

PADRONIZAÇÃO DOS FORMATOS DE ARQUIVO: UM CAMINHO PARA PRESERVAR DOCUMENTOS ARQUIVÍSTICOS DIGITAIS

*Henrique Machado dos Santos - UFSM
Dhion Carlos Hedlund - UFRG
Daniel Flores - UFSM*

Resumo: A preservação de documentos arquivísticos digitais é uma necessidade que surge conforme estes documentos vão ganhando relevância, agregando valor social, histórico e informativo. Logo, deve-se pensar em estratégias e políticas de preservação digital que garantam o acesso em longo prazo. Neste contexto, a preservação de muitos documentos arquivísticos irá depender de estratégias baseadas na conversão de formatos, onde se escolhe um novo formato para a transferência da sequência de *bits*. O momento da escolha do novo formato é tido como um ponto fundamental, isto porque pode gerar perdas de informação ou retrabalho de conversão futura. Tendo em vista a falta de critérios para a escolha dos formatos de arquivo para a preservação, este artigo tem por objetivo definir aspectos relevantes para avaliação

1 INTRODUÇÃO

Os avanços das tecnologias da informação e comunicação modificaram os hábitos da sociedade no que tange a troca de informações. Dentro do campo de atuação da Arquivologia não foi diferente, documentos começaram a ser produzidos em meio digital, e, devido a sua relação orgânica com os demais

e escolha destes formatos. Para isto, realiza-se um estudo de natureza aplicada, com abordagem qualitativa, baseado no levantamento bibliográfico. Desta forma, destacam-se entre os resultados o uso de formatos de código aberto, amplamente utilizados e não comprimidos. Por fim, ainda deve-se chamar a atenção para a qualidade do conversor utilizado no processo, além preservar os metadados do documento original, bem como incluir os metadados referentes ao registro da sequência de conversões realizadas.

Palavras-chave: Documento arquivístico digital; Preservação digital; Acesso em longo prazo; Formatos de arquivo.

documentos, passaram a ser denominados arquivísticos digitais. A preservação de documentos arquivísticos digitais tornou-se necessária conforme estes documentos ganharam relevância social, histórica e informativa. Entretanto, este contexto de crescente proliferação de documentos digitais vem ameaçando a capacidade de continuar utilizando os arquivos como fontes confiáveis. Tais fatos são justificados pelos desafios impostos

pela preservação da integridade e da autenticidade em longo prazo (THOMAZ, 2005). Devido a isto há uma necessidade de se pensar em estratégias e políticas de preservação digital que garantam o acesso contínuo em longo prazo a documentos arquivísticos digitais íntegros autênticos, satisfazendo assim, as expectativas da comunidade de pesquisa.

Inicialmente a preservação de documentos arquivísticos digitais irá depender da definição das políticas institucionais que prevejam ações e responsabilidade para a preservação do patrimônio digital. Em um segundo momento, vem à escolha das estratégias que serão contempladas no plano de preservação digital, de forma que satisfaçam os requisitos diplomáticos.

Uma política de preservação ideal deve explicitar de forma clara quais as estratégias escolhidas para efetuar a preservação dos documentos em seus diferentes níveis de abstração (físico, lógico e conceitual). Mas além destes, também há outros níveis considerados superiores, que são o social, o econômico e o organizacional (FERREIRA, 2006). É preciso considerar tanto as especificidades dos documentos, quanto a sua implicação no contexto onde está situado. Logo, há diferentes estratégias que podem ser implementadas para as variações nos níveis sociais, econômicos e organizacionais.

Dentre as possíveis estratégias que o plano de preservação poderá contemplar, pode-se citar, por exemplo, a migração, a emulação, o encapsulamento, a preservação de

tecnologia e o refrescamento. Delimita-se o tema deste artigo para as estratégias de migração, abordando aspectos como a necessidade de conversão de formatos, migração de versão e os formatos escolhidos para a preservação em longo prazo.

Entende-se que a preservação de uma quantidade considerável de documentos arquivísticos irá depender das estratégias fundamentadas na conversão de formatos, onde se escolhe um novo formato para a transferência da sequência de *bits*, procedimento que ocorre, por exemplo, quando o formato atual encontra-se sob o risco de obsolescência tecnológica. Ressalta-se que o momento da escolha do novo formato para preservação é um ponto fundamental, pois qualquer falha no conversor utilizado, conversão para um formato com compactação ou conversão para formato com tendência obsoleta, pode gerar perdas de informação ou retrabalho em conversões futuras.

Tendo em vista a falta de critérios para a escolha dos formatos de arquivo na preservação de longo prazo, este artigo tem por objetivo realizar uma reflexão sobre as características relevantes para avaliação e escolha destes formatos. Do ponto de vista metodológico, este estudo se caracteriza como sendo de natureza aplicada, abordagem qualitativa, com base no levantamento bibliográfico de materiais previamente publicados. Desta forma, apresenta uma breve revisão de literatura sobre o tema buscando embasar os principais entraves teóricos (GIL, 1991; LUNA, 1997; SILVA; MENEZES, 2005).

2 PRESERVAÇÃO DIGITAL

Desde que a informação começou a ser registrada, seja no desenho ou na escrita, houve também o nascimento de novos suportes para registro da memória. A partir desse momento, mais facilmente poderiam surgir às lembranças e permaneceriam fixadas na memória por muito mais tempo. Graças a esses suportes, grandes acontecimentos políticos, sociais e culturais das mais diversas civilizações puderam ser preservados.

Porém, muitos desses registros se perderam ao longo do tempo, ou estão em fase terminal de deterioração devido à despreocupação e/ou a inexistência de métodos científicos adequados de preservação, aplicados aos documentos, de acordo com o material que é constituído. Assim também acontece nos dias de hoje com os documentos digitais em suportes eletrônicos.

Essas são algumas das causas de deterioração e até mesmo da perda da informação contida no documento. Sendo assim, a preservação de documentos tem por objetivo assegurar a integridade, autenticidade e acessibilidade em longo prazo dos suportes documentais e da informação contida neles.

A expressão “Preservação Digital”, que vem sendo utilizada desde os anos 90 deste século (THOMAZ, 2004), enfoca tanto documentos nato-digitais, aqueles gerados diretamente em meio digital, quanto àqueles convertidos para o formato digital. Desta forma, as atividades de preservação digital têm como base a elaborar o planejamento, a fim de alocar recursos e aplicar os métodos de preservação e as

tecnologias necessárias para garantir o acesso em longo prazo à informação digital (HEDSTROM, 1998).

Os problemas que envolvem a preservação digital crescem rapidamente nos dias de hoje. A fragilidade intrínseca do armazenamento digital, mais especificamente a degradação física do suporte é umas das principais ameaças ao documento digital. Contudo, não basta preservar apenas o suporte/mídia onde estão armazenados os documentos digitais, pois tal atitude não garante a acessibilidade deste em longo prazo. Outros fatores dificultam a preservação desses documentos, como por exemplo, o formato de arquivo utilizado. Assim, percebemos que a rápida obsolescência da tecnologia digital, quanto ao *hardware*, *software* e formatos de arquivos, coloca em risco a acessibilidade e a integridade do documento. Dessa forma, pode-se questionar: quem irá garantir que o *software* decodificador deste documento permanecerá estável ao longo do tempo possibilitando a leitura do documento? Ou se não houver mais suporte a esse *software* ou formato de documento armazenado? E se o sistema operacional utilizado não estiver mais disponível no mercado?

Tanto a urgência, quanto as dimensões do problema não envolvem somente a comunidade arquivística. Deve-se ressaltar que os produtores de informação e os fabricantes de *software* precisam levar em consideração métodos de garantir a preservação em longo prazo (THOMAZ, 2005). Há a necessidade de tratar o problema em sua raiz, fazendo assim, que a preservação seja

vislumbrada desde a produção dos conteúdos digitais.

Essas discussões têm sido impulsionadas pelos gestores da informação que, habituados a trabalharem com esse tipo de documentação, viram-se obrigados a encontrar uma solução, no mínimo para suprir as necessidades do momento. Para isso, a utilização de metadados e padrões abertos têm-se mostrado eficiente, proporcionando flexibilidades na gestão documental e facilitando a adoção de estratégias para a preservação digital. Igualmente, se torna um recurso acessível para as instituições, tendo em vista a complexidade e custos da preservação digital. Todavia, é necessário que os sistemas incorporem os conceitos arquivísticos e suas implicações na construção e utilização desses, visando assegurar as características fundamentais de confiabilidade e autenticidade dos documentos.

Diante dessa problemática, muitas instituições preocupadas em salvaguardar seus acervos documentais e órgãos do governo federal responsáveis pela criação e divulgação de políticas de gestão e preservação de documentos, vêm desenvolvendo e auxiliando inúmeras pesquisas, onde procuram abranger também o universo dos documentos digitais.

Entretanto, há de se considerar a falta de preparo das organizações, que de maneira geral, não estão habituadas a preservar documentos digitais ou analógicos. Tendo em vista a exceção de bibliotecas, arquivos e centros de documentação, ainda é difícil imaginar quais as organizações que estão preparadas para preservar documentos digitais em longo prazo. A princípio, não é um exagero realçar

que as próprias organizações responsáveis pela preservação da memória estão com dificuldades de adaptação à realidade digital. É imprescindível reformular conceitos e desenvolver novas teorias que contemplem os documentos de natureza digital (SANTOS; FLORES, 2015a).

Antes de se fazer qualquer aprofundamento teórico, a necessidade da sociedade contemporânea reconhecer as dimensões do problema da preservação digital. Neste sentido, o Conselho Nacional de Arquivo (CONARQ) aprovou e divulgou em 2004, a “Carta para a Preservação do Patrimônio Arquivístico Digital - Preservar para garantir o acesso”, que tem por objetivo alertar os produtores de informação e pesquisadores da área para a urgente necessidade de salvaguardar o patrimônio arquivístico digital. Desta forma, objetiva-se garantir o acesso contínuo aos conteúdos dos registros eletrônicos, considerado uma condição fundamental para democratizar a informação arquivística no Brasil, bem como preservar sua memória (CONARQ, 2004).

2.1 POLÍTICAS DE PRESERVAÇÃO DIGITAL

Para enfrentar os problemas em relação à acessibilidade e à manutenção da integridade do documento digital em longo prazo, se faz necessário estabelecer políticas de uso e manutenção de *hardware* e *software*, além de adotar estratégias de preservação digital.

Entende-se por política de preservação digital, o conjunto de normas, procedimentos e estratégias

que devem ser seguidas para alcançar uma estrutura técnica capaz de manter os documentos digitais íntegros, autênticos e acessíveis em longo prazo, garantindo a sua confiabilidade. Ao mesmo tempo, deve estabelecer como estes métodos serão implementados.

A política de preservação digital deve envolver todos os setores da instituição que sejam afetados pelo processo de preservação, isto compreende as unidades de gestão documental, as administrativas e de tecnologia da informação. Dentre as definições a serem consideradas, destacam-se algumas questões pertinentes como a necessidade de avaliar e selecionar os materiais utilizados; identificar os padrões de metadados apropriados; definir as estratégias de preservação para cada classe de objeto digital; criar planos de sucessão caso as atividades de preservação sejam interrompidas; e utilizar modelos sustentáveis para financiamento do projeto de preservação (FERREIRA, 2006). Acredita-se que através destes pontos seja possível articular o planejamento da preservação digital de forma segura e eficaz.

Após a definição da política de preservação digital, é possível elaborar o plano de preservação especificando as estratégias com eficácia comprovada e de conhecimento do acervo. Durante a elaboração do plano devem-se considerar as propriedades significativas dos documentos que precisam ser preservadas. Desta forma, atenta-se para a conformidade

com os conceitos de forma fixa¹ e conteúdo estável², considerados fundamentais para a preservação de documentos digitais fidedignos (SANTOS; FLORES, 2015d).

2.2 ESTRATÉGIAS DE PRESERVAÇÃO DIGITAL

As estratégias para a preservação digital representam as ações concretas a serem executadas no âmbito das organizações com vistas à preservação de documentos em longo prazo. Podem-se dividir as estratégias de preservação digital em duas categorias: estratégias estruturais e estratégias operacionais (THOMAZ, 2004).

As estratégias estruturais dizem respeito às políticas da instituição, estabelecendo um ambiente para a criação de documentos arquivísticos digitais. As ações que essa categoria abrange são: aderência a padrões; elaboração de manuais; adoção de metadados; investimento em infraestrutura; e formação de rede de relações. Já as estratégias operacionais se referem às atividades concretas para a preservação do documento digital. As ações que essa categoria abrange são: seleção do meio adequado de armazenamento; aplicação da migração; aplicação da emulação; impressão em papel ou microfilme; conservação da tecnologia; aplicação da arqueologia digital (MÁRDERO ARELLANO, 2004; THOMAZ, 2004).

O foco de preservação das estratégias se concentra nos níveis físico, lógico ou conceitual do objeto

¹ Documento arquivístico o qual assegura a mesma apresentação quando recuperado (INTERPARES, 2007).

² Documento arquivístico que torna imutáveis as informações e os dados nele contidos,

exigindo assim, que eventuais alterações sejam feitas por meio do acréscimo de atualizações ou através da produção de uma nova versão (INTERPARES, 2007).

digital (THIBODEAU, 2002). Mesmo assim, nenhuma estratégia se mostrou eficaz a ponto ser aplicada a todos os tipos de documentos digitais (CONAR, 2004; FERREIRA, 2006). Isto vem reforçando a necessidade de empreender e explorar as estratégias existentes para estabelecer um conhecimento sólido de suas aplicabilidades (SANTOS; FLORES, 2015b).

Haverá casos em que os documentos não estão nos formatos desejados o mais comum é se proceder a conversão de formatos, caracterizada como uma espécie de migração. Desta forma é possível adotar padrões para a preservação em longo prazo.

2.2.1 Estratégias de migração

As estratégias migração consistem em transferir de forma periódica os objetos digitais de uma dada configuração de *hardware/software* para outra, ou de uma geração de tecnologia para uma subsequente (TASK FORCE ON ARCHIVING OF DIGITAL INFORMATION, 1996 apud FERREIRA, 2006).

As estratégias de migração, embora não possam ser aplicadas para todos os objetos digitais³, configuram-se como a melhor alternativa para a preservação digital. Isso porque possibilitam que os objetos digitais oriundos de plataformas antigas possam ser migrados e interpretados em plataformas atuais. Mesmo que apresente perdas, a migração/conversão possui vantagens relevantes, como é o caso

da possibilidade de transposição de um objeto criado em um contexto do passado para a atualidade. Além disso, um sistema que implemente padrões de metadados, para documentar a sua custódia, e ainda possibilite retroceder ao objeto digital original, aumentará significativamente os níveis de confiança nesta estratégia, sendo possível a sua implementação em longo prazo (SANTOS; FLORES, 2015c, p. 248).

Por muitas vezes os termos migração e conversão são empregados como sinônimos, porém há uma pequena diferença a fim de adotada neste trabalho. Estes são procedimentos muito semelhantes em termos de preservação digital, ambas consistem em atualizar os objetos digitais para tecnologias atuais a fim de que possa interpretar corretamente o documento.

A migração consiste na atualização sistemática dos objetos digitais que integram o documento digital. Consiste na atualização das versões dos documentos produzidos por um determinado *software*, podendo variar o formato do objeto lógico, desde que este continue vinculado ao mesmo *software* o qual o originou o documento pioneiro. A migração poderá se manifestar de duas formas: mudando a versão ou o formato. No caso da mudança de versão pode-se citar a migração do formato (.doc) na versão do *Word* 1997 para o (.doc) na versão do *Word* 2003, a versão do *software* mudará, mas o formato continuará sendo o mesmo. Já no caso migração do (.doc) para (.docx) formato e a versão do *software* mudarão, mas o *software*

³ Definição genérica que corresponde a todos os tipos de materiais representados em meio digital.

Desta forma abrangem os documentos e os seus componentes digitais.

interpretador continuará sendo o mesmo.

A conversão por sua vez, é entendida como o procedimento de atualização de formato, onde ocorrerá obrigatoriamente a mudança do formato do objeto digital e do *software* utilizado para interpretação. Um exemplo será a conversão do formato (.docx) originado e interpretado no *software Word* para o formato (.odt) para ser interpretado no *software Writer*, assim o formato e o *software* interpretador mudarão sua base tecnológica.

2.2.2 O papel dos metadados na preservação de longo prazo

A implementação de estratégias é uma condição para preservar de documentos digitais em longo prazo. Porém não podem ser executadas de maneira arbitrária, é preciso conhecer as especificidades dos documentos e dos respectivos métodos a serem empregados. Neste sentido, salienta-se o uso de padrões de metadados, os quais vão auxiliar na manutenção do contexto de preservação dos documentos.

Ao implementar estratégias de preservação digital, de natureza operacional, como a migração, torna-se prudente combiná-la com uma estratégia estrutural: a adoção de metadados. Durante a atualização de versões ou conversão dos formatos será necessário preservar informações referentes a todos os componentes digitais que integram o documento.

Os metadados exercem um papel importante em qualquer estratégia de migração para obtenção de bons resultados. A migração dependerá dos metadados criados para registrar a história das ações realizadas sobre os

componentes do documento digital. Além disso, há necessidade de preservar a informação do contexto para que no futuro, para que assim, os usuários compreendam o ambiente tecnológico em que o documento digital e seus respectivos componentes foram criados (MÁREDRO ARELLANO, 2008).

Pode-se afirmar que a adoção de padrões de metadados é fundamental nas atividades de preservação digital, isto porque possibilitam a verificação da integridade e a garantia de autenticidade, uma vez que documentam toda e qualquer alteração realizada sobre os documentos e seus respectivos componentes digitais.

Metadados, portanto, se constituem em componentes do documento eletrônico arquivístico e em instrumentos para sua análise diplomática. É através do domínio desse tipo de análise que será possível estabelecer métodos que garantam a fidedignidade e a autenticidade do documento eletrônico arquivístico. [...], além disso, tanto a decomposição analítica dos documentos arquivísticos, como os mecanismos de garantia de sua integridade pressupõem uma estrutura de procedimentos que gera outros metadados e que se constitui no chamado sistema de gerenciamento arquivístico de documentos (RONDINELLI, 2005, p. 2).

Desta forma, entende-se que os sistemas de gestão e preservação devem contemplar e preservar os metadados dos componentes que integram os documentos arquivísticos digitais. A migração de versão ou a conversão para novos formatos implica na preservação dos

metadados, desta forma, objetiva-se a comprovação da autenticidade.

3 FORMATOS DE ARQUIVO RECOMENDADOS PARA PRESERVAÇÃO

As atividades de preservação digital pressupõem a implementação de estratégias, dentre elas as estratégias operacionais. Como consequências destas ações, deve-se identificar e avaliar, as próprias estratégias escolhidas, o *software*, o suporte e os respectivos formatos de arquivo para preservação.

Neste contexto, o Conarq (2004) manifesta explicitamente o compromisso de definir e/ou recomendar o uso de padrões e protocolos abertos e sua ampla aceitação para criação, uso, transmissão e armazenamento de documentos digitais (CONARQ, 2004). Os padrões abertos, sejam eles formatos de arquivos ou *softwares*, são aqueles que possuem que suas especificações publicadas e disponibilizadas gratuitamente. Podendo significar que são amplamente utilizados e/ou livres de patentes. (INTERPARES, 2007). Logo, chama-se a atenção para que tanto o *software* originador quanto o formato de arquivo originado possuam código aberto.

Dentre a variedade de formatos de arquivo disponíveis, devem-se escolher formatos que tenham uma expectativa de acesso em longo prazo, exemplo disto são os formatos utilizados em larga escala. Recomenda-se também os formatos de arquivo de código aberto, isto se deve a possibilidade de reconstrução do formato em caso de obsolescência.

Utilizar *softwares* específicos como padrões fechados torna-se um

problema para a preservação digital. Tal fato se justifica, pois o padrão fechado transforma o documento em uma 'caixa preta', não havendo nenhuma possibilidade de acesso as suas especificações por parte acervo. Neste sentido, devem-se eliminar as dependências específicas, em especial os *softwares* proprietários de código fechado. Optar por *softwares* de padrões abertos reduzirá o risco da perda de documentos digitais nos momentos críticos de migração tecnológica (INNARELLI, 2007). Logo, emuladores, conversores, padrões de metadados, *softwares* de produção, gestão, repositórios digitais e demais *softwares* utilizados no acervo, devem ser prioritariamente de código aberto.

Estudos e iniciativas estão surgindo no mundo todo, com estes, conclui-se que devem ser usados padrões e converter os documentos para formatos livres, para que sejam acessados após a obsolescência de *hardware* e *software* em que foram criados. O uso de padrões abertos permite seu estudo e sua conversão para novos padrões (MÁRDERO ARELLANO, 2004). Este é um ponto que aumenta as chances de sucesso da migração para novas tecnologias.

Dentre os estudos em preservação digital, com relação aos formatos de arquivo, destaca-se o Arquivo da Web Portuguesa (AWP), uma iniciativa da Fundação para a Computação Científica Nacional (FCCN), de acesso gratuito com objetivo de preservar o acesso em longo prazo da informação na Web de interesse à comunidade portuguesa. Este apresenta uma lista de formatos de arquivo, classificados de acordo com um grau de recomendação (alta, média e baixa), para documentos na forma de texto, imagem, áudio, vídeo e outros formatos (AWP, 2007).

De acordo com AWP (2007) um formato adequado para preservação deve ser:

- Livre de quaisquer direitos que restrinjam sua utilização;
- Uma norma emitida por organismo oficial;
- Documentado com o seu respectivo código fonte aberto;
- Amplamente utilizado;
- Multiplataforma com os respectivos códigos abertos;
- Um formato sem compressão ou de compressão sem perdas.

De tal forma, um formato com características fracas de preservação será aquele que:

- Possui direito proprietário com especificação fechada;
- É pouco usado/difundido;
- É lido e escrito através de poucas plataformas de *software* de código fechado;
- Possui compressão com perda de informação;
- Seja composto por elementos internos como, por exemplo, macros.

Desta forma, identificou-se que os critérios utilizados são de extrema relevância para a recomendação de formatos. Embora sejam recomendados para a preservação do conteúdo de *sites*, podem ser adotados, por exemplo, para a preservação de documentos pessoais e institucionais. Dentre os fatores a se destacar estão o código aberto, a multiplataforma, a ampla utilização, a compressão sem perdas e as normas, entende-se que estes devem ser requisitos para a definição de qualquer padrão. Além disso, há de se

entender que a *Web* possui as suas especificidades, e para seu melhor desempenho necessita de formatos “leves”, fator que em alguns aspectos pode deixar a preservação em segundo plano, priorizando o acesso.

Em linhas gerais, a proposta de formatos recomendados para a preservação do AWP (2007), vai de encontro ao recomendado por Márdero Arellano (2004), Innarelli (2007), projeto Interpares (2007) e Conarq (2004). A preservação em longo prazo necessita de padrões abertos, usados em larga escala, sem compressão ou de compressão sem perdas, acessíveis em multiplataformas e normalizados. Este é o perfil de formatos de arquivo que os administradores dos repositórios digitais devem solicitar aos produtores de conteúdos.

4 PRODUÇÃO DE DOCUMENTOS DIGITAIS VISLUMBRANDO A PRESERVAÇÃO

A preservação digital envolve todo o ciclo de vida dos documentos, desde a sua produção, passando pela tramitação até a sua guarda permanente. Logo, nestas cadeias de custódia, de fase corrente e intermediária, e fase permanente, deverão ser implementados procedimentos para a preservação dos objetos digitais.

As fases corrente e intermediária devem ser gerenciadas por um Sistema Informatizado de Gestão Arquivística de Documentos (SIGAD), este deverá contar com ferramentas que garantam a manutenção da integridade e da autenticidade dos documentos digitais. O SIGAD deve ser um sistema integrado com o repositório digital, desta forma, chama-se a atenção para

a necessidade de garantir a interoperabilidade entre estas tecnologias. Além disso, deve-se entender que os documentos de fase corrente e intermediária possuirão um determinado período de permanência no SIGAD até serem recolhidos ao repositório. Este período poderá ser o suficiente para que ocorra algum problema de obsolescência tecnológica, desta forma, chama-se a atenção para a implementação de repositórios digitais em fase corrente/intermediária.

Ressalta-se que os documentos digitais não podem ser armazenados em formatos obsoletos, para serem transferidos para os repositórios digitais somente após longos períodos. Tal fato se justifica, pois há o risco da tecnologia ser incompatível e assim não conseguir recuperar as informações. Esta é uma preocupação que envolve tanto os produtores de conteúdos digitais, quanto os responsáveis pela sua preservação de longo prazo (MÁRDERO ARELLANO, 2008).

A obsolescência tecnológica é um reflexo da constante evolução, a qual influencia diretamente os documentos digitais. Logo, com intervalos de evolução tecnológica, cada vez menores, torna-se fundamental a ação imediata a cada alteração dos componentes do sistema de arquivamento digital (THOMAZ; SOARES, 2004). Desta forma, pode-se dizer que a implementação de repositórios digitais na fase corrente/intermediária é altamente recomendável, adicionando confiabilidade aos documentos digitais.

Além disso, grande parte do “sucesso” das atividades de preservação digital dependerá dos

formatos de arquivo utilizados na produção. Nesta fase, deve-se atentar para o uso de padrões abertos, normalizados, sem compressão, ou de compressão sem perdas e que sejam amplamente utilizados.

Para assegurar a gestão de documentos digitais e o seu acesso em longo prazo, padrões de formatos devem ser planejados e implementados. Estes padrões, além de garantirem a gestão e o acesso aos documentos de guarda permanente, garantirão a autenticidade dos documentos digitais no seu ambiente ou nos processos de migração (INNARELLI, 2007; 2012). A partir do momento em que os documentos são produzidos de acordo com os formatos recomendados para a preservação, diminuem-se os riscos de obsolescência, perda ou corrupção de dados.

Partindo das recomendações analisadas, identificou-se que o formato (.odt) é um dos ideais para a preservação de documentos textuais (AWP, 2007). Sendo assim, e a produção de documentos textuais for padronizada para este formato, teoricamente, diminuem as necessidades de migração e de outras estratégias que sejam necessárias. A partir do momento em que se faz a submissão de um documento que já está em conformidade com o plano de preservação do repositório, procede-se o seu armazenamento, sem necessidade de convertê-lo para outro formato, diferentemente do que aconteceria com um documento em formato (.doc). A mesma linha de pensamento se propõe para os documentos de outras naturezas como áudio, vídeo, imagem, entre outros.

Além de produzir documentos em formatos adequados para a preservação, é fundamental

implementar padrões de metadados que documentem toda e qualquer estratégia envolvida no processo de preservação digital (SARAMAGO, 2004). Entretanto, deve-se compreender que os padrões podem mudar, e desta forma, será preciso converter os formatos. Logo, a criação de padrões de metadados de preservação, e a sua adaptação, se configura como um componente-chave para a implementação de grande parte das estratégias de preservação digital (THOMAZ; SOARES, 2004). Tudo está em constante mudança: os *softwares*, os formatos de arquivos e os padrões de metadados.

Por fim, a produção de documentos digitais deve ser projetada em conformidade com os padrões de preservação. Desta forma, minimizam-se as atividades técnicas e os esforços financeiros para a preservação destes documentos. Entretanto, é preciso compreender que estes formatos podem mudar com o tempo, então o monitoramento das tendências se torna fundamental, a fim de evitar a produção em formatos potencialmente obsoletos. Por isso, é altamente recomendável o uso de formatos abertos, amplamente utilizados ou padrões ISO, estes formatos proporcionarão maior chance de recuperação em caso de obsolescência tecnológica.

5 CONCLUSÃO

Tendo em vista as necessidades de preservar o patrimônio arquivístico em meio digital, fazem necessários procedimentos de ordem estrutural e operacional. Dentre estes, destaca-se a importância da definição de uma política de preservação que satisfaça os requisitos arquivísticos e

diplomáticos. Estratégias como a migração, baseadas na conversão de formatos e atualização de versão, têm por necessidade a adoção de metadados para garantir a sua autenticidade. Da mesma forma, precisam ser executadas por um conversor eficaz e um profissional qualificado. Além disso, a definição dos formatos de arquivo a serem utilizados no acervo é uma questão fundamental na preservação em longo prazo.

A adoção de *software* livre e formatos de arquivo abertos são fundamentais ao se implementar estratégias de preservação digital. A guarda de documentos digitais está implicitamente ligada à utilização de padrões, entre eles, os formatos de arquivo recomendados para preservação e os *softwares* que geram estes formatos. Logo, a adoção de padrões abertos garante o acesso ao código fonte de ambos, o que torna possível a sua reconstrução caso seja necessário. Padrões recomendados por normas, ou amplamente utilizados possibilitam acesso ao conteúdo documento, além da perspectiva de longevidade.

Na medida em que se produzem documentos arquivísticos, a instituição deverá pensar na sua preservação em longo prazo e adotar formatos padronizados desde a produção. Desta forma, ao se produzir documentos seguindo os padrões de preservação, simplifica-se substancialmente as atividades de preservação dos documentos digitais. Logo, a produção deverá ser voltada para a preservação, podendo adotar os mesmos formatos recomendados para a preservação. No contexto dos repositórios, os preservadores devem solicitar formatos padronizados juntamente aos produtores,

potencializando as ações de preservação.

Em linhas gerais, as comunidades de preservação digital, as instituições e pessoas envolvidas, devem estar cientes de que a preservação digital necessita de políticas que contemplem requisitos arquivísticos e diplomáticos definidos *a priori*. Pode-se afirmar que o uso da tecnologia isoladamente não resolverá os problemas de preservação digital, por isto, deve haver o equilíbrio, uma sincronia entre as políticas de preservação digital e a implementação de tecnologias. Na preservação de documentos arquivísticos digitais, a tecnologia da informação é apenas um meio para atingir os objetivos, a base se encontra nos referenciais da Arquivologia e da Diplomática, ambos apresentam os requisitos necessários para preservar a fidedignidade dos documentos.

STANDARDIZATION OF FILE FORMATS: A WAY TO PRESERVING THE DIGITAL ARCHIVAL DOCUMENTS

Abstract: *The preservation of digital archival documents is a need that arises as these documents are gaining relevance, adding value social, historical and informative. Therefore, one should think of digital preservation strategies and policies that guarantee access in the long term. In this context, the preservation of archival documents will depend on many strategies based on conversion format, which is choosing a new format for the transfer bit sequence. The moment of choice of the new format is seen as a key point, because this can cause loss of information or rework future conversion. Given the lack of criteria*

for the choice of file formats for preservation, this article is to define relevant to evaluation and choice of these formats. For this, we make a study of applied research, with a qualitative approach, based on the literature. Thus, we highlight the results of the use of open source formats, widely used and uncompressed. Finally, I must draw attention to the quality converter used in the process, and preserve the metadata of the original document and include the metadata for the record of the sequence of conversions performed.

Keywords: *Digital archival document; Digital preservation; Access in long-term; File formats.*

Sobre os autores

Henrique Machado dos Santos

henrique.hms.br@gmail.com
Graduado em Arquivologia pela
Universidade Federal de Santa Maria

Dhion Carlos Hedlund

dhion.hedlund@gmail.com
Graduado em Arquivologia pela UFSM
Mestre em Patrimônio Cultural pela
Universidade Federal de Santa Maria
Professor Assistente do Curso de
Arquivologia da Universidade Federal do
Rio Grande

Daniel Flores

dfloresbr@gmail.com
Professor Associado do Departamento de
Documentação da Universidade Federal
de Santa Maria. Docente do Mestrado
Profissional em Patrimônio
Cultural/PPGPPC-UFSM, Docente
Colaborador do Mestrado Profissional
em Gestão de Documentos e Arquivos -
PPGARQ - UNIRIO
Doutor em Ciência da Informação

REFERÊNCIAS

ARQUIVO DA WEB PORTUGUESA (AWP). **Formatos adequados para preservação**. Portugal: Lisboa, 2007. Disponível em: <<http://sobre.arquivo.pt/colaboracoes/recomendacoes-para-autores-de-sitios-web/formatos-adequados-para-preservacao>>. Acesso em: 04 nov. 2014.

CONSELHO NACIONAL DE ARQUIVOS (CONARQ). Câmara Técnica de documentos eletrônicos. **Carta para a Preservação do Patrimônio Arquivístico Digital**. Rio de Janeiro: Arquivo Nacional, 2004. Disponível em: <<http://www.conarq.arquivonacional.gov.br/Media/publicacoes/cartapreservpatrimarqdigitalconarq2004.pdf>>. Acesso em: 10 ago. 2014.

FERREIRA, Miguel. **Introdução à preservação digital**: conceitos, estratégias e atuais consensos, Portugal: Escola de Engenharia da Universidade do Minho, 2006. Disponível em: <<https://repositorium.sdum.uminho.pt/bitstream/1822/5820/1/livro.pdf>>. Acesso em: 02 ago. 2014.

GIL, Antônio Carlos. **Como elaborar projetos de pesquisa**. São Paulo: Atlas, 1991.

HEDSTROM, Margaret. Digital preservation: a time bomb for digital libraries. **Computer and the humanities**, Netherlands, n. 31, p. 189-202, 1998. Disponível em: <http://deepblue.lib.umich.edu/bitstream/2027.42/42573/1/10579_2004_Article_153071.pdf>. Acesso em: 01 out. 2014.

INNARELLI, Humberto Celeste, **Instrumenta 2**: Preservação de Documentos Digitais. Associação dos Arquivistas de São Paulo. São Paulo: ARQ-SP, 2012. 64p.

INNARELLI, Humberto Celeste. **Preservação digital e seus dez mandamentos**. In: SANTOS, Vanderlei Batista (Org.). Arquivística: temas contemporâneos, classificação, preservação digital, gestão do conhecimento. Distrito Federal: SENAC, 2007. p. 21-75.

INTERPARES 2 PROJECT. **Diretrizes do Produtor**. A elaboração e a manutenção de materiais digitais: diretrizes para indivíduos. TEAM Brasil. Tradução: Arquivo Nacional e Câmara dos Deputados. 2002 – 2007. Disponível em: <http://www.interpares.org/ip2/display_file.cfm?doc=ip2_creator_guidelines_booklet--portuguese.pdf>. Acesso em: 09 ago. 2014.

LUNA, Sergio Vasconcelos de. **Planejamento de pesquisa**: uma introdução. São Paulo: EDUC, 1997.

MÁRDERO ARELLANO, Miguel Ángel. **Critérios para a preservação digital da informação científica**. 2008. 354 p. Tese (Doutorado em Ciência da Informação) - Universidade Federal de Brasília. Brasília: Departamento de Ciência da Informação, 2008. Disponível em: <http://bdtd.bce.unb.br/tesesimplificado/tde_busca/arquivo.php?codArquivo=4547>. Acesso em: 15 jun. 2014.

MÁRDERO ARELLANO, Miguel Ángel. Preservação de documentos digitais, **Ciência da Informação**, Brasília, v. 33, n. 2, p. 15-27, maio/ago. 2004. Disponível em: <<http://revista.ibict.br/ciinf/index.php/ciinf/article/view/305/1452>>. Acesso em: 25 jul. 2014.

RONDINELLI, Rosely Curi. **Gerenciamento arquivístico de documentos Eletrônicos**: uma abordagem teórica da diplomática arquivística contemporânea. 4. ed. Rio de Janeiro: Editora FGV, 2005. 160 p.

SANTOS, Henrique Machado dos; FLORES, Daniel. As vulnerabilidades dos documentos digitais: Obsolescência tecnológica e ausência de políticas e práticas de preservação digital. **Biblios: Revista de Bibliotecologia y Ciencias de la Información**, Lima, n. 59, p. 45-54, 2015a. Disponível em: <<http://biblios.pitt.edu/ojs/index.php/biblios/article/view/215>>. Acesso em: 04 jul. 2015.

_____. Estratégias de preservação digital para documentos arquivísticos: uma breve reflexão. **Cadernos BAD**, Lisboa, 2015b, n. 1, jan./jun., p. 87-101. Disponível em: <<http://www.bad.pt/publicacoes/index.php/cadernos/article/view/1225>>. Acesso em: 01 ago 2015.

_____. Preservação de documentos arquivísticos digitais autênticos: reflexões e perspectivas. **Acervo**, Rio de Janeiro, v. 28, n. 1, p. 241-253, jan./jun. 2015c. Disponível em: <<http://www.revistaacervo.an.gov.br/seer/index.php/info/article/view/740>>. Acesso em: 01 jul. 2015.

_____. Repositórios digitais confiáveis para documentos arquivísticos: ponderações sobre a preservação em longo prazo. **Perspectivas em Ciência da Informação**, Belo Horizonte, v.20, n.2, p. 198-218, abr./jun. 2015d. Disponível em: <<http://portaldeperiodicos.eci.ufmg.br/index.php/pci/article/view/2341>>. Acesso em 03 jul. 2015.

SARAMAGO, Maria de Lurdes. Metadados para preservação digital e aplicação do modelo OAIS. In: CONGRESSO NACIONAL DE BIBLIOTECARIOS, ARQUIVISTAS E DOCUMENTALISTAS, 8., 2004, Estoril. **Anais eletrônicos...** Disponível em: <<http://www.bad.pt/publicacoes/index.php/congressosbad/article/view/640/637>>. Acesso em: 04 jul. 2014.

SILVA, Edna Lúcia da; MENEZES, Estera Muszkat. **Metodologia da pesquisa e elaboração de dissertação**. 4. ed. rev. Florianópolis: UFSC, 2005. Disponível em: <https://projetos.inf.ufsc.br/arquivos/Metodologia_de_pesquisa_e_elaboracao_de_teses_e_dissertacoes_4ed.pdf>. Acesso em: 13 jun. 2014.

THIBODEAU, Kenneth. Overview of Technological Approaches to Digital Preservation and Challenges in Coming Years: The State of Digital Preservation: An International Perspective. **Anais eletrônicos...** Washington: CLIR and Library of Congress, 2002. Disponível em: <<http://www.clir.org/pubs/reports/pub107/thibodeau.html>>. Acesso em: 19 ago. 2014.

THOMAZ, Kátia de Pádua. **A preservação de documentos eletrônicos de caráter arquivístico**:

novos desafios, velhos problemas. 2004. 389f. Tese (Doutorado em Ciência da Informação) - Universidade Federal de Minas Gerais. Belo Horizonte: Escola de Ciência da Informação, 2004. Disponível em: <http://www.bibliotecadigital.ufmg.br/dspace/bitstream/handle/1843/VALA-68ZRKF/doutorado__katia_de_padua_thomaz.pdf>. Acesso em: 28 jul. 2014.

_____. Documentos eletrônicos de caráter arquivístico: fatores

condicionantes da preservação. **Perspectivas em Ciência da Informação**, Belo Horizonte, v.10 n.1, p. 34-53, jan./jun. 2005. Disponível em: <www.brapci.ufpr.br/download.php?dd0=13204>. Acesso em: 07 set. 2014.

_____; SOARES, Antônio José. A preservação digital e o modelo de referência Open Archival Information System (OAIS). **DataGramZero**, v.5 n.1, fev./2004. Disponível em: <www.dgz.org.br/fev04/Art_01.htm>. Acesso em: 10 jun. 2014.