

Howard Besser

Doutor em Biblioteconomia e Ciência da Informação.
Diretor da Escola de Cinema da Universidade de Nova Iorque.

Longevidade Digital

Este artigo descreve o problema do desaparecimento da informação digital e analisa os cinco fatores causadores de problemas na longevidade digital: o da visualização, o do embaralhamento, o da inter-relação, o da custódia e o da tradução. Propõe ações possíveis para reduzir o problema de preservar os documentos digitais para o futuro e alerta sobre a importância da cooperação internacional para a preservação da informação digital.

Palavras-chave: preservação digital; longevidade digital; arquivos; cooperação internacional.

This article describes the problem of the disappearance of digital information and analyzes the five factors causing the problems in the longevity digital: the viewing, the scrambling, the interrelation, the custodial, and the translation problem. It proposes a series of actions to reduce the immense problem to preserve the digital works into the future and highlights on the importance of international cooperation for the preservation of digital information.

Keywords: digital preservation; digital longevity; archives; international cooperation.

Por duas décadas temos reformado mais e mais trabalhos para o formato digital. E durante a década passada o número de obras “nascidas digitais” (trabalhos originalmente criados em formato digital) superou o número das obras criadas em formato analógico. Como o trabalho digital em breve se tor-

nará a maioria dos trabalhos em nossas instituições culturais, precisamos começar a pensar em um modo de assegurar que a informação digital continue acessível durante longo período.

Neste artigo, primeiro descreverei o problema geral do desaparecimento da

informação em formato digital. A partir daí, analisarei de forma mais detalhada os cinco fatores principais que causam problemas a sua longevidade digital. Há mais de doze anos expus esses fatores pela primeira vez e nos anos seguintes muitas pessoas os acharam úteis na tentativa de entender a complexidade da preservação digital. Finalmente, analisarei uma série de sugestões adequadas para o aperfeiçoamento da longevidade da informação digital, com enfoque principal nos metadados. Uma versão mais recente deste artigo foi publicada em 2000,¹ especificamente para a comunidade de imagem digital, no entanto essas observações podem ser úteis para todas as comunidades que desejem garantir a longevidade de qualquer tipo de informação digital.

À CURTA VIDA DA INFORMAÇÃO DIGITAL

Embora o armazenamento eletrônico seja relativamente novo, uma quantidade substancial de informações armazenadas em formato eletrônico já se deteriorou e desapareceu. Arquivos de fita de vídeo e de áudio, tais como entrevistas relativamente recentes realizadas para registrar os últimos vestígios culturais dos anciãos da tribo de índios Navajo, podem não ser resgatáveis.² Ainda que a maioria das pessoas tenda a pensar que (ao contrário da informação analógica) a informação digital durará para sempre, não conseguimos perceber a fragilidade dos objetos digitais. Grandes

porções de informação digital (tais como partes importantes da missão espacial *Viking Mars*) foram perdidas devido à deterioração das fitas magnéticas em que se encontravam. Mas o problema de deterioração do meio de armazenamento é irrelevante se comparado aos problemas causados pela rápida substituição dos dispositivos de armazenamento e pelas modificações nos formatos dos arquivos. Hoje, é quase impossível ler os arquivos contidos nos disquetes flexíveis de oito polegadas, populares há apenas trinta anos, assim como tentar decodificar os arquivos de *Wordstar* de vinte anos atrás pode se tornar um pesadelo. Grandes quantidades de informação digital de apenas trinta anos atrás está, para todos os efeitos, perdida. Para prevenir novas perdas, precisamos entender os problemas de longevidade do mundo digital. Precisamos pensar que a preservação no mundo digital difere daquela a que estamos acostumados no mundo analógico. No mundo analógico, todos os esforços de preservação enfocam a obra, o objeto como artefato. Quando começamos a nos engajar na preservação da informação em formato digital, foi preciso dar um salto conceitual, mudando o enfoque na preservação física do objeto para a preservação do conteúdo informativo, que pode estar completamente dissociado de qualquer artefato físico.

Nos itens seguintes apresentarei os cinco principais fatores causadores dos problemas na longevidade digital: o problema da visualização, o do embaralhamento,

o da inter-relação, o da custódia e o da tradução.

O PROBLEMA DA VISUALIZAÇÃO

A informação digital criada no passado requer a manutenção de uma infraestrutura e de uma base de conhecimento para que possa ser visualizada. Por exemplo, para visualizar um arquivo de um processador de texto mais antigo, necessitamos de um *software* que compreenda os esquemas de codificação do *software* original e possa exibi-los na tela adequadamente. Sem isso, a informação será mostrada de forma desordenada. Mas, para manter esses arquivos vivos ao longo do tempo, precisaremos, ainda, manter o *software* capaz de executá-los ou o conhecimento dos esquemas de codificação e sermos capazes de produzir os *softwares* que utilizem esses esquemas de codificação para que nossos arquivos digitais possam ser exibidos na tela de maneira adequada.

No mundo analógico, os formatos mais antigos perduram no tempo. A escrita cuneiforme, os papiros e os livros irão existir até que alguma coisa ou alguém (incêndio, terremotos etc.) de alguma maneira os destruam. Porém, a regra para a informação digital é a de não sobreviver a menos que alguém conscientemente aja de modo a fazê-la perdurar. No passado, diversas vezes antigos manuscritos ou livros foram encontrados escondidos em porões ou sótãos. Mas, arquivos de processadores de texto encontrados em sótãos e

porões no futuro não poderão ser lidos, a menos que seus autores tomem medidas efetivas para fazê-los durar. Mesmo que os disquetes flexíveis que encontrarmos possam ser lidos, de forma a descobriremos que existem arquivos neles, nós não conseguiremos decifrar esses arquivos, nem visualizá-los corretamente.

Quando trabalhos analógicos antigos são descobertos, ao menos eles, assim como sua estrutura, podem ser visualizados, ainda que a habilidade de decodificação do idioma tenha se perdido. Assim, descobertas posteriores, como a Pedra de Roseta, nos permitem decodificar sua estrutura e seu significado. Quando descobrimos filmes antigos (tanto de imagens fixas quanto em movimento), ainda que não tenhamos o projetor adequado para aquele formato, ainda somos capazes de visualizar seu conteúdo, colocando-os contra a luz.

Porém, a informação digital requer um conjunto elaborado de conhecimento e/ou ambiente informático para que possa ser decifrada. Normalmente, a informação digital é codificada e para que seja visualizada requer um *software* aplicativo que roda em determinado sistema operacional, que, por sua vez, necessita de uma plataforma de *hardware* específica. Normalmente, é armazenada em dispositivos físicos (como um *drive* de disco rígido, disco flexível ou CD-ROM), que requer um tipo específico de *driver* conectado a um tipo específico de computador.

Cada peça dessa infraestrutura se modifica em incrível velocidade – de um modo que

possibilita à indústria informática vender, repetidamente, o mesmo tipo de produto para a mesma pessoa (pois o usuário “necessita” de uma versão mais nova e mais rápida). As rápidas mudanças de versões de *hardware* e *software* criam uma enorme dor de cabeça para aqueles que se preocupam com a longevidade digital. Isso inclui problemas com os formatos dos arquivos, dispositivos de armazenamento, sistemas operacionais e *hardware*.

A maioria dos processadores de texto de hoje não conseguem ler arquivos criados em processadores de texto mais antigos. Muitas instituições têm problemas até mesmo para abrir os arquivos criados com o processador de texto mais popular há apenas vinte anos (*Wordstar*). De fato, os processadores de texto mais populares hoje (como o *Microsoft Word*) não conseguem sequer ler arquivos criados em versões mais antigas do mesmo processador de texto (e, geralmente, somente conseguem ler arquivos criados nas últimas duas versões mais recentes). Como podemos esperar que os arquivos que criamos hoje sejam passíveis de leitura no ambiente de informação de daqui a cem anos?

Quando os processadores de texto de hoje são capazes de abrir os arquivos a partir de versões mais recentes, geralmente esses arquivos perdem formatação e os negritos, os sublinhados, a centralização e as margens se alteram ou desaparecem. Ao menos, a maioria dos nossos arquivos de processadores de texto mais antigos são, em sua grande maioria, textos, no

padrão ASCII (Código Padrão Americano para Troca de Informações), entremeados com comandos de formatação. Assim, nas tentativas de recuperação desses arquivos há, pelo menos, alguma esperança de que palavras e frases contidas nesses arquivos possam ser encontradas. Mas, para formatos de arquivo não baseados no padrão ASCII (tais como os formatos de arquivo multimídia), há pouca esperança de que os arqueólogos, daqui a um século, sejam capazes de decifrar qualquer coisa contida nesses arquivos. Formatos como TIFF, AVI, as várias versões de MPEG etc. trarão ainda mais problemas de longevidade que os arquivos de processadores de texto.

A mudança dos dispositivos de armazenamento também trará problemas no futuro. Em menos de trinta anos atravessamos a fase dos dispositivos de armazenamento removíveis, incluindo: disquetes flexíveis de oito polegadas, disquetes flexíveis de 5.25 polegadas, CD-ROMs, DVDs e *Blu-rays* (e com o aumento da densidade de armazenamento, há pouca esperança de que a busca por novos dispositivos de armazenamento cesse em breve). Hoje, quando encontramos um disquete flexível de oito polegadas, temos que primeiro encontrar um *drive* de leitura adequado, acoplá-lo em um computador e a um sistema operacional que possua um *driver* apropriado para lê-lo. Depois de tudo isso feito, ainda teremos os problemas antes elencados para decifrar o formato do arquivo. Com as modificações dos sistemas operacio-

nais (CP/M, MS DOS, *Windows*, *Windows 95*, *Windows NT*, *Windows 2000*, *Windows XP*, *Visat* etc.) e plataformas de *hardware* (8088, 8086, 286, 386, 486, *Pentium*, *Pentium II*, *Pentium III* etc.), estamos criando uma verdadeira torre de babel na proliferação de combinações necessárias para a visualização de um arquivo.

Embora a longevidade digital pareça necessitar disso, como podemos sequer esperar que sejamos capazes de lidar com todas essas mudanças e combinações? Pense em todos os formatos que teríamos que preservar, ou todas as emulações de que precisaríamos para que pudéssemos decifrar apenas a quantidade de arquivos existentes hoje.

O PROBLEMA DO EMBARALHAMENTO

A fim de solucionar problemas a curto prazo, resultantes do uso da tecnologia digital, nos engajamos em práticas que possam resultar em riscos a longo prazo. Dois exemplos dignos de nota são a forma que temos lidado com as limitações de armazenamento e com o comércio digital.

No passado, tendo em vista que o armazenamento em larga escala era caro e a largura de banda era consideravelmente estreita, muitos repositórios respondiam a essas limitações comprimindo as imagens mestras ou de multimídia. De acordo com o raciocínio que predominou até pouco tempo, arquivos mestres comprimidos utilizam menor espaço de armazenamento, são mais fáceis de entregar aos usuários

com conexões de rede lentas e são mais práticos de manusear internamente. Nos últimos anos um número de instituições tem questionado essa ideia, já que os custos de armazenamento despencaram e as velocidades de conexão aumentaram consideravelmente. Porém, a noção de que mesmo os arquivos mestres devem ser comprimidos ainda persiste em muitas instituições.

Há inúmeros problemas criados pela compressão. Em primeiro lugar, ainda não sabemos verdadeiramente quais os efeitos a longo prazo causados pela compressão. Formatos comuns de compressão com perda de dados,³ tais como o JPEG, basicamente tentam jogar fora qualquer informação que não seja perceptível ao olho humano (cores que são próximas uma das outras são combinadas e faixas espectrais além da capacidade de percepção do olho humano são eliminadas). Porém, ainda não sabemos se qualquer desses dados eliminados será útil para os futuros aplicativos que utilizarão visão artificial (ao invés da humana) – para realizar funções tais como análise de cor, comparando e sobrepondo imagens etc. O uso de compressão com perda de dados hoje pode obstar certos usos dessas imagens no futuro.

Outro importante fator é que tanto a compressão com perda de dados, quanto à compressão sem perda de dados, aumentam o grau de complexidade na codificação de um arquivo, tornando ainda mais difícil para os arqueólogos do futuro decifrar seu conteúdo.

De modo semelhante, os inúmeros esforços para intensificar o comércio digital podem trazer ameaças à longevidade. Esquemas de criptografia para impedir o uso desautorizado aumentam o grau de complexidade da codificação dos arquivos, novamente aumentando o problema para os arqueólogos do futuro que tentarem decifrar o conteúdo do arquivo. E é difícil acreditar que todas essas peças, dos complexos esquemas de comércio digital como a arquitetura de estrutura de dados (que se baseia tanto na criptografia quanto na existência contínua de um administrador que aprove o pagamento de uma transação e libere a chave apropriada para decifrar o arquivo), sobreviverão tempo suficiente para assegurar o acesso aos arquivos digitais por mais do que uma década.

Muitos desses esquemas de embaralhamento são patenteados, e muitos não aderem aos padrões largamente aceitos. O nível de complexidade que o embaralhamento proporciona torna difícil acreditar que qualquer pessoa seja capaz de decodificar os arquivos embaralhados de hoje, mesmo daqui a meros cinquenta anos.

O PROBLEMA DA INTER-RELAÇÃO

No mundo digital, a informação é cada vez mais inter-relacionada à outra informação. A Internet é o principal exemplo de como qualquer trabalho pode incorporar ou direcionar para inúmeros outros trabalhos. E, geralmente, um trabalho pode na verdade consistir de mais do que um arquivo distinto, que pode

ou não ser visualizado como se fosse um único arquivo (tal como quando um usuário visualiza o que parece ser uma única página, que na verdade é composta de um arquivo de texto HTML e arquivos distintos para cada imagem digital da página).

Hoje os *web designers* são estimulados a empregar a “boa prática”, utilizando-se dos aspectos do hipertexto da Internet para quebrar documentos em pequenos pedaços, cada um armazenado em um arquivo distinto. Esses pedaços podem ser, então, reagrupados no momento da visualização para que se pareçam com o documento original completo, ou os vários pedaços podem ser recontextualizados em diferentes formas para diferentes finalidades. Isso significa que até mesmo os trabalhos “simples” podem conter muitos arquivos, e que qualquer arquivo pode ser parte de mais de um trabalho.

Na Internet de hoje é difícil conseguir com que nossos próprios trabalhos persistam quando direcionam para e/ou integram outros trabalhos de terceiros. Tendo em vista que nossos atuais esquemas para referenciamento de arquivos de Internet (o URL) são baseados na locação de arquivos e toda vez que a locação de um arquivo referenciado muda, nossos *links* se quebram e os usuários enfrentam a mensagem de erro mais comum na Internet: “404 Not found” (arquivo não encontrado). Geralmente, esse problema é causado por simples reorganizações no site objeto do direcionamento (a renomeação de um arquivo, de uma pasta/diretório em um local

de armazenamento hierarquicamente superior ou a renomeação do servidor etc.). Mas, esse simples ato de gerenciamento de arquivo/site instaura o caos em qualquer trabalho que direcione para ou incorpore arquivos daquele site.

Outro subconjunto crítico de problemas de inter-relação é a questão da determinação dos limites de um conjunto de informação (ou mesmo de um objeto digital). Hoje os limites de um trabalho digital não são mais confinados em um único arquivo. Frequentemente, uma página da *web* incorpora imagens, gráficos e botões que são armazenados em arquivos distintos (algumas vezes até em servidores separados gerenciados por diferentes organizações). Mesmo os trabalhos mais tradicionais como artigos de jornais, notícias ou ensaios são, geralmente, quebrados em vários arquivos distintos, que ou são agrupados juntos no momento da visualização pelo navegador do usuário, ou (para fins de estilo, para não serem visualizados pelo usuário em tamanhos que excedam duas telas cheias de extensão) permanecem arquivos vinculados distintos, em que o usuário terá que clicar, separadamente.

Se quisermos agir para preservar um desses complexos trabalhos, precisaremos desenvolver diretrizes sobre onde se situam os limites desse trabalho. Se um trabalho incorpora pedaços pertencentes ou gerenciados por outra instituição (ícones, logos, imagens, textos etc.), realizar cópias desses itens trará problemas relativos à propriedade intelectual? Se quisermos ser

capazes de mostrar aos futuros pesquisadores que tipo de informação era organizada e distribuída por uma instituição de hoje, devemos tentar preservar as *home-pages* das instituições e todas as páginas a que essa *homepage* se redirecione? E quanto às páginas vinculadas àquela por outras páginas? Onde estão os limites? Isso não é diferente do problema enfrentado hoje por conferencistas que queiram mostrar seus *websites* em uma sala de conferências onde não haja conexão de internet. Esses conferencistas terão de decidir quantas camadas de arquivos serão inter-relacionados, baixando-os para um equipamento de demonstração.

Definições dos prazos de longevidade digital

A abordagem técnica mais importante para a manutenção da informação digital viva ao longo do tempo foi primeiramente descrita em um relatório de 1996.⁴

- *Atualização* envolve transferência periódica de arquivos de um meio de armazenamento físico para outro, para evitar deterioração física ou obsolescência daquele meio. Tendo em vista que dispositivos de armazenamento físico (mesmo CD-ROMs) sofrem deterioração e que as mudanças tecnológicas tornam os dispositivos de armazenamento (tais como os *drives* de disquete flexível de oito polegadas) inacessíveis para computadores novos, formas contínuas de atualização serão necessárias durante muitos anos no futuro.

- *Migração* é uma abordagem que envolve a transformação periódica de arquivos de

um formato de codificação de arquivos em outro, que seja executável em computadores mais modernos. (Um exemplo seria transformar um arquivo de *WordStar* em um arquivo de *WordPerfect*, daí para *Word 3.0*, então para *Word 5.0*, e depois para *Word 97*.) A migração busca diminuir o problema dos arquivos codificados em uma ampla variedade de formatos que tenham existido ao longo do tempo, por meio de uma gradativa transformação dos formatos antigos em um número mais restrito de formatos contemporâneos.

- *Emulação* busca resolver um problema semelhante ao enfocado pela migração, mas sua abordagem enfoca os *softwares* de aplicativos, no lugar dos arquivos que contêm a informação. Os partidários da emulação querem desenvolver um *software* que imite cada tipo de aplicativo que tenha sido construído para cada tipo de formato de arquivo e fazer com que eles os executem em qualquer ambiente informático atual que se encontre. Assim, com os emuladores adequados, aplicativos como o *WordStar* e *Word 3.0* poderiam efetivamente ser executados nas máquinas atuais.

Ambas as abordagens, migração e emulação, requerem atualização.

O PROBLEMA DA CUSTÓDIA

Embora várias práticas tenham se estabelecido com relação a que instituições deveriam assumir a responsabilidade pela preservação e manutenção dos vários tipos de material

analógico (correspondência, manuscritos, material impresso), ainda não existe nenhuma prática para o material digital. Como resultado, muitos materiais atuais, criados originalmente em formato digital, se degradam completamente, sendo improvável que possam ser acessados pelas gerações futuras.

Por exemplo, bibliotecários de coleções especiais que ativamente buscam o desenvolvimento de uma coleção impressa em suas áreas específicas de especialização alegam que seria de responsabilidade das equipes de computação das instituições buscarem o desenvolvimento de uma coleção de materiais criados originalmente em formato digital.⁵ Contudo, essa equipe de computação alega que a responsabilidade em buscar o desenvolvimento da coleção de materiais digitais deva ser do especialista no assunto-alvo. Enquanto isso, muito desse frágil material simplesmente não é colecionado.

Outro exemplo é a correspondência, que no mundo analógico deixa um rastro de papel. Muitas instituições seguem diretrizes para preservar quantidades significativas de correspondência em papel. Porém, poucas organizações desenvolveram diretrizes semelhantes para preservação de correspondência eletrônica, e poucos indivíduos têm a mais remota ideia de como devem agir para preservar suas próprias correspondências pessoais, ainda que queiram. Esse problema está se tornando tanto mais sério quanto mais e mais importantes são as correspondências criadas na forma digital.

Um exemplo final é o do domínio da criação literária. No mundo analógico, autores costumam deixar importantes rastros de seu processo criativo na forma de inúmeros esboços, manuscritos com anotações de revisão e correspondência. Hoje, eles usam processadores de texto e email para esboços e correspondência. E, geralmente, salvam poucos de seus esboços e nenhuma de suas correspondências.

A questão crucial a ser enfrentada nos anos vindouros é: Quem será responsável pela preservação do material em formato eletrônico? Os indivíduos devem se responsabilizar? Ou as entidades sociais (tais como empresas, bibliotecas, arquivos e associações profissionais) devem intervir drasticamente na preservação do material? E como será a decisão sobre o que deve ser preservado?

Outra questão crítica é: Como eles devem agir para preservá-los? Nosso campo ainda necessita desenvolver diretrizes e “boas práticas” para que as instituições e indivíduos que queiram se engajar na tentativa de fazer com que a informação digital perdure, saibam como fazê-lo.

Uma das principais funções dos arquivos é a de assegurar a autenticidade de um trabalho. Eles fazem isso através do acúmulo de “provas” e da manutenção da “cadeia de custódia”. Mas quando os trabalhos são submetidos a reiteradas ações de atualização, tais como a maioria das abordagens para fins de longevidade digital (ver Definições dos prazos de longevidade digital), esses meios tradicionais de garantia da

autenticidade se rompem. Arquivos reiteradamente copiados para novos estratos enfrentam a probabilidade de que essas mudanças sejam incorporadas a esses arquivos, e nós conhecemos pouco sobre como controlar essa mutabilidade causada por repetidas atualizações.

O PROBLEMA DA TRADUÇÃO

Quando o conteúdo é traduzido para novos dispositivos de entrega (tal como os meios digitais), a mudança de forma geralmente altera parte do seu significado. Conversões de analógico para analógico enfrentam esse problema, assim como as conversões de analógico para digital (uma fotografia de uma pintura não é a mesma coisa que a pintura, e a representação digital de um objeto não é a mesma coisa que o objeto).⁶

O fato de podermos fazer cópias idênticas de arquivos digitais tem levado algumas pessoas a acreditar, de forma equivocada, que a conversão de digital para digital não sofre os mesmos problemas de tradução que a conversão analógica para digital sofre. Isso não é verdade, pois embora os *bits* nos conteúdos dos arquivos possam ser idênticos, o ambiente do aplicativo utilizado para visualizar a maioria dos arquivos certamente será diferente. De fato, a verdadeira razão para a conversão do arquivo é a impossibilidade de manter o ambiente do aplicativo ao longo do tempo.

Muitas pessoas experimentam isso quando seus processadores de texto importam “com êxito” um documento criado por

uma versão mais recente do mesmo processador de texto, e perdem a formatação (como alteração de centralização, sublinhado e de fontes) ou a pontuação (perda de apóstrofes ou aspas). Isso pode ocorrer ainda em ambientes de emulação, pois os criadores desses ambientes têm que escolher quais aspectos do ambiente eles querem simular, pois não podem emular todos os aspectos. Por exemplo, há uma década, uma emulação de um dos mais recentes jogos de computador (*Moon Dus*) foi apresentada para seu criador original, Jason Lanier, que argumentou se tratar de um jogo completamente diferente daquele que ele criara, pois o *pacing* (ritmo na transmissão de dados) era diferente.

Na preservação de um trabalho, é crucial que preservemos as partes do ambiente desse trabalho que podem não ser, em um primeiro momento, óbvias. Por exemplo, qualquer um é capaz de reconhecer que nós temos que preservar a imagem de cada página em um livro digitalizado. Mas, para que o livro possa ser utilizado, precisamos guardar os “procedimentos” importantes daquele livro, tais como os metadados e os “procedimentos” associados que permitirão que futuros usuários mudem de página, pulem do sumário para um capítulo específico ou consigam ir e voltar do corpo do texto para as citações ou notas de rodapé.⁷ Salvar apenas as imagens das páginas do livro sem esses “procedimentos” seria o mesmo que salvar um *videogame* com as interações em algum tipo de representação, mas faltando uma das mais importantes funções.

Com trabalhos iniciados em meios digitais, precisamos compreender melhor os aspectos do ambiente original desse trabalho, que são cruciais para sua visualização, e precisamos, também, descobrir meios de manter todos os importantes “procedimentos” de um trabalho para que possamos mover esse conteúdo por gerações de migrações ou emulações.⁸ É necessário, ainda, compreender como cada novo ambiente de visualização afeta a natureza do trabalho. Por exemplo, muitos diretores de cinema alegam que o filme deles é radicalmente modificado quando mostrado na tela de vídeo. Como os criadores de multimídia de hoje se sentiriam com o fato de seus trabalhos estarem sendo apresentados em ambientes futuros, onde os tubos de raio catódico não mais existirão para a visualização?

CAMINHOS PARA APERFEIÇOAMENTO DA LONGEVIDADE DIGITAL

Devido a esses terríveis problemas, como podemos esperar que consigamos assegurar a longevidade dos trabalhos, de hoje, que queremos preservar? Algumas dessas abordagens foram primeiramente delineadas em 1998,⁹ mas a informação a seguir tem sido acrescida, pelos mais recentes pensamentos e desenvolvimentos.

Abordagens amplas

Em primeiro lugar, precisamos reconhecer que existe uma grande quantidade de conhecimento sobre como realizar a preservação de *bits* através do tempo. Por meio

século, a comunidade de processamento de dados tem praticado a movimentação de grandes corpos concentrados de *bits* de um meio de armazenamento físico para outro. Nossa comunidade precisa estudar a experiência dos departamentos de processamento de dados de empresas e universidades, para aprender com as experiências deles, calcular seus custos, para então analisarmos como esses devem ser utilizados para os corpos menos altamente concentrados de informação digital que nossa comunidade possui.

Enquanto estudamos essa experiência, precisamos ainda ter em mente que a preservação de *bits* é apenas uma pequena parte do problema. Isso é minimizado pelos muitos problemas maiores enfrentados para assegurar que os formatos de arquivos sejam acessíveis e pelos problemas envolvendo instituições, políticas, funções e responsabilidades.

Nos milhares de anos desde que a Biblioteca de Alexandria foi destruída, a redundância tem sido a chave para a preservação da informação. A existência de múltiplas cópias de um trabalho geograficamente distribuídas entre inúmeros *sites* tem ajudado a preservação de trabalhos contra desastres naturais e/ou criados pelo homem (abrangendo incêndios, terremotos e a destruição acidental de um conjunto de trabalhos). Qualquer estratégia de preservação para informação digital a longo prazo deve incorporar relacionamentos de cooperação entre locais geograficamente distantes e as instituições. Precisamos

desenvolver projetos de cooperação internacional em que instituições armazenem e atualizem cópias redundantes de trabalhos que estejam na verdade sob custódia de outras instituições.

As atuais leis de propriedade intelectual proíbem que um arquivo ou biblioteca preserve informação em formato digital, especialmente porque muitas informações digitais sob sua custódia são licenciadas e não de sua propriedade. Um estudo sobre direitos autorais feito pela National Academy of Science (National Research Council 2000) recomenda firmemente que as leis de propriedade intelectual sejam alteradas de forma a permitir que essas instituições legalmente preservem informação em formato digital, e que fundos expressivos sejam alocados para a preservação digital. Precisamos continuar a monitorar as alterações nas leis de propriedade intelectual¹⁰ e pressionar para que as alterações permitam que possamos nos engajar na preservação digital sem enfrentar punições criminais.

Necessitamos de mais experiência nas duas estratégias concorrentes para preservação digital – emulação e migração (ver Definições dos prazos de longevidade digital). A abordagem da emulação é altamente experimental e precisamos analisar os dois estudos internacionais experimentais que exploraram essa área há uma década: NEDLIB, patrocinado pela Comunidade Europeia (*Networked European Deposit Library website*), e o Projeto CEDARS (*CURL Exemplars in Digital Archives*

website), patrocinado pelo Britain's Joint Information Systems Committee e pela US National Science Foundation.

O que nós podemos fazer enquanto comunidade

Enquanto ninguém resolveu, ainda, a ampla gama de problemas envolvendo longevidade digital, há inúmeras ações individuais que podemos realizar para aumentar a probabilidade de que o trabalho que procuramos preservar permaneça acessível durante longos períodos de tempo. Existe ainda uma série de ações que nossa comunidade, como um todo, deve se engajar a fim de reduzir esse imenso problema.

Nossa comunidade precisa persistir na busca por meios padronizados e inteligíveis para que um objeto digital auto-identifique seus formatos e os aplicativos necessários para sua visualização. Com um padrão para a fixação do nome do aplicativo de visualização em um local específico dentro do cabeçalho de imagem, os arqueólogos do século XXII, ao descobrirem os arquivos de hoje, ao menos conseguirão saber quais são os aplicativos necessários para visualizar o arquivo. Influenciados por isso, inúmeros problemas relacionados à longevidade das imagens digitais surgiram, como parte da *Spring 1999*, uma conferência patrocinada pela Commission on Preservation & Access, pela National Information Standards Organization (NISO) e pelo Research Libraries Group, em 2006, que resultou na Z39.87, a norma técnica de metadados NISO para imagens digitais fixas.¹¹

Nossa comunidade necessita compreender melhor como a informação se relaciona com outra.¹² Em especial, precisamos de esclarecimentos adicionais sobre o que são os "limites" dos objetos de informações. Quando tentamos preservar algo (especialmente um hipertexto ou objeto de hiperídia), precisamos saber quantos pedaços realmente necessitamos preservar. Grupos como o Internet Archive (Archive-It) e California Digital Library (Web Archiving Service) estão desenvolvendo ferramentas para arquivar partes selecionadas da *Web*, enquanto a Conferência Anual sobre Preservação de Objetos Digitais (iPres) destaca as últimas ferramentas e práticas que estão sendo desenvolvidas.

Finalmente, nossa comunidade precisa desenvolver um conjunto de diretrizes concretas, que possam ser utilizadas por pessoas e instituições que desejem fazer com que a informação perdure.

Na decisão de preservar digitalmente um grupo de trabalhos, a instituição deve, primeiro, compreender as necessidades especiais do tipo de trabalho contido naquela coleção. Isso significa compreender como a reformatação desses objetos para outros formatos pode afetar a compreensibilidade e a usabilidade desses trabalhos. Isso significa, também, entender os limites desses trabalhos e quais pedaços devem ser preservados (talvez até incluindo-se pedaços contextuais), e o que significa preservar os "procedimentos" do trabalho, e não simplesmente seu "conteúdo".

A função do metadado

A esta altura, metadados extensivos são o nosso melhor recurso para minimizar os riscos de que um objeto digital se torne inacessível. Adequadamente utilizados, os metadados podem: identificar o nome do trabalho, quem o criou, quem o reformatou e outras tantas informações descritivas; fornecer identificação única e *links* às instituições, arquivos ou bases de dados que tenham metadados descritivos mais extensivos sobre esse trabalho (isso é especialmente importante na eventual ocorrência de que o arquivo digital e seus metadados externos se separem); informar o ambiente técnico necessário para visualizar o trabalho, incluindo aplicativos e números de versões necessárias, esquemas de descompressão e outros arquivos que sejam necessários para vinculá-los etc.

Existem vários tipos de metadados que parecem insignificantes hoje, mas podem se mostrar cruciais para a visualização adequada desses arquivos no futuro. Por exemplo, informação guardada sobre o perfil de cores de um escâner será crucial para o futuro sistema de gerenciamento de cores calcular as diferenças do dispositivo de visualização e mostrar as cores

adequadamente em um dispositivo específico. Uma boa regra de manuseio é salvar qualquer metadado que seja barato/fácil de capturar ou que tenha sido indicado que pode vir a ser importante.

Nos últimos anos a comunidade bibliotecária desenvolveu o padrão PREMIS (<http://www.loc.gov/standards/premis/>) para acompanhar as alterações feitas em um trabalho por meio de atualização e migração.

INFORMAÇÃO ADICIONAL

Aqueles envolvidos no planejamento da longevidade digital deveriam ler os textos recentes que demarcaram os problemas nos nossos campos: o relatório da Commission on Preservation and Access (Task Force 1996), a Conferência Getty's Time & Bits sobre preservação digital¹³ e outros itens referidos na página de longevidade de Besser.¹⁴ Ainda devem monitorar continuamente os *websites* da NDIIPP (Library of Congress' National Digital Information Infrastructure & Preservation Program), as publicações da Commission on Preservation & Access (*Commission on Preservation & Access website*) e o trabalho do Internet Archive (*Internet Archive website*).

N

O

T

A

S

1. Versão publicada em um livro editado por Maxine Sitts intitulado *Handbook for digital projects: a management tool for preservation and access*. Andover, Massachusetts: Northeast Document Conservation Center, 2000, p. 155-166.
2. SANDERS, Terry. *Into the future: preservation of information in the electronic age*. Santa Monica: American Film Foundation (16 mm film, 60 minutes), 1997.

3. Por definição, quando a imagem comprimida “sem perda de dados” é descomprimida, ela é idêntica à imagem antes da compressão. Mas, quando uma imagem comprimida “com perda de dados” é descomprimida, ela é diferente da imagem original, pois algumas informações foram eliminadas como parte do processo de compressão.
4. COMMISSION ON PRESERVATION AND ACCESS AND RESEARCH LIBRARIES GROUP. *Preserving Digital Information*. Task Force on Archiving of Digital Information (1996). Disponível em: <http://www.rlg.org/ArchTF/tfadi.index.htm>.
5. Collecting at the Margins: Social Protest and Counterculture Materials. In: COLLECTION DEVELOPMENT LIBRARIANS OF ACADEMIC LIBRARIES DISCUSSION, 1999. Conferência da Filadélfia, American Library Association Midwinter, 30 de janeiro de 1999.
6. BESSER, Howard. The changing museum. In: CHEN, Ching-Chih (ed.) *Information: the transformation of society*. Proceedings of the 50th Annual Meeting of the American Society for Information Science. Medford, NJ: Learned Information Inc., 1987, p. 14-19.
7. Making of America II White Paper, 1998. Disponível em: <http://sunsite.berkeley.edu/moa2/>.
8. BESSER, Howard; GILLILAND-SWETLAND Anne. *Multimedia: issues in using visual material in cultural heritage organizations*. Spring 1999 class and website. Disponível em: <http://www.sims.berkeley.edu/impact/s99/>.
9. LYMAN, Peter; BESSER, Howard. Defining the problem of our vanishing memory: background, current status, models for resolution. In: MACLEAN, Margaret; DAVIS, Ben H. (eds.). *Time & bits: managing digital continuity*. Los Angeles: J. Paul Getty Trust, 1998, p. 11-20.
10. BESSER, Howard. *Copyright* (website). Disponível em: <http://besser.tsoa.nyu.edu/howard/Copyright/>.
11. BESSER, Howard. *Image Metadata: meeting summary*, 1999. Disponível em: <http://besser.tsoa.nyu.edu/howard/image-meta.html>.
12. BESSER, Howard; GILLILAND-SWETLAND Anne, op. cit.
13. MACLEAN, Margaret; DAVIS, Ben H. (eds.). *Time & bits: managing digital continuity*. Los Angeles: J. Paul Getty Trust, 1998.
14. BESSER, Howard. *Digital longevity* (website). Disponível em <http://besser.tsoa.nyu.edu/howard/longevity/>.

Recebido em 21/1/2011

Aprovado em 2/3/2011