criação de uma infra-estrutura muito profunda (*deep infrastructure*) capaz de suportar um sistema distribuído de dados. Na disposição de uma tal estrutura, criar-se-á um processo de certificação de organizações capazes de estocar, de migrar e abastecer o acesso às coleções numéricas. Estas organizações certificadas teriam o direito legal de intervir pela salvaguarda de documentos depositados alhures, em caso de perigo de destruição, seja por uma ameaça física à integridade dos documentos, seja por uma mudança de políticas de conservação em outro lugar, devido à privatização de um arquivo, por exemplo.

Obstáculos à automação

Nesta parte, resumem-se brevemente alguns obstáculos atuais

à automação dos arquivos. Em nível das infra-estruturas, a banda freqüentada terá necessidade de ser acrescida consideravelmente antes que se possa responder convenientemente às necessidades dos usuários, cujo número não cessa de crescer. Devemos prever eventualmente o acesso universal à Internet e seus sucessores, como é o caso do correio à escala internacional, ou ainda do telefone. Lembremos que no momento somente uma ínfima parte da população global está em linha, e que mesmo nos países industrializados falamos apenas de dez ou quinze por cento da população.

A questão da “prisão ASCIL”, expressão de Mitchell Kapor para designar o problema das línguas não inglesas que lutam para ostentar suas marcas diacríticas no meio informático, é extremamente importante no contexto das redes. A “consortium Unicode” trabalha há vários anos para desenvolver um código informático que dê conta de todas as línguas escritas, mas esse código toma 16 bits de memória para cada caracter comparado a 7 ou 8 para os dados codificados em ASCIL, e os produtores de sistemas operacionais não os adotam muito rapidamente. Todavia, com o desenvolvimento das soluções a baixo custo dos problemas de estocagem e de tratamento, este problema importante vai, sem dúvida, ser solucionado num futuro não muito distante. Um passo importante: a “World Wide Web Consortium” vem de anunciar (julho 1997) a publicação da primeira versão de trabalho da HTML 4, que adota como jôgo de caracteres a “Unicode”.

Em termos de suportes físicos, o obstáculo principal é sua instabilidade. É necessário encontrar soluções neste nível para evitar que estejamos eternamente condenados a substituir a intervalos relativamente curtos a totalidade de arquivos que possuímos. No momento, não há nada além de tecnologias experimentais, mas é preciso crer que o problema será resolvido eventualmente. Passam sob silêncio os problemas de deterioração química e biológica.

Em termos de estocagem, assinalemos os problemas de integridade e autenticidade dos dados. Os arquivistas precisam ter confiança de que os documentos informáticos dos quais eles têm a guarda não podem ser alterados, e que o documento que eles oferecem aos usuários por consulta é o mesmo que eles

receberam por arquivo.

Em termos de sistemas operacionais, tem-se a necessidade de desenvolver os sistemas melhor integrados às necessidades dos arquivistas e dos usuários, tanto ao nível das linhas diretas entre os sistemas de gestão e documentos quanto da interação pessoa-máquina. Visto desta última perspectiva, os sistemas operacionais disponíveis atualmente são bastante penosos, não recorrem senão de maneira muito primitiva aos aparelhos cognitivos dos usuários.

Como vimos, a conversão dos fichários permanece um problema importante por várias razões. Podemos assinalar particularmente as dificuldades de conversão de fichários de ordem técnica e aqueles de ordem econômica. Além disso, teríamos vantagem em considerar como inaceitável a prática atual de versar os dados nos sistemas de informação sem controle de qualidade, ao dizer que se fará correções mais tarde. Muito frequentemente vimos que as condições econômicas não permitem essas correções. As pessoas que administram os orçamentos têm a impressão de que os trabalhos estão completos, e são os usuários que sofrem a utilização desses dados não verificados e não corrigidos. Como é o caso em qualquer outro lugar, o controle de qualidade é importante no arquivamento de dados eletrônicos.

Terminando, analisaremos o problema considerável da pilha de fichários necessária para a estocagem de imagens em movimento quando estas últimas são numerosas, os problemas arquivísticos associados à compreensão de imagens para melhor estocá-las, o trabalho considerável requerido para efetuar os trabalhos de conversão, os custos implicados nesse processo, e o problema em termos de infra-estrutura incapaz de tratar convenientemente esses enormes fichários.

Soluções a longo prazo

Apesar dos numerosos e importantes problemas associados atualmente aos arquivos automatizados, podemos ainda assim esperar ver melhoras consideráveis a curto, médio e longo prazo. A importância dos trabalhos em curso nos deixa crer que se verá o controle dos dados desde sua criação até sua disposição eventual, seja por eliminação, seja por sua instituição como arquivos permanentes. Nossos métodos, nossos processos, nossas práticas, nossas normas serão estabilizadas eventualmente. O turbilhão

tecnológico no qual nos encontramos hoje dará lugar aos métodos normalizados, sobre os quais trabalhamos atualmente.

No que concerne aos computadores, esses instrumentos de trabalho tão importantes à nossa vida, veremos bem eventualmente a chegada de computadores melhor “educados” para responder às nossas necessidades. Eles serão capazes de detectar um problema de funcionamento que experimentamos, por exemplo, e intervir de maneira interativa para nos apontar as soluções possíveis. Veremos disponíveis em linha de demonstrações, vídeo para nos mostrar como executar tal função, como executar tal tarefa, efetuar tal manobra informática. Além disso, o desenvolvimento de tipos de memória viva e morta que não se apagam automaticamente ou que não se corrompem em função de uma falha de eletricidade nos permite diminuir nossos temores psicológicos face a nossas relações com esses instrumentos que têm uma importância tão grande em nossas vidas. Veremos eventualmente a automatização de procedimentos de salvaguarda, de formação de usuários, de migração de dados e de outras funções arquivísticas. Finalmente, com o tempo assistiremos sem dúvida ao desenvolvimento de suportes informáticos tão inabaláveis quanto o velho papel.

Terminando, será bom lembrar que nós nos encontramos atualmente no meio desse turbilhão tecnológico, que o papel que representamos neste momento é de uma grande importância histórica, pois é a presente geração de arquivistas que assegura a transição entre dois mundos tecnológicos fundamentalmente diferentes um do outro. É na gestão dessa transição que nós podemos tirar vantagem de nossas atividades para os próximos anos.

*Tradução de Andréa Araújo do Vale, Carla da Silva
Miguelote e Rejane Moreira.*