



PRESERVAÇÃO DIGITAL DE DOCUMENTOS DE ARQUIVO À LONGO PRAZO: ESTRATÉGIAS E INICIATIVAS

Leonardo Mendes Padilha, Maurício Barcellos Almeida

Resumo. Documentos digitais estão presentes na maioria das instituições, e encerram parcela significativa do patrimônio cultural da sociedade nos últimos 30 anos. Documentos de arquivo em formato digital têm trazido novos desafios para profissionais da informação. Destacam-se estudos sobre como manter documentos por longos períodos em um ambiente efêmero como o digital. Alternativas têm sido pesquisadas sobre a denominação de “preservação digital”. O presente artigo aborda estratégias para preservação digital e descreve iniciativas internacionais para a manutenção a longo prazo de documentos de arquivo digitais. O objetivo é explorar os recursos existentes sobre preservação digital para uso em projetos nacionais, bem como fomentar o estudo e a disseminação do tema entre profissionais de informação e arquivistas.

Palavra chave: preservação digital, arquivos, metadados

Abstract. *Digital documents are pervasive in institutional environments, and retain valuable portion of the society cultural heritage produced during the last 30 years. Archival documents in a digital format have given raise to new challenges for information scientists and archivists. Studies about how to preserve documents for a long period of time in an ephemeral digital environment have been emphasized. Alternatives have been studied under the label “digital preservation”. The present paper advances strategies to digital preservation and describes international projects about maintaining archival documents in the long term. The goal is to explore resources and well-established projects with the aims of fostering the dissemination of that matter among archival professionals and information scientists.*

Key-words: *digital preservation, archives, metadata*

1. Introdução

A preservação de documentos de arquivo é questão que recebe atenção de instituições públicas e privadas em todo o mundo, em função da necessidade de manter o patrimônio cultural e a memória da sociedade. A falta de registro do contexto de produção de documentos de arquivo tem sido um problema para a preservação ao longo da história. Nos últimos 30 anos, documentos de arquivo têm sido substituídos em larga escala por documentos digitais em formatos diversos, produzidos diretamente em ambiente digital (nati-digitais) ou provenientes de digitalização. Apesar do avanço tecnológico, o problema de manter o contexto se mantém. Além disso, o ambiente digital revela outras questões peculiares, as quais dificultam o trabalho de preservação.

A dificuldade em preservar documentos digitais é consequência da volatilidade das mídias utilizadas para registro dos dados e da rápida obsolescência da tecnologia. Mídias como o papel podem durar até 500 anos, mas não se consegue precisar um período equivalente para as mídias ópticas e magnéticas. Além disso, os documentos de arquivos se tornam dependentes do formato



digital, o qual depende do *software* e do *hardware*. Assim, o acesso a documentos se torna dependente do conjunto tecnológico que os produziu, e nada garante que a mesma tecnologia estará presente no longo prazo.

Estudos sobre o impacto do formato digital na produção e manutenção à longo prazo de documentos de arquivo em formato têm revelado questões multidisciplinares. Essas questões têm sido pesquisadas sob a denominação de *preservação digital* (DURANTI, 2007). Uma alternativa amplamente utilizada para a preservação de conjuntos de dados é considerar, além da manutenção de dados, a descrição dos metadados que explicam como acessar aqueles conjuntos. Documentos digitais e metadados correspondentes são fortemente conectados: um não tem valor sem o outro.

O presente artigo se insere nesse contexto, apresentando iniciativas sobre a preservação de documentos de arquivo em formato digital. A partir daqui, objetivando simplicidade, utiliza-se o termo “documento” para se referir a “documentos de arquivo” e o termo “preservação” para se referir a “preservação à longo prazo”. A seção 2 apresenta uma visão geral sobre a preservação digital, a seção 3 apresenta as principais iniciativas de pesquisa pioneiras no assunto e a seção 4 traz considerações finais.

2. Uma visão geral sobre preservação digital

No processo de disseminação da informação escrita, a invenção da imprensa elevou significativamente a quantidade de registros escritos e promoveu uma verdadeira revolução social. O século XVIII foi a época em que os registros em papel ganharam força e independência (CLANCHY, 1979). No final do século XX, analogamente, a emergência da Internet promoveu a disseminação em larga escala de registros digitais (LANZINGER, 2001).

A maioria dos documentos digitais, entretanto, não tem recebido qualquer tipo de tratamento visando preservação. Isso ocorre mesmo considerando que parte significativa corresponde a documentos científicos e de valor histórico. Problemas diversos dificultam tal tratamento, com destaque para a crescente dependência do *hardware* e do *software* ao longo do ciclo de vida do documento (THOMAZ, 2004).

Outro fator impactante no processo de preservação é a mídia utilizada para armazenamento. A migração entre diferentes tipos de mídias altera metadados, modifica a forma de gravar os dados e pode resultar em perdas. Além disso, o tempo de duração de uma mídia é maior do que o tempo de sua obsolescência tecnológica (UNITED KINGDON, 2003).

A adoção de um conjunto de requisitos básicos visando a preservação é o primeiro passo de um longo processo, uma vez que a tecnologia é incapaz de proporcionar solução definitiva. De fato, tais requisitos são definidos no escopo de uma política de preservação digital e parte das diretrizes arquivísticas de uma instituição, contemplando: i) planejamento de estratégias; ii) definição de funções para seleção de registros; iii) definição de funções para adoção de formatos de armazenamento de dados e respectivos metadados (BULLOCK, 1999).

Um padrão de metadados define o contexto para caracterização do documento, atributo essencial



para garantir o acesso à longo prazo (IKEMATU, 2001). Tal garantia, envolve ainda atributos que informam ao *software* ou *hardware* os requisitos para operações inerentes à preservação. As técnicas para preservação digital consistem de mecanismos que, em última instância, buscam combinar a estrutura lógica do registro ao seu suporte físico. As técnicas mais conhecidas são *migração*, *emulação* e *encapsulamento* (HEDSTROM, 1997).

A *migração* consiste na estratégia de substituir a mídia de armazenamento e o *software* associado, antes que o registro se torne inacessível. Garante-se assim acesso à longo prazo ao acompanhar a evolução tecnológica. Em geral, a transferência de documentos entre mídias ocorre em períodos determinados em políticas. Migrações sucessivas sem planejamento resultam em alterações na formatação do registro, tornando-o inacessível.

A segunda estratégia, denominada *emulação*, cria um *software* (o emulador) o qual recria as condições do *software* original. A emulação dispensa a migração ao recriar o ambiente do sistema anterior, mas mantém a necessidade de migração do suporte. Requer metadados do documento para que o emulador possa ler um arquivo escrito anteriormente. Não é uma técnica de simples execução: não é possível prever os metadados necessários e nem o curso das mudanças tecnológicas (LUSENET, 2002).

A terceira estratégia, conhecida como *encapsulamento*, consiste em agrupar o conjunto de *softwares* e a descrição das condições necessárias para a reprodução do documento a ser preservado. É criado ainda um sistema administrador de que assegura atualizações *software* e *hardware*. Trata-se de uma alternativa adequada, por exemplo, para a preservação de documentos digitais complexos, com anexos ou *hyperlinks*, como ocorre na web. Identificam-se os elementos participantes da rede de conexões e encapsulam-se os recursos necessários para a visualização do conjunto.

3. Iniciativas internacionais para a preservação digital

Ao definir uma estratégia para a preservação, as instituições se beneficiam da experiência de iniciativas e de casos de sucesso. A *Norma Brasileira de Descrição Arquivística (NOBRADE)*, por exemplo, é uma importante iniciativa brasileira arquivística. Embora não trate diretamente de preservação, trabalha com metadados arquivísticos ligados a descrição. Tais metadados são fundamentais já que além de dos aspectos computacionais a que se enfatizar os arquivísticos. A NOBRADE baseia-se na *International Standard Archival Description*, um padrão internacional traduzido por *Norma Geral Internacional de Descrição Arquivística* (BRASIL, 2008).

3.1. Metadata Requirements Project

O *Metadata Requirements Project*²³ é uma iniciativa desenvolvida na *University of Pittsburgh*, que propõe um modelo de referência para intercâmbio de documentos visando a preservação digital no contexto de negócios. O projeto busca por um padrão de metadados para garantia de autenticidade de documentos em transações comerciais.

23 Disponível na Internet em: < <http://www.archimuse.com/papers/nhprc/> >. Acesso em: 22/07/2010.



O conjunto de metadados definido possibilita acesso seguro a documentos de transações comerciais, e, portanto valor de prova legal, por um longo período de tempo. Os metadados foram organizados em um modelo de referência composto por módulos funcionais – organização, sistemas, aquisição de registros, utilização e manutenção – cada qual com o seu conjunto próprio de metadados.

3.2. Dublin Core Metadata Initiative

Na tentativa de reduzir a ambiguidade de metadados criados automaticamente, a *Dublin Core Metadata Initiative*²⁴ (DMCI) propõe um conjunto mínimo de atributos para descrever metadados. O DCMI é mantido por um organismo sem fins lucrativos – a *Online Computer Library Center Research* – a qual se fundamenta no padrão *Dublin Core* (DC) para estabelecer regras de gestão de documentos.

A descrição do DMCI (2004) define termos com fins diversos: i) termos básicos de descrição (nome, identificação, identificador web, tipo de termo); ii) termos de descrição adicionais (comentário, referência, versão); e iii) termos para definição de contexto (esquemas de vocabulário, sintaxe de esquemas de classificação).

3.3. Open Archives Information System Reference

A *Open Archives Information System Reference Model*²⁵ (OAIS) é talvez a mais importante iniciativa sobre preservação digital. Trata-se de padrão ISO desenvolvido pelo *Consultative Committee for Space Data Systems* da NASA (CCSDS).

O OAIS é aplicável a qualquer tipo de documento digital e determina as relações entre o OAIS, os *produtores de informação* e os *consumidores de informação*. No âmbito do OAIS, dados em formato digital são armazenados em uma sequência codificada (*bit stream*), a qual é interpretada por esquemas de representação. O conjunto composto por dados, codificação e representação da informação constitui um *objeto digital*, o qual, *uma vez acessível a uma pessoa*, é então preservado (CCSDS, 2002).

Os dados do objeto adicionados à metadados de preservação são reunidos na uma entidade *pacote de informação*. O pacote abriga dois tipos de dados: o *informação de conteúdo*, que contem os dados do conteúdo propriamente dito, e a *informação descritiva de preservação*, que contem metadados para o acesso futuro. Ainda existe outro nível de metadados, denominado *informação descritiva*. Os dados de preservação são definidos por princípios de proveniência, inviolabilidade contexto e referência.

3.4. CURL Exemplars in Digital Archives Project

O *CURL Exemplars in Digital Archives Project (Cedars)*²⁶ é uma iniciativa de consórcio

24 Disponível na Internet em : <<http://dublincore.org/>>. Acesso em: 30/04/ 2010.

25 Disponível na Internet em : <<http://public.ccsds.org/publications/>>. Acesso em: 21/05/ 2011.

26 Disponível na Internet em : <<http://www.ukoln.ac.uk/services/elib/projects/>>. Acesso em: 1/02/2010.



universitário inglês, que resultou em protótipo de um sistema baseado no sistema OAIS. O conjunto de metadados proposto representa a memória descritiva, administrativa, técnica e legal, bem como a estrutura necessária para criar um pacote de informação para documento de arquivo digitais.

O *Cedars* possui dois desdobramentos principais: o primeiro, sobre propriedade intelectual, que lida com o direito de propriedade intelectual no contexto de trabalho de bibliotecários e arquivistas; o segundo, sobre estratégias, que consideram que a preservação digital envolve preservar o acesso, a qual, por sua vez, envolve agregar metadados (de preservação, de migração, de contexto, dentre outros).

3.5. Preservation Metadata for Digital Collections Initiative

A iniciativa conhecida como *Preservation Metadata for Digital Collections Initiative*²⁷ (PMCD) é um projeto da *National Library of Australia* que objetiva manipular, para fins de preservação, coleções de bibliotecas, arquivos de publicações digitais *on-line*, publicações digitais e coleções de áudio, coleções de textos e imagens.

O PMDC também é baseado no OAIS: metadados descrevem pacotes de informação e a história dos objetos no processo de preservação. A iniciativa distingue ainda parâmetros de metadados para documentos nati-digitais e também para aqueles convertidos via digitalização. O principal item do sistema de preservação é o histórico de preservação (NLA, 1999).

3.6. Metadata Encoding and Transmission Standard

O *Metadata Encoding and Transmission Standards*²⁸ (METS) é uma iniciativa desenvolvida pelo *Network Development and MARC Standards Office* da *U.S. Library of Congress*. A METS se fundamenta no padrão OAIS e propõe um padrão de metadados em formato *Extended Markup Language* (XML) para gerenciar coleções de objetos digitais. Consiste em um padrão de codificação descritiva, administrativa e estrutural de metadados, criado para bibliotecas digitais.

Os componentes METs para caracterizar metadados são o cabeçalho METS, os metadados descritivos, os metadados administrativos, os arquivos (*files*), o mapa estrutural, a ligação estrutural e o comportamento.

3.6. Projeto Interpares

O projeto *Interpares* é um projeto internacional organizado em três fases. Tem como objetivo desenvolver conhecimento teórico e metodológico para a preservação permanente de arquivos gerados eletronicamente (DURANTI, 2007). Com base nesse conhecimento, formularam-se modelos de estratégias, políticas e padrões capazes de assegurar a preservação. O projeto está em sua terceira fase, e já se desenvolveram modelos para elaboração de políticas para preservação digital que abrangem a

27 Disponível na Internet em : <<http://www.nla.gov.au/preserve/pmeta.html>>. Acesso em: 22/07/2010.

28 Disponível na Internet em : <<http://www.loc.gov/standards/mets/>>. Acesso em: 22 /10/2010.



captura do objeto digital, sua organização, trâmite, preservação e garantia de autenticidade.

4. Considerações finais

O presente trabalho apresentou requisitos básicos necessários para uma política de preservação de documentos de arquivo em formato digital, bem como iniciativas pioneiras nesse sentido. Espera-se contribuir explorando um assunto ainda pouco difundido entre profissionais da informação e arquivistas.

A importância de buscar soluções para a preservação digital está associada a manutenção de documentos de arquivo de caráter permanente, fundamentais para compreensão das instituições pelas gerações futuras, e mesmo para questões legais. O problema torna-se mais complexo quando o período de preservação exigido é o longo prazo. Muito embora o caráter interdisciplinar desse estudo a contribuição dos arquivistas, bibliotecários e demais profissionais da informação é fundamental na medida em que a omissão desses profissionais pode levar a confusão na construção de sistemas de gerenciamento que não observem requisitos necessários a contextualização dos documentos digitais.

Esse trabalho é uma contribuição da Ciência da Informação em assunto que, muitas vezes, tem sido tratado como computação, mesmo que o domínio teórico seja próprio da arquivística (preservação). De fato, metadados são fundamentais para um processo eficiente de preservação digital. São parte intrínseca do processo e representam a única forma de contextualizar os documentos de arquivo, garantindo princípios fundamentais como a proveniência. Esses princípios se mantêm fundamentais no ambiente digital. A falta de metadados adequados inviabiliza a preservação digital, fato que explica a grande ênfase ao assunto.

5. Referências

BRASIL. (2008). *CONARQ: Norma Brasileira de Descrição Arquivística*. Disponível em: <<http://www.conarq.arquivonacional.gov.br/media/publicacoes/nobrade.pdf>>. Acesso em 21/07/2010.

CLANCHY, M. T. *From memory to written record. England, 1066 – 1307*. Cambridge: Harvard University Press, 1979.

CONSULTATIVE COMMITTEE FOR SPACE DATA SYSTEMS (CCSDS). (2002) *Reference model for an Open Archival Information System (OAIS)*, recommendation. Disponível na Internet em: <<http://www.ccsds.org/documents/>>. Acesso: 14/09/2008.

DURANTI, L. (2007). *An Overview of InterPARES 3 (2007-2012)*. Disponível em: <<http://www.interpares.org/>>. Acesso em 10/01/2011.

HEDSTROM, M. Digital preservation: a time bomb for digital libraries. *Language Resources and Evaluation*, Springer (Netherlands), v. 31, n.3, p. 189 – 202, May. 1997.

IKEMATU, R. Si. 2001. *Gestão de Metadados: sua evolução na tecnologia da informação*. In: *Datagramazero – Revista de Ciência da Informação*, v. 2, num. 6, Instituto Brasileiro de Adaptação e Inserção na Sociedade da Informação.



LAZINGER, S. S. *Digital preservation and metadata: history, theory, practice*. Englewood: Libraries Unlimited, 2001. 359p.

LUSENET, Y. (2007). *Towards Open Access in Europe*. Disponível em: <<http://www.knaw.nl/ecpa/publications.html>>. Acesso em: 01/02/2010.

NATIONAL LIBRARY OF AUSTRALIA (NLA) (1999). *The NLA Preservation Metadata Working Group*. Disponível em na Internet em: <http://www.nla.gov.au/>. Acesso em: 02/11/2008.

THOMAZ, K. P. *A preservação de documentos eletrônicos de caráter arquivístico*. 2004. 388f. Doutorado em Ciência da Informação – Escola de Ciência da Informação. Belo Horizonte. 2004.

UNITED KINGDOM. (2003). National Archives. Uk Government. Digital Preservation Guidance. Preservation guide of National Archives of UK. Disponível em: <<http://www.nationalarchives.gov.uk/documents/>>. Acesso em: 05 mar. 2009.