UNIVERSIDADE DE SÃO PAULO
ESCOLA DE COMUNICAÇÃO E ARTES
DEPARTAMENTO DE INFORMAÇÃO E CULTURA

Guilherme Marques de Mattos

Políticas e Estratégias de Preservação Digital: Uma Abordagem Tecnológica

São Paulo 2020

UNIVERSIDADE DE SÃO PAULO
ESCOLA DE COMUNICAÇÃO E ARTES
DEPARTAMENTO DE INFORMAÇÃO E CULTURA

Guilherme Marques de Mattos

Políticas e Estratégias de Preservação Digital: Uma Abordagem Tecnológica

Trabalho de Conclusão de Curso apresentado como exigência parcial para a obtenção do título de Bacharel em Biblioteconomia, ao Departamento de Informação e Cultura da Escola de Comunicação e Artes da Universidade de São Paulo.
Orientador: Prof. Dr. Francisco Carlos Paletta.

São Paulo 2020

Catalogação na Publicação
Serviço de Biblioteca e Documentação
Escola de Comunicações e Artes da Universidade de São Paulo
Dados inseridos pelo(a) autor(a)

Mattos, Guilherme Marques de
Politicas e Estratégias de Preservação Digital: Uma Abordagem Tecnológica / Guilherme Marques de Mattos ; orientador, Francisco Carlos Paletta. -- São Paulo, 2020.
70 p.: il.

Trabalho de Conclusão de Curso - Departamento de Informação e Cultura/Escola de Comunicações e Artes / Universidade de São Paulo.
Bibliografia
Versão corrigida

1. Preservação Digital 2. Ciência da Informação 3. Preservação da memória 4. Biblioteconomia I. Carlos Paletta, Francisco II. Titulo.

CDD 21.ed. - 020

Elaborado por Alessandra Vieira Canholi Maldonado - CRB-8/6194

MATTOS, Guilherme Marques de

Políticas e Estratégias de Preservação Digital: Uma Abordagem Tecnológica.

Trabalho de Conclusão de Curso apresentado como exigência parcial para a obtenção do título de Bacharel em Biblioteconomia, ao Departamento de Informação e Cultura da Escola de Comunicação e Artes da Universidade de São Paulo.
Orientador: Prof. Dr. Francisco Carlos Paletta.

São Paulo, 23 de junho de 2020

Banca Examinadora

Orientador: Prof. Dr. Francisco Carlos Paletta

Membro: Pedro Luiz Côrtes

Membro: Ivan Claudio Pereira Siqueira

AGRADECIMENTOS

Nenhuma monografia é feita de maneira totalmente solitária, por conta disso agradeço:

- à minha mãe, Vânia Marques Ribeiro, pelo constante apoio durante a graduação;
- ao meu orientador Prof. Dr. Francisco Carlos Paletta pela contumaz disponibilidade e orientação durante a produção deste trabalho;
- às minhas amigas e colegas Adriana de Araujo Neitze e Nádia Lara Muniz Manchini por sua disponibilidade e apontamentos de revisão;
- à minha companheira Cristina Araújo Nogueira do Vale por sua paciência e dedicação em ler e revisar este trabalho;
- a toda comunidade uspiana por ter me mudado para sempre nesses anos de graduação.

RESUMO

MATTOS, Guilherme Marques de. **Políticas e Estratégias de Preservação Digital: Uma Abordagem Tecnológica**. Trabalho de Conclusão de Curso (Bacharelado em Biblioteconomia) - Escola de Comunicação e Artes, Universidade de São Paulo, São Paulo, 2020.

Este trabalho faz uso de revisão bibliográfica para a identificação das melhores práticas de preservação digital que podem ser empregadas nas organizações, tanto privadas quanto públicas, levando em consideração suas especificidades. Tais práticas aliam conceitos biblioteconômicos e de tecnologia da informação, indicando que ambas as áreas de atuação têm de colaborar em conjunto para que a memória de nosso tempo não se perca, visto que os documentos vêm sendo cada vez mais produzidos e distribuídos em meio digital. Dentre as melhores práticas de preservação, este trabalho propôs-se a discutir a gestão de suportes, de *software*, de *hardware*, virtualização para a preservação, cópias de segurança, gestão de metadados e controle de autenticidade de documentos digitais por meio de *hashing,* buscando compreender qual seria o uso ideal dessas técnicas de preservação digital, de acordo com a bibliografia especializada.

Palavras-chave: Preservação Digital. Ciência da Informação. Tecnologia da Informação.

ABSTRACT

MATTOS, Guilherme Marques de. **Policies and Strategies in Digital Preservation: A Technological Approach**. Term Paper (Graduation in Libraryship) - School of Communication and Arts, University of São Paulo, São Paulo, 2020.

This work makes use of bibliographic review to identify the best digital preservation practices that can be used in organizations, in private or public ones, taking into account their specificities. These best practices combine libraryship concepts with information technology concepts, showing that both areas of activity need to collaborate to preserve our time memory, as the registries of our time are gradually being used and distributed in digital way. Among the best conservation practices, this work discussed management of supports, of software, and hardware, virtualization for preservation purposes, backup copies, metadata management and hashing for digital document authentication control, and sought to understand the ideal use of these digital preservation techniques according to the specialized bibliography.

Keywords: Digital Preservation. Information Science. Information Technology.

SUMÁRIO

INTRODUÇÃO

Podemos dizer que a explosão informacional ocorrida no século XX, mediante o desenvolvimento da computação, modificou para sempre a forma com que a humanidade passou a produzir suas pesquisas científicas, assim como alterou significativamente seu modo de se comunicar. O compartilhamento de informação com a rede mundial de computadores acarretou uma produção informacional sem precedentes (BARROS; CASTRO; ARELLANO, 2018), obrigando as organizações a paulatinamente criarem estratégias, ou políticas, de preservação digital e, consequentemente, ocasionando a criação de diferentes instituições de memória. Espalhadas pelo mundo, tais instituições unem esforços no desenvolvimento de pesquisas, na produção e na disseminação de metodologias e práticas de preservação digital. Contraditoriamente, tal esforço acadêmico pode, no entanto, não estimular efetivamente a aplicação de tais práticas; ainda assim, há que se considerar sua relevância para a manutenção da memória da humanidade em uma era cada vez mais digital e interconectada.

Com o advento da era digital, novas modalidades de se criar, modificar e desenvolver documentos alteraram a forma com que as organizações lidam com seus documentos e, de forma mais ampla, com a informação, permitindo também a criação de diferentes formas de interação por parte dos usuários, que conseguem expressar a informação que carregam em suas atividades laborais diárias e, assim, produzem grandes quantidades de dados não estruturados por meio do uso das TICs - Tecnologias da Informação e Comunicação (CONSELHO NACIONAL DE ARQUIVOS - CONARQ, 2011). Quando estruturados pelo uso de ferramentas de análise de dados massivos *big data*, tais dados geram resultados que podem auxiliar em tomadas de decisão. Ademais, com a disseminação da arte digital e eletrônica, um outro desafio se coloca para as equipes de curadoria das instituições de memória, já que as diferentes expressões artísticas podem ter características de *hardware* e *software* únicas produzidas pelo artista.

Para dar um exemplo dessas transformações, um documento digital não possui vínculo com o suporte que o expressa, diferentemente de um documento de suporte informacional analógico (CONSELHO NACIONAL DE ARQUIVOS -

CONARQ, 2011). Um livro físico possui dimensões físicas, tinta e papel, passa por deterioração química e física, enquanto um livro digital de conteúdo equivalente possui apenas dimensões lógicas, sendo seu volume quantificado pelos *bits* que ocupa em um dado dispositivo de armazenamento digital. É o dispositivo de armazenamento, e não seu conteúdo, que está exposto a variáveis ambientais de deterioração. O volume lógico de um documento digital é o mais perto que se pode chegar de sua descrição física, tratando-se de eletricidade organizada em lógica binária em um dispositivo eletrônico (LUZ, 2018).

Tais considerações podem nos conduzir ao seguinte questionamento inicial: como classificar um documento digital quando o encontramos, visto que ele não possui as dimensões de um objeto físico nem nada que descreva preliminarmente seu conteúdo, ou mesmo a forma de acessar a informação contida nele? Ora, antes de termos acesso ao conteúdo de um objeto digital, temos acesso aos seus metadados (título, autoria, data de criação, data de edição, entre outros), e estes são pistas para identificarmos preliminarmente seu conteúdo e também verificarmos sua autenticidade (CONSELHO NACIONAL DE ARQUIVOS, 2005). Como o documento digital é efêmero, a autenticidade também tem sua efemeridade. Ao se acessar um documento digital, portanto, lida-se com a assim chamada presunção de autenticidade, que, segundo o Arquivo Nacional (2016, p.31), significa: “Inferência da autenticidade de um documento arquivístico feita a partir de fatos conhecidos sobre a maneira como aquele documento foi produzido e mantido”. Há que se observar, entretanto, que documentos digitais nem sempre são interpretados pelo usuário da mesma maneira que seus pares analógicos. No caso de um *e-book,* por exemplo, o conteúdo será similar ao de um livro impresso, mas sua forma de leitura será diferente (em função de seu formato de apresentação distinto do formato de códice de um livro comum), mais se assemelhando ao uso prático de um rolo de pergaminho, já que precisamos rolar para baixo ou para cima para lermos seu conteúdo textual (LOGAN, 2012a). Destarte, podemos afirmar que arquivos digitais possuem formas de apresentação da informação distintas (ainda que similares em muitos aspectos) de suas versões analógicas. Outro ponto relevante é que um documento digital só pode ser devidamente interpretado por um dispositivo digital com um *software* que traduza sua composição de *bits*. Por

exemplo, no livro digital que nossos olhos veem, o *software* depende de um sistema operacional compatível, que depende de um *hardware* adequado, que está exposto a uma miríade de fatores de mercado, tais como a obsolescência programada. Seja qual for a especificidade do objeto digital, haverá a necessidade de se preservar não apenas o arquivo, mas seu *software* de interpretação e também o *hardware* de suporte (ARELLANO; ANDRADE, 2006). Cumpre destacar que, apesar de não ser tão frequentemente atualizado quanto os *softwares*, o *hardware* requer políticas de preservação a longo prazo, sob o risco de impactar todos os outros processos da preservação digital.

Quando o *hardware* não pode ser mais mantido, seja por conta de mudanças radicais nas tecnologias dos dispositivos ou da falta de componentes eletrônicos para reposição quando os mesmos apresentam defeito, uma possibilidade de solução é a emulação do *hardware*, ou seja, a simulação daquele *hardware* por meio de um *software*, para que se possa executar um sistema operacional fora de uso em conjunto com os *softwares* que também estão em análoga situação de obsolescência (SCHÄFER; CONSTANTE, 2012). Tal prática requer que o *hardware* e as dependências de *software* necessárias para a execução das máquinas virtuais onde ocorrem a emulação estejam também dentro da política de preservação, visto que, a longo prazo, estarão da mesma forma expostas à degradação tecnológica antes mencionada, e invariavelmente sofrerão substituição.

JUSTIFICATIVA

Enquanto ciência que visa dentre seus objetivos a preservação informacional, a Biblioteconomia tem como desafio atual compreender, aplicar e aperfeiçoar metodologias de preservação digital, metodologias estas que são muito distintas das de preservação analógica. A preservação digital nas organizações foi o tema proposto com base na percepção de que, com grande frequência, as práticas de preservação digital nas organizações são confundidas com a mera realização de cópias de segurança, visto que o viés dominante sobre o documento digital é meramente instrumental, colocando a autenticidade e a confiabilidade de sua informação em dúvida sob um olhar mais apurado. Há, portanto, relevância na

discussão do tema sobretudo pela comunidade dos profissionais da informação, pois sem nossa presença na criação de políticas de preservação, a longo prazo, a memória digital tende a se perder e, com ela, nossa história. Consequentemente, a justificativa deste trabalho de conclusão de curso de graduação em Biblioteconomia é viabilizar a discussão, por meio de revisão bibliográfica, das melhores práticas de preservação digital nas organizações. Em nossa passagem pela academia, tornou-se evidente a importância da informação para a manutenção da sociedade, esta que cada vez mais está trafegando sua informação por meio digital de maneira exponencial. A revisão bibliográfica foi importante para conceituar as melhores práticas, já que, como cada organização é um microcosmos, seria imprudente tomar o modelo de determinada organização como o melhor a ser seguido. Uma política de preservação digital terá maior chance de sucesso se moldada aos processos da organização que a define, por isso a seleção dessa abordagem de pesquisa se mostrou mais adequada.

OBJETIVO

Geral: Descrever e analisar as melhores práticas para o desenvolvimento de políticas e estratégias, além de métodos para seleção de *softwares,* que auxiliem o profissional da informação a preservar a informação contida em meio digital nas organizações, assim como as motivações que levam ao desenvolvimento dessas políticas.

Específicos: Levantar bibliografia especializada que apresentem abordagens para criação de políticas ou estratégias de preservação digital nas organizações, buscando, assim, identificar as melhores práticas e as tendências metodológicas mais relevantes sobre o tema.

METODOLOGIA

Em meio ao quadro no qual se encontram as organizações no esforço de preservar sua produção digital, haverá sempre o dilema de como devem preservar, para que possam de fato resguardar a memória ou suas produções para posterior

recuperação de maneira confiável. As organizações se encontram em posição de refletir sobre a institucionalização de políticas de preservação digital que se façam coerentes e tenham um embasamento mais confiável no campo teórico e metodológico. Este trabalho traz metodologias que têm como referência bibliografias baseadas em produção teórico-científica, assim como alguns exemplos de políticas de preservação digital presentes em instituições brasileiras. Por tal característica, este trabalho põe em pauta a reflexão sobre o quanto as instituições, públicas ou privadas, têm dado a devida relevância à aplicação de políticas de preservação da informação em meio digital. Observamos que a bibliografia teórica é farta; faltam, porém, exemplos práticos, o que pode sinalizar as dificuldades de aplicação prática das políticas de preservação digital existentes. Para tanto, busca-se conceituar nos capítulos 1 (As Diferentes Formas de se Definir Informação) e 2 (O Papel da Informação Digital no Século XXI) o que é informação e como ela se tornou objeto tão presente na sociedade contemporânea. Nos capítulos 3 (Em Quais Organizações a Preservação Digital se faz Necessária?) e 4 (Políticas e Estratégias de Preservação Digital), são apresentadas as características que constituem as melhores práticas para o desenvolvimento de políticas de preservação digital e como podem ser aplicadas nas organizações.

CAPÍTULO 1 - AS DIFERENTES FORMAS DE SE DEFINIR INFORMAÇÃO

Apesar de o termo *informação* estar muito presente nas vidas das pessoas no século XXI, conceituar o que é informação é um intrigante desafio. Após o advento das TICs, houve uma explosão na quantidade de profissionais da área de tecnologia da informação que, progressivamente, aproximaram-se das discussões sobre comunicação não apenas entre máquinas, mas também entre pessoas e máquinas. Percebe-se, no entanto, que as definições correntes no meio dos profissionais da TI ainda estão presas a uma forma tecnicista de enxergar informação. Isto implica que esta crescente categoria profissional suporta uma definição de informação embasada pela teoria da informação de Shannon (1948), de assimilação mais simples em nossa sociedade pós-industrial.

De acordo com Shannon, a informação é uma composição de dados, que são transmitidos de um emissor para um receptor e, caso a mensagem tenha sido recebida e compreendida pelo receptor, o ciclo de informar foi concluído. Devemos notar que Shannon era um matemático, estudioso de telecomunicações, que publicou sua obra singular logo após a Segunda Guerra Mundial, momento no qual os esforços na computação militar do lado aliado marcaram significativamente os rumos da história e da tecnologia da informação. Podemos também apontar que o termo *informação* é mais antigo que Shannon ou a era industrial, entretanto, com significados bem distintos. De acordo com Logan (2012b), o termo *informação* foi introduzido na terminologia anglófona no século XIV pelos franceses, que derivaram a palavra *inform*, que significava transmitir a forma de uma mente à outra, complementando-a com a terminação *ation*, que denota uma ação contínua. Assim, o termo *information* era usado na época para caracterizar um treinamento, ou uma mudança de opinião, por meio da exposição de um novo fato. No século XVII, o termo *informação* começou a ser usado com o sentido de fonte de conhecimento, pois, quando informado, o ouvinte conseguiria obter uma capacidade mais ampla de tomada de decisão. Essa nova definição pode ser atribuída ao amadurecimento do mercantilismo e ao desenvolvimento das colônias europeias nas Américas e na África. Já no século XVII, ainda de acordo com Logan (2012b), não ocorreram alterações significativas no significado do termo *informação*. Foi apenas com o

advento da abstração matemática da informação no século XX, aliada ao amadurecimento das tecnologias de telecomunicação e computação, e, finalmente, com a obra seminal de Shannon, que a Teoria da Informação teve sua primeira forma consolidada.

Shannon (1948) diz que a informação possui uma relação estatística entre os atos de ser emitida e de ser recebida, pois, em um universo cheio de variáveis físicas entre emissor e receptor da mensagem, é pressuposta a presença de "ruídos" na comunicação, que afetarão a recepção da mensagem. Em dois extremos, num caso, se apenas um trecho da mensagem for compreendido, alguma informação foi recebida; noutro caso, numa mensagem que foi recebida por completo, o meio físico certamente atribuiu ruídos à comunicação, sendo assim possível afirmar que nenhuma transmissão de informação seria perfeita nas telecomunicações, o que dá à informação uma característica efêmera. Para reduzir os impactos do ruído na transmissão de informação, Shannon percebeu que a mensagem a ser enviada poderia ser codificada de maneira simbólica, o que facilitaria sua transmissão entre máquinas. No caso, o sistema binário poderia ser usado para encapsular a informação em pacotes menores do que frases ou palavras. Quanto menores os pacotes, mais fácil de se perceber as perdas por ruído e melhor a identificação de maneiras para a recuperação da integridade da mensagem. O foco de Shannon seria, então, identificar modos de atenuar ruídos na transmissão e na recepção de mensagens entre dispositivos de telecomunicações, em que nem sempre o ser humano está presente. Consequentemente, o significado da mensagem não é relevante para Shannon, que afirma, em tradução livre (Shannon, 1948):

> O problema fundamental da comunicação é a reprodução exata ou aproximada de uma mensagem selecionada de um ponto a outro. Frequentemente as mensagens têm significado, isso significa que elas se referem ou são correlacionadas com algum sistema, de acordo com certas entidades físicas ou conceituais. Esses aspectos semânticos da comunicação são irrelevantes para o problema de engenharia.

Esta definição não agradou a todos na comunidade científica, como aponta Logan (2012b). Em meio às discussões na *Macy Conference*, evento que deu origem aos estudos da cibernética, debatia-se sobre a definição de Shannon.

Donald MacKay defendia não ser possível simplesmente descartar o valor semântico das mensagens. O problema da argumentação de MacKay é que não se poderia quantificar matematicamente semântica e relevância na transmissão da informação, o que dificultava o aceite de sua premissa pela comunidade científica da época. Ironicamente, Shannon, considerado o pai da Teoria da Informação, afirmou que sua teoria não respondia a todas as questões referentes à informação, direcionando a compreensão de seus estudos para questões mais técnicas na engenharia das telecomunicações. Por outro lado, MacKay (1969), anos depois, introduziu mais elementos para conceituar informação. Para ele, a informação possui dois campos de propagação: o campo físico, por exemplo, o caminho percorrido pelas ondas sonoras que viajam a partir de nossas laringes para o ar e em seguida para nossos ouvidos; e o campo dos significados, contidos em um objeto como um livro ou propagado pelo telefone. No campo dos significados, o cérebro deve reconhecer o valor semântico dos termos que recebe e, para que a informação faça sentido, a mensagem necessita possuir coesão dentro de um sistema de linguagem compartilhado com o receptor. Após isso, a mensagem terá de passar por um crivo de significados preexistentes de modo que, quando apresentado um novo conceito, este possa se encaixar dentro de um paradigma. Portanto, a compreensão da mensagem depende do seu contexto.

A obra de Shannon deu origem a intensos debates sobre o conceito de informação. O sentido do termo muda com o passar das décadas e com o desenvolvimento científico de diversas áreas do conhecimento, aumentando o teor de complexidade para se explicar algo que, para alguns, parece óbvio e simples: “o que é informação?”. Em sua obra, Logan (2012a) cita as nuances do significado de *informação* em diferentes esferas, as quais classifica em biosfera, simbolosfera, tecnosfera e econosfera. Logan defende que a informação é o tecido que firma o contexto dos seres humanos; na simbolosfera existiria a linguística e a cultura; na tecnosfera, a tecnologia e a ciência; na econosfera, a economia e a organização política; e na biosfera, a propagação da informação pela vida[1]. O autor relaciona o uso da informação nessas diferentes esferas, que possuem características

[1] LOGAN, Robert K. Que é informação?: A propagação da organização na biosfera, na simbolosfera, na tecnosfera e na econosfera. Rio de Janeiro: Editora PUC-Rio, 2012.

específicas de propagação, de linguagem e de significados. Significados estes que, por sua vez, só os humanos são capazes de decifrar com sua subjetividade, capacidade desenvolvida pelo fato de ser indispensável para a sobrevivência de uma espécie social. Com exceção da biosfera, todas as outras estão relacionadas com a subjetividade humana, aperfeiçoando-se para a sobrevivência da espécie, das sociedades e dos indivíduos em si. Assim, por sermos seres sociáveis e dotados de individualidade conceitual, a informação permite aos seres humanos sobreviverem.

Ainda conforme Logan, a informação subjetiva dos seres humanos difere da informação biológica contida nos diversos seres vivos, informação esta que define a organização e a reprodução do ser e, mesmo depois de sua morte, segue parcialmente recuperável, por análise genética de seu material remanescente, por até mesmo centenas a milhares de anos, dependendo das condições ambientais (ALLENTOFT, 2012). Os seres humanos, quando perecem, perdem a informação subjetiva que acumularam, pois esta está contida em suas mentes, construída com base no seu contexto de vida. Por conta disso, o ser humano sentiu a necessidade de registrar sua subjetividade (seja por razões religiosas, organizacionais, artísticas etc.) nos mais diversos suportes, como listou Innarelli (2015):

> Se fizermos uma retrospectiva sobre a história da humanidade e as formas encontradas para registrar sua memória, é possível ver que a informação e o conhecimento foram assegurados para as futuras gerações em função dos diferentes suportes documentais: ora a pedra, ora o osso, ora a argila, a madeira, o bambu, o couro, o tecido, o metal, o pergaminho, o papiro, o papel, o plástico, até chegar aos dias de hoje aos registros em suportes digitais. Vale lembrar que até meados do século passado, as formas de registros utilizadas eram baseadas em tecnologias analógicas, também chamadas nesta tese de convencionais.

Sobre o documento, podemos afirmar que seu objetivo é o registro da informação, transcendendo a comunicação oral para o registro documental. Bellotto (2006) também contribui para o entendimento da definição de *documento*:

> Segundo a conceituação clássica e genérica, documento é qualquer elemento gráfico, iconográfico, plástico ou fônico pelo qual o homem se expressa. É o livro, o artigo de revista ou jornal, o relatório, o processo, o dossiê, a carta, a legislação, a estampa, a tela, a escultura, a fotografia, o filme, o disco, a fita magnética, o objeto utilitário etc., enfim, tudo o que seja produzido por motivos funcionais, jurídicos, científicos, técnicos, culturais ou artísticos, pela atividade humana.

Podemos inferir, então, que o ser humano sente a necessidade de preservar seu conhecimento para as gerações futuras, fato que podemos identificar durante toda a história da civilização. Consequentemente, dentro das organizações pós-industriais, com a constante saída e entrada de colaboradores, o conhecimento adquirido e desenvolvido pelos que saem necessita ser preservado para uso por parte dos novos colaboradores, seja por razões estruturais ou estratégicas. Na administração pública, por exemplo, a informação deve ser preservada para garantir aos cidadãos os serviços de que eles necessitam, com base no que se tem documentado por organizações de censo como o IBGE ou o DIEESE. Em contextos como esse, o documento se torna um objeto informacional confiável para tomada de decisão dentro de um sistema burocrático. Ainda segundo Logan (2012a), no século XX, juntamente com o advento de máquinas mais rápidas e precisas que permitiram aumentar a produtividade industrial, também se testemunhou o aumento da produção de documentos. De acordo com Chiavenato (2003), isso se deu através do que Weber chamou de pensamento burocrático, o controle racional das atividades pela administração com a finalidade de delimitar seus resultados. Somadas a intensificação da capacidade de produção com novas formas de complexidade organizacional do século XX, as organizações de memória começaram a se desenvolver, tendo como objetivo assegurar que, numa sociedade em que a produção informacional nas organizações possui uma execução racional, também a preservação dessa informação seja feita de forma racional, ou seja, com metodologia bem definida cientificamente e adequadamente aplicada. Após duas guerras mundiais, podemos dizer que as organizações de memória das nações também tiveram estímulos para consolidar práticas de preservação, valendo-se de consórcios internacionais de cooperação por terem vivenciado perdas ou saques de objetos culturais ou informacionais durante os conflitos.

CAPÍTULO 2 - O PAPEL DA INFORMAÇÃO DIGITAL NO SÉCULO XXI

O termo *documento*, de acordo com o Dicionário Brasileiro de Terminologia (Arquivo Nacional, 2005), refere-se à “Unidade de registro de informações, qualquer que seja o suporte ou formato”, e sua forma digital se popularizou de fato no século XXI, pela sua facilitada manipulação e difusão, como uma forma evolucionária dos processos de trabalho nas organizações oriundas do século XX, como define o Conselho Nacional de Arquivos (CONARQ, 2005):

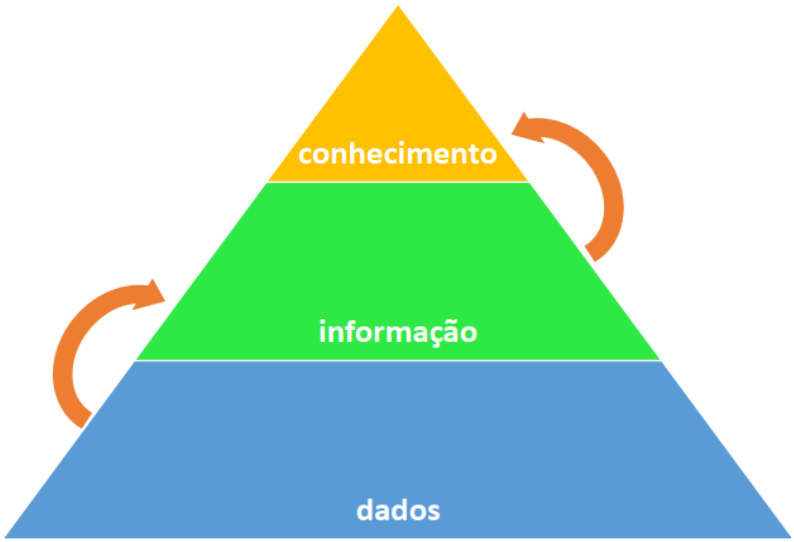
> As facilidades proporcionadas pelos meios e tecnologias digitais de processamento, transmissão e armazenamento de informações reduziram custos e aumentaram a eficácia dos processos de criação, troca e difusão da informação arquivística.
> O início do século XXI apresenta um mundo fortemente dependente do documento arquivístico digital como um meio para registrar as funções e atividades de indivíduos, organizações e governos.

A informação em meio digital possui características próprias, ensejando novas formas de se trabalhar com sua forma atomizada, o dado. Cunha e Cavalcanti (2008) definem *dado* como:

> A menor representação convencional e fundamental de uma informação (fato, noção, objeto, nome próprio, número, estatística, etc.) sob forma analógica ou digital passível de ser submetida a processamento manual ou automático.

Dados, portanto, são fragmentos não estruturados que, em conjunto, formam a informação. Eles são a única forma, até então, de a informação ser manipulável pela computação, devido ao seu valor quantificável e abstrato, já que computadores são primariamente calculadoras de alto desempenho e não possuem a capacidade de manipular a informação por meio semântico ou conceitual, apenas de maneira lógico-matemática. O avanço da computação permitiu que cada vez mais dados fossem processados, aumentando a complexidade de conhecimento que se pode obter a partir deles. Resumindo, dados com contexto se tornam informação e informação com significado se torna conhecimento (LITERIS, 2015).

Figura 1 – Pirâmide da informação



Fonte: Descomplicando a Gestão do Conhecimento – Literis[2].

Castells (2010) recupera a história do uso dos dispositivos digitais, relacionando a popularização do microcomputador pela Apple em 1975 com o desenvolvimento dos protocolos de rede pela ARPA (Agência de Projetos de Pesquisa Avançada do Departamento de Defesa Norte-Americano), que dariam os primeiros e revolucionários passos nas tecnologias necessárias para o desenvolvimento do computador pessoal e da internet. Devemos lembrar que a ARPA, no contexto da Guerra Fria, desenvolveu as tecnologias de base para a construção da rede mundial de computadores, então com o objetivo de resguardar as comunicações e informações estratégicas em caso de ataques da União Soviética. É curioso notar que, duas décadas depois, com a vinda da *World Wide Web*, seus precursores civis defendiam a disseminação da informação pelo globo como uma forma de estreitar as relações entre diferentes pessoas e culturas.

Voltando à década de 1970, com o aparecimento dos microcomputadores combinado ao avanço do desenvolvimento de sistemas operacionais pela Apple e pela Microsoft, os microcomputadores pouco a pouco foram sendo incorporados às atividades laborais comuns, tanto de pequena quanto de grande complexidade, substituindo materiais analógicos ou outros computadores de modelos menos

[2] Para mais informações: <https://literis.com.br/blog/descomplicando-a-gestao-do-conhecimento/> Acesso em: 19 jan 2020.

amigáveis ou muito mais dispendiosos, conquistando a preferência dos usuários, segundo Santos e Flores (2015). Consequentemente, reduziram-se gastos com armazenamento físico, porém muitos documentos físicos ainda se mantinham nos processos de trabalho. A digitalização de documentos surgiu, anos depois, como uma forma de trazer os documentos analógicos para a realidade digital, flexibilizando a difusão e o acesso aos documentos dentro das instituições. Com o tempo, desenvolveram-se redes conectando espaços geográficos cada vez maiores, até abrangerem todo o globo, por volta do final do século XX. Segundo a análise de Castells, as nações do primeiro mundo e dos países emergentes investem cada vez mais em suas redes de infraestrutura de telecomunicações, pois a difusão da informação se tornou vital para a soberania dos países. A informação como uma necessidade estratégica se soma aos esforços da administração pública em prover infraestrutura e regulamentações para alimentar as necessidades de mercado do sistema capitalista.

A informação em formato digital pouco a pouco se expandiu dos centros de pesquisa para as organizações privadas e então para os lares das pessoas, alterando não só os processos nas organizações, mas também a forma com que as pessoas lidam com a informação em seu dia a dia. O uso das tecnologias digitais fora do âmbito de trabalho impulsionou as TICs, fazendo com que a relação entre pessoas e sistemas computacionais não fosse mais apenas uma relação de produtor de conteúdo com uma ferramenta, mas também de difusão dessa produção, aumentando as formas possíveis de expressão.

No começo da primeira década dos anos 2000, essa produção informacional em meio digital se tornou muito mais complexa e aderente aos costumes, alterando significativamente as formas de apropriação da informação. As redes digitais tornaram-se espaço para expressar e compartilhar subjetividade, fazendo com que a produção de conteúdo digital não fosse mais exclusividade das organizações, e sim passasse a fazer parte dos comportamentos sociais. A produção de conteúdo em meio digital pela sociedade em seu tempo de lazer também fortalece o gosto pelo uso de documentos digitais nas organizações, tornando a preservação digital nas organizações uma prática de suma importância para preservação da informação e do conhecimento contemporâneos.

CAPÍTULO 3 - EM QUAIS ORGANIZAÇÕES A PRESERVAÇÃO DIGITAL SE FAZ NECESSÁRIA?

Como aponta Ferreira (2018), a produção de conteúdo digital dentro das organizações, sejam elas públicas ou privadas, necessita de políticas de preservação digital para constituir uma memória, que pode ser nacional ou local, ou então, no caso das empresas, relacionada a seus processos, tecnologias e estratégias. Nota-se, no entanto, a ausência de políticas de preservação digital na maioria das instituições, visto requererem a compreensão abrangente dos processos da instituição para definir claramente o que preservar. Tal compreensão exige um nível alto de comprometimento, investimento e domínio técnico para sua devida execução. A complexidade da criação de políticas de preservação se torna ainda maior quando o objetivo é a preservação da memória coletiva, da cultura de uma sociedade, em que não só as pessoas estão produzindo conteúdo em quantidades colossais e em velocidade inédita, mas também em plataformas privadas a que as instituições de memória não têm acesso (e, mesmo se o tivessem, enfrentariam dificuldades legais e técnicas para manipular os dados das pessoas).

Há visões diversas sobre a revolução da informação digital no cotidiano da sociedade do século XXI. Segundo Arellano (2004), durante os últimos anos do século XX, apenas as bibliotecas, os arquivos, os centros de pesquisa e os organismos governamentais teriam produzido conteúdo digital relevante. Castells (2010), por outro lado, atribui relevância não apenas ao documento digital respaldado pelas instituições, mas também ao produzido por anônimos na *web* e para a *web*. Desse ponto de vista, uma página na *web* de décadas atrás ou um código fonte primevo de algum serviço de internet poderiam ser considerados documentos digitais relevantes, por sua relevância histórica.

Ao abordarmos a questão da relevância em documentos digitais, a primeira pergunta que precisa ser feita é se a organização responsável pelo documento, seja pública ou privada, tem interesse em preservar o conhecimento sobre o qual tem domínio, seja por razões administrativas, estratégicas, históricas etc. De todo modo, faz-se necessária uma conscientização da importância da aplicação de

políticas de preservação digital. Barros, Castro e Arellano (2018), por exemplo, defendem que, se informações em meio digital podem ser recuperadas e acessadas futuramente para a geração de qualquer conhecimento, estas devem ser preservadas. Entretanto, é um exercício complexo determinar o que torna uma classe de documentos relevantes ou não, pois cada organização terá necessidades e especificidades diversas. Além disso, visto que nem todos os documentos digitais são de fácil atribuição de relevância, a definição do que preservar e o processo para tal podem requerer o olhar técnico de equipes interdisciplinares (compostas, entre outros, de bibliotecários, arquivistas, historiadores, e técnicos de TI).

Ainda que cada política de preservação deva seguir uma trilha de desenvolvimento diferente, de acordo com as especificidades de cada organização, podemos identificar certos fatores chave na composição de políticas de preservação digital mais pertinentes às organizações públicas e outras mais pertinentes às organizações privadas. A democratização do acesso à informação como uma ideologia defendida no pós-guerra criou a necessidade da produção de metodologias de preservação da informação pois, para que a informação seja democraticamente acessível, ela necessita ser acessível para os usuários contemporâneos e também futuros. Tendo como pressuposto a perspectiva humanitária da preservação da informação como um direito universal, encabeçada pela ONU no século XX, a UNESCO / NLA (2003) consolidou um guia que orienta como e por que produzir políticas de preservação digital, intitulado, em tradução livre, “Diretrizes de Preservação do Patrimônio Digital”. Também em tradução livre, este afirma:

> A herança digital é composta por materiais com suporte baseados em computador, de valor duradouro que devem ser mantidos para as gerações futuras. O patrimônio digital emana de diferentes comunidades, indústrias, setores e regiões. Nem todos os materiais digitais são de valor duradouro, apenas aqueles que exigem abordagens de preservação ativa para manter a continuidade do patrimônio digital.

Isto considerado, seria correto afirmar que nem toda produção digital é de fato relevante, pois o objetivo principal da preservação é tornar a informação acessível para as futuras gerações. Por exemplo, as instituições públicas produzem informações que podem se tornar dados estatísticos para uso futuro na

administração pública ou então podem conter a memória de diferentes grupos que compõem uma nação. No Brasil, ambos os casos são amparados pelas leis de acesso à informação[3]. Além disso, a produção científica apoiada pela administração pública nos centros de pesquisa e nas universidades públicas também deve ser preservada, por razões estratégicas, pois são de grande valia para o desenvolvimento de uma nação. Analogamente, é possível afirmar que os mesmos objetivos de políticas de preservação digital nas organizações públicas podem encontrar espaço nas organizações privadas. Ainda que, dentro de uma realidade capitalista, a competitividade dos mercados também se aplique às nações, as organizações privadas veem muito menos necessidade em preservar sua memória, já que esta supostamente não interfere diretamente nos negócios e nos lucros que visam obter. Salvo algumas exceções, podemos afirmar que a grande maioria das políticas de preservação, quando existentes, visa ao conhecimento estratégico, que pode abranger produções científicas ou tecnológicas, documentos registrando processos administrativos, contábeis, financeiros, jurídicos, publicações internas, memorandos, comunicação via e-mail, etc. A produção intelectual para fins de tomada de decisão e a manutenção da competitividade nas organizações privadas são os principais objetivos que determinam a criação de políticas de preservação nestes ambientes. É possível afirmar que, via de regra, o nível de complexidade de uma organização é proporcional à sua produção informacional. Se levarmos em consideração que, nos dias de hoje, todos os processos de trabalho, em algum momento de sua atividade, envolvem alguma produção em formato digital, podemos inferir que a necessidade de preservação digital de uma organização seja também diretamente proporcional à sua complexidade. Em sua maioria, organizações não estão instrumentalizadas com boas políticas de preservação digital, e é muito comum que profissionais leigos no assunto confundam políticas de preservação com políticas de *backup.* Da mesma forma, para o senso comum, o profissional de tecnologia da informação seria o único a articular soluções para

[3]BRASIL. Lei nº 12.527 de 18 de novembro de 2011. Regula o acesso a informações previsto no inciso XXXIII do art. 5o, no inciso II do § 3o do art. 37 e no § 2o do art. 216 da Constituição Federal; altera a Lei no 8.112, de 11 de dezembro de 1990; revoga a Lei no 11.111, de 5 de maio de 2005, e dispositivos da Lei no 8.159, de 8 de janeiro de 1991; e dá outras providências. Diário Oficial da União, Brasília, DF, 18 nov. 2011a.

as necessidades que estamos discutindo. Isto decorre, entre outros fatores, de como a sociedade enxerga os profissionais da informação. Ainda assim, é necessário reconhecer as deficiências de cunho tecnológico na formação dos profissionais bibliotecários e arquivistas, consolidando assim um hiato profissional no que tange à preservação efetiva dos documentos digitais e à forma com que a informação é encarada e tratada. Uma situação hipotética ideal seria o profissional de tecnologia da informação ser responsável pela aplicação e manutenção da técnica por trás da política, enquanto o profissional da informação seria o arquiteto da política, definindo o desenho técnico da solução e como esta seria aplicada. Arellano (2004), ao mesmo tempo que destaca a necessidade da implementação imediata de políticas de preservação digital, também aponta para a deficiência de profissionais capacitados para tal implementação:

> Alguns estudos sobre a preservação digital têm estabelecido que a imediata implementação de políticas de preservação digital é a forma mais efetiva de garantir o armazenamento e uso de recursos de informação por longos períodos de tempo. A falta dessas políticas nos projetos de repositórios digitais sugere a carência de conhecimentos técnicos sobre a importância das estratégias de preservação digital existentes. Essa lacuna informacional por parte dos responsáveis pelas políticas de implementação de informação digital precisa ser destacada.

Assim, os estudos sobre preservação apontam ser de suma importância a aplicação de políticas de preservação digital em quaisquer que sejam as organizações. Entretanto, se tomarmos como referência as orientações do documento da UNESCO / NLA, para o qual o intuito da preservação digital é a preservação da herança, entendida como preservação da memória coletiva, podemos inferir que são as instituições públicas que, prioritariamente, devem possuir políticas de preservação digital, já que estas são de importância coletiva e afetam todos os cidadãos, que nela depositam seus impostos e confiança. A realidade, no entanto, mostra que, quando existentes, as políticas de preservação das instituições públicas não são devidamente aplicadas, ou, quando o são, não são devidamente explicadas ao público. No universo das organizações privadas, é compreensível o equívoco do uso facultativo da preservação digital nos ambientes organizacionais, já que, apesar de os documentos preservados serem úteis para tomada de decisão e produção de conhecimento em longo prazo, é um tanto difícil

aplicar políticas de preservação digital quando as necessidades de se preservar documentos são mensuradas tendo como critério apenas a obtenção de lucro. A um leigo pode parecer que, no curto prazo, não compensariam os altos investimento em uma solução visando ao acesso e à preservação dos documentos digitais no longo prazo.

Outro campo em que a importância da preservação digital deve ser considerada, tanto nas organizações públicas ou privadas, é o campo jurídico, visto que, segundo Luz (2018), fatores como metadados, controles de acesso e cadeia de custódia aumentam a confiabilidade de um documento digital, que pode ser considerado íntegro e usado como evidência documentada de algum fato. A importância da preservação digital no âmbito legal vem tomando proporções maiores na década da publicação deste trabalho, no Brasil particularmente com a futura entrada da Lei de Proteção de Dados, a LGPD[4], segundo a qual, para que um documento digital seja considerado confiável, as tratativas de preservação devem ser colocadas sob análise em auditorias. Este cenário pode abrir novos mercados para profissionais da informação, visto que, como a LGPD é destinada à proteção legal dos dados pessoais dos cidadãos brasileiros e a regulamentação da posse desses dados é feita por instituições privadas, a legitimidade do documento digital deve se tornar uma questão para todas as organizações, acompanhando a tendência mundial emergente. Portanto, pode-se afirmar que quanto mais complexa uma organização é, seja pública ou privada, maior é a relevância da presença de políticas de preservação bem definidas, salientando-se a importância de as instituições públicas as manterem, já que estas prestam serviços essenciais à nação. A relevância do documento digital nos processos das organizações já é amplamente identificada, porém seu controle de autenticidade será um problema para as organizações que não implementarem políticas de preservação digital em médio prazo.

[4] BRASIL. Lei Nº 13.709, de 14 de agosto de 2018. Lei Geral de Proteção de Dados Pessoais (LGPD). **Diário Oficial da União**, Brasília, DF, 15 ago. 2018.

CAPÍTULO 4 - POLÍTICAS E ESTRATÉGIAS DE PRESERVAÇÃO DIGITAL

Como forma de ilustrar a necessidade de se criar políticas de preservação digital, Silva e Flores (2018) fizeram um estudo de caso em instituições públicas de ensino superior, por meio do qual constataram que 65% das universidades federais não possuem políticas de preservação digital e nem uma equipe de desenvolvimento para esse fim. Isto demonstraria a vulnerabilidade de nossas redes de ensino mais relevantes para o desenvolvimento científico e o risco a que nossas gerações futuras estão expostas de não terem acesso a informação crítica para tal desenvolvimento no longo prazo.

Como identificaram Silva Júnior e Mota (2012) em seu trabalho sobre a implementação de políticas de preservação digital, são poucas as instituições brasileiras que aplicam alguma política de preservação digital e, mesmo quando estas possuem políticas, não fica claro se elas estão de fato promovendo a preservação. Innarelli (2016) cita, de acordo com sua expertise, o que seriam os dez mandamentos da preservação digital:

> Os dez mandamentos da preservação digital apresentam uma breve contextualização da preservação digital e de seus dez mandamentos, tomando como base os três elementos do documento digital: o hardware (elemento físico); o software (elemento lógico) e; o suporte (elemento de armazenamento). Esta contextualização permite a apresentação de cada um dos dez mandamentos da preservação digital, os quais indicam os fundamentos básicos que poderão apoiar a elaboração de uma política de preservação digital. Os mandamentos são: **I -** manterás uma política de preservação; **II -** não dependerás de hardware específico; **III -** não dependerás de software específico; **IV -** não confiarás em sistemas gerenciadores como única forma de acesso ao documento digital; **V -** migrarás seus documentos de suporte e formato periodicamente; **VI -** replicarás os documentos em locais fisicamente separados; **VII -** não confiarás cegamente no suporte de armazenamento; **VIII -** não deixarás de fazer backup e cópias de segurança; **IX -** não preservarás lixo digital; **X -** garantirás a autenticidade dos documentos arquivísticos digitais.

Neste trabalho, propusemo-nos a tomar os mandamentos apresentados por Innarelli como pontos de reflexão, por serem bastante assertivos em definir premissas básicas para o desenvolvimento de políticas de preservação digital. Tais mandamentos colaboraram para a elaboração dos capítulos a seguir, bem como os demais autores citados no decorrer dos mesmos, para enriquecimento dos tópicos abordados.

4.1. POLÍTICA DE PRESERVAÇÃO DIGITAL

Para que práticas de preservação digital não estejam fadadas ao fracasso logo de início, para que sejam sustentáveis em longo prazo, é mister que sejam definidas políticas que as orientem, com um escopo muito bem determinado. Para tal, deve-se ter ciência dos altos gastos com infraestrutura tecnológica e capacitação profissional do pessoal técnico envolvido diretamente com a preservação e da necessidade de conscientização de todos os colaboradores quanto à importância da preservação e o papel de cada um nesse processo. Na prática, o poder aquisitivo de dada instituição, somado ao empenho da gestão e de seu corpo técnico, são muito relevantes para que uma política de preservação de fato cumpra o seu papel de preservar o conhecimento em meio digital. A crescente necessidade de se averiguar a autenticidade e a confiabilidade do documento digital também deve ser levada em conta ao se definir o que deve ser preservado ou não, e de maneira muito criteriosa. No tocante à preservação da memória, como define Arellano (2004), quanto mais rápido se aplicarem as políticas de preservação digital, melhor isso será para a preservação da informação em meio digital para as gerações futuras. Portanto, tal aspecto não se refere somente à manutenção de uma política de preservação digital em longo prazo, mas também à criação dessa política, quando ela ainda inexiste.

De acordo com Bullock (1999), certos requisitos devem ser cumpridos na criação de uma política de preservação digital. São eles:

> Fixar os limites do objeto a ser preservado; preservar a presença física; preservar o conteúdo; preservar a apresentação; preservar a funcionalidade; preservar a autenticidade; localizar e rastrear o objeto digital; preservar a proveniência; preservar o contexto. (tradução por Arellano, 2004)

Entende-se, com essa lista de requisitos, que, para a preservação de um documento digital ser efetiva, não basta se preocupar apenas com a preservação de seu conteúdo; também é necessário preservar a forma como tal conteúdo é apresentado, assegurar sua autenticidade e também seu contexto e origem. Estendendo mais um pouco as definições de Arellano sobre o assunto, compreende-se como contexto não somente o que está presente no conteúdo do

documento mas a tecnologia em que ele foi produzido, e quais tecnologias podem ser usadas para que seja interpretado. Tais detalhes precisam estar inseridos na política de preservação digital, como afirmam Schäfer e Constante (2012):

> preservar o contexto tecnológico utilizado originalmente na criação e uso dos objetos digitais que visam ser preservados. Significa que o hardware e software necessários para o acesso e apresentação do objeto digital devem ser conservados, recebendo manutenção constante.

Isso significa que, além do *software* de leitura do arquivo, quaisquer *softwares* que sejam usados para conversão, encapsulamento, verificação de autenticidade, gestão de metadados, entre outros, precisam também estar amparados pela política de preservação. Isso se estende aos *softwares* que interagem com versões convertidas dos documentos para formatos de arquivos *open-source,* ou soluções de emulação, que muitas vezes dependem de versões específicas de programas auxiliares para seu funcionamento correto. Porém, de acordo com Flores e Santos (2014), a migração para outros formatos pode alterar o contexto ou a interatividade do documento digital, sendo a emulação a forma mais segura de manter o contexto de dado documento digital. Além disso, a política de preservação necessita gerir e preservar não apenas o *software*, mas sim a versão específica do *software* que está envolvida no contexto daquele documento, e mesmo suas dependências de *software*, se houver.

4.2. O *HARDWARE* É A PEDRA FUNDAMENTAL

O *hardware* é a pedra fundamental de uma política de preservação digital. Novamente, os documentos digitais necessitam de *softwares* para sua leitura; estes, por sua vez, necessitam de um sistema operacional compatível, e este, por fim, dependente de um *hardware* compatível. Devem ser selecionadas configurações de *hardware* que não sejam de fabricação exclusiva, evitando-se, por exemplo, os produtos de marca, protegidos por propriedade intelectual, pois com o tempo seu suporte será extinto. Nos casos de dependência de *hardware* proprietário, quando ocorrer um problema técnico em algum dispositivo obsoleto, não será possível encontrar peças de troca diretamente com os fabricantes, o que

implicará obter peças e manutenção no mercado paralelo ou então depender da substituição periódica dos suportes indicados pelos próprios fornecedores do *hardware* proprietário. Isto torna a política de preservação digital em torno dos suportes proprietários muito mais dispendiosa e dependente de fatores externos e de mercado. É muito mais interessante a obtenção de soluções de *hardware* que contenham uma arquitetura de fabricação livre e acreditada para a montagem de servidores (repositórios digitais), para as rotinas de *backup* e para a verificação de integridade dos documentos. O uso de TICs ou dispositivos de *hardware* proprietários é interessante apenas para a preservação da usabilidade e experiência da leitura; por exemplo, a leitura de dado *e-book* feito exclusivamente para dada plataforma preserva a experiência pensada para o usuário. É claro que, no caso de ser interessante à instituição preservar também a experiência de usuário, o escopo da política de preservação digital deve abranger essa especificidade.

Porém, com o tempo, todo *hardware* irá se degradar, e a experiência do acesso a determinado documento via seu dispositivo digital de origem poderá então ser preservada via emulação de suas especificidades. De acordo com Ferreira (2006):

> As estratégias de emulação baseiam-se na utilização de software especial, designado vulgarmente por Emulador, que é capaz de reproduzir o comportamento de uma plataforma de hardware e/ou software, inicialmente incompatível (Rothenberg, Commission on Preservation and Access, & Council on Library and Information Resources, 1999). A grande vantagem desta abordagem está na capacidade de preservar, com um elevado grau de fidelidade, as características e as funcionalidades da representação digital pois esta será manipulada recorrendo ao software originalmente utilizado na sua criação ou reprodução (Lee et al., 2002).

De acordo com Beagrie e Greenstein (1998), algumas questões devem ser levadas em conta na manutenção de servidores de repositórios digitais. São elas, em tradução livre:

> … verificações periódicas da integridade, funções e consistência dos recursos; atualização do meio de armazenamento e cópia do recurso para superar qualquer instabilidade no meio de armazenamento ao longo do tempo; migração do recurso para novas mídias de armazenamento ou para novos formatos de arquivo; fornecimento de cópias de contingência e armazenamento em vários locais para proteger contra danos ou perdas; manter uma cópia do recurso em seu formato principal antes de qualquer migração para verificação e validação futuras e, se necessário, recuperação de dados.

Tendo em vista essas questões, podemos dizer que é necessário determinar políticas de avaliação periódica de atualização de *hardware* que abarque as especificidades dos suportes abrangidos pela política de preservação. Quando o assunto é preservar os suportes de *hardware*, quanto mais específico o *hardware,* maiores serão os custos para mantê-lo, o que, em longo prazo, torna-se inviável pela obsolescência programada da indústria tecnológica. Deve-se também levar em consideração que fazer emulação de *hardware* proprietário pode ser caro, pois exigirá maior expertise técnica e pagamento de *royalties*. Se o objetivo da política de preservação proposta é preservar a experiência dos usuários para determinados dispositivos, esse fator deve ser levado em consideração para tomada de decisão. Barros, Castro e Arellano (2018), sintetizando as ideias de Ferreira (2006), ressaltam a importância da preservação do objeto físico, lógico e conceitual do documento digital:

> O objeto digital pode surgir de duas formas: ser criado em ambiente digital ou ser oriundos da digitalização. Em qualquer de suas acepções o objeto digital se torna fator importante para a preservação digital. Sendo assim, Ferreira (2006) ressalta que um objeto digital possui diferentes níveis de abstração – objeto físico, objeto lógico e objeto conceitual. O nível físico significa que o objeto digital tem seu início a partir de um suporte físico contendo um conjunto de símbolos ou sinais inscritos. Dessa forma o hardware transforma os símbolos inscritos no suporte físico em dados legíveis pelos softwares. O nível lógico é justamente a interação que existe entre hardware e software. Por fim, o nível conceitual é caracterizado pela imagem que é formada na mente do receptor (livros, filmes, etc.)

Por outro lado, a política de preservação do *hardware* não estará somente atrelada aos dispositivos que os usuários porventura usarão. A preservação do *hardware* também implica a preservação dos servidores que abarcarão o armazenamento dos documentos digitais. Outros processos envolvidos são a preservação dos serviços de *backup*, das análises periódicas de confiabilidade, e inúmeros outros serviços e instâncias de acesso a usuários ou técnicos da preservação. É importante ressaltar que, nas melhores práticas de preservação, os servidores que os técnicos e os usuários acessam devem estar em redes separadas fisicamente por razões de segurança, pois os vírus de computador normalmente possuem o comportamento de se espalharem em rede; além disso

eventuais *hackers,* ao terem acesso a uma rede, terão mais trabalho para comprometer outra rede que, no caso, seja autônoma.

Também deve ser levado em consideração que é o *hardware* que deve guiar a escolha dos *softwares,* e não o contrário. Primeiro, porque a seleção de *software* é mais fácil quando o *hardware* já está definido. Segundo, normalmente os acervos digitais crescem mais do que diminuem, e deve ser levado em consideração que tanto a capacidade de armazenamento quanto a capacidade de processamento devem ser periodicamente analisadas e, quando for previsto que em determinado momento não darão mais conta de seus objetivos, o *hardware* deve passar por atualização. A política de preservação deve ser rígida em entender a progressão de aumento que seu acervo possui, principalmente para manter a boa saúde de seu armazenamento, antevendo atualizações antes que o sistema comece a apresentar problemas de armazenamento e coloque em estado crítico todas as outras funções. Beagrie e Greenstein (1998) mencionam certos pontos que devem ser levados em consideração na manutenção do armazenamento, em tradução livre:

> O armazenamento de dados envolverá decisões sobre a preservação a curto prazo da integridade e funcionalidade do documento, que normalmente envolverá uma combinação do seguinte:
> • verificações periódicas da integridade, função e consistência do documento;
> • atualizar a mídia de armazenamento e copiar o documento para superar qualquer instabilidade na mídia ao longo do tempo;
> • migração do documento para nova mídia de armazenamento ou para novos formatos fornecendo cópias de contingência com armazenamento em vários locais para proteger contra danos ou perdas;
> • manter uma cópia do documento em seu formato original antes de qualquer migração para futuras verificações e validações e, se necessário, recuperação de dados.

Como o autor descreve, são necessários ciclos de curto prazo de ações de manutenção dos suportes de armazenamento. Apesar de ser um texto de 1996, seus argumentos ainda são válidos em 2020. A evolução da tecnologia apenas melhorou a automatização dos processos e as capacidades de processamento e armazenamento, mas as premissas continuam as mesmas. Felizmente hoje dispomos de tecnologias de criação de *backups* redundantes. Isso significa que um servidor de *backup* conectado ao servidor principal pode controlar a autenticidade dos documentos, avisando quando estes apresentarem anomalias. Uma opção

dentre os processos de *backups* redundantes, dependendo da disponibilidade de recursos financeiros, seria dividi-los em *live storage* e *cold storage. Live storage* são os suportes de armazenamento que estão ligados ao servidor e que, como estão sendo usados constantemente, invariavelmente apresentarão falhas mais rapidamente do que os discos no *cold storage,* que são discos rígidos que ficarão guardados *off-line*. O *cold storage* é interessante para documentos que não exigem serem acessados com frequência. Ele pode conter uma cópia periódica de todo o armazenamento em *live* para fins de maior segurança, porém deve-se ter a clareza de que o *cold storage* deve também estar dentro das políticas de análise de autenticidade dos documentos.

A bibliografia diverge muito no assunto de quanto tempo leva a degradação física dos dispositivos de armazenamento digital. O CONARQ - Conselho Nacional de Arquivos (2011) - destaca que a degradação dos dispositivos digitais é o fator que mais fragiliza a preservação dos documentos digitais, pela seleção de dispositivos de *hardware* de baixa qualidade, mau acondicionamento ou manipulação descuidada. Para a seleção do *hardware,* devem ser levadas em consideração, inicialmente, as especificações do fabricante, mas, para além disso, é importante observar não só as capacidades tecnológicas do dispositivo de *hardware* mas também do suporte técnico oferecido pelo fabricante, assim como sua capacidade de responder questões sobre a vida útil dos dispositivos, o que deve permitir a adequada atualização da política de preservação.

Há também os suportes digitais não dinâmicos, como DVDs ou CDs. Sua preservação é um caso à parte, já que esses suportes já se encontram em muitos acervos em processo de degradação física, que afeta consequentemente seu conteúdo digital. Não são mais habitualmente usados, sendo muito sensíveis às variáveis ambientais, e exemplificam bem o caso de quando determinado suporte digital entra em obsolescência. É necessário realizar a migração de seus conteúdos para outros suportes mais duráveis, como *cold storages* ou servidores de preservação, nos quais os processos iniciais da inclusão de novos documentos à base são centralizados. Entretanto, deve-se considerar também a preservação do suporte original, acondicionando-o devidamente e fazendo checagens periódicas de seus conteúdos para confirmar sua integridade.

4.3. *SOFTWARES* DEVEM SER DIVERSIFICADOS E LIVRES

Devemos não somente diversificar os *softwares* usados nos processos de preservação digital, mas também nos assegurar de que o formato dos arquivos não seja dependente de um *software* específico. Consequentemente, a única forma de cumprir essa boa prática é com o uso de formatos de arquivos abertos e *softwares* de código aberto, também conhecidos como *softwares* e formatos *Open Source* ou livres.

Via de regra, *softwares Open Source* se originam de comunidades cujos colaboradores trabalham em rede, em prol do desenvolvimento e suporte de determinado *software*. O uso desse tipo de *software* é altamente recomendável para que a política de preservação digital tenha maior capacidade de preservar seus documentos em longo prazo. No caso do uso de *softwares* de código proprietário ou fechado, corre-se o risco de que determinado formato de documento perca suporte técnico e seja substituído por outro formato, por interesses comerciais dos detentores dos direitos intelectuais daqueles formatos. Isso obriga as instituições mantenedoras a depender do mercado e dos desenvolvedores do *software* para conseguir manter sua política de preservação, o que pode ser problemático especialmente para as instituições públicas, que terão gastos maiores e mais risco de não terem a devida confiabilidade de preservação digital em longo prazo. As instituições públicas muitas vezes são os alvos preferidos de ações de *marketing* das empresas de *software*, já que tais ações podem resultar em um número substancial de compras de licenças de uma só vez nos processos de licitação.

Ainda que em situações muito específicas, o *software* livre também permite sua modificação por meio da alteração de seu código fonte. Isso possibilita a criação de soluções de *software* customizadas, aumentando ainda mais a capacidade de uso dessas ferramentas, já que a própria instituição se torna mantenedora da solução que desenvolveu. No entanto, a alteração e a manutenção de uma ferramenta própria podem ser onerosas, pois requerem manter projetos de engenharia de *software* e um quadro de programadores.

A gestão dos *softwares* usados na política de preservação digital deve indicar quais devem ser usados para leitura, gestão e verificação dos documentos. A documentação deve apresentar versão e propósito dos executáveis, assim como suas dependências de *software*, que são programas ou *plugins* necessários para seu funcionamento, e, por fim, deve registrar quando tais executáveis começaram a ser usados pela instituição. Em caso de necessidade de conversão do documento para formatos livres e/ou padronizados (o que será melhor abordado no item 4.5), considerando-se que nem todos os formatos são de fácil conversão, as mesmas tratativas para gestão de *software* devem ser aplicadas aos *softwares* de conversão. Dependendo da especificidade do acervo que a instituição deseja preservar, deve-se também implementar uma política de verificação constante do suporte ao formato junto às comunidades de desenvolvedores dos *softwares*, que deve também ser documentada. A temporalidade desse tipo de verificação vai depender de um estudo mais detalhado sobre a velocidade com que os *softwares* de leitura dos documentos são descontinuados. Quanto mais específico o tipo de documento que está sendo preservado, menor a comunidade e menor o número de atualizações de *software*, o que pode levar à necessidade de conversão para um formato de documento que esteja recebendo mais suporte. Nestes casos, pode ser feito um teste amostral com as novas versões do *software* de leitura, para verificar se este ainda possui suporte e se sua comunidade de usuários e desenvolvedores continua ativa. Se ativa, a nova versão do *software* deve seguir as mesmas tratativas de *software* preconizadas pela política de preservação digital. Se não ativa, um estudo aprofundado deve identificar se o formato é passível de ser convertido para outros, sem perda de conteúdo ou de experiência de usuário, o que, em alguns casos, pode ser desafiador ou tecnicamente impossível.

Quando não for possível a migração de *software* e/ou formatos, a solução é a preservação dos *softwares* em sua última versão confiável, assim como todas as dependências de *software* necessárias para a emulação computacional daquele ponto. Devemos ter em mente que a evolução dos *softwares* acompanha a evolução dos sistemas operacionais, que por sua vez acompanham a evolução do *hardware*. Quando não houver mais formas de continuar a constante migração de *softwares* e formatos, podemos também inferir que dado formato ou *software* estará

fadado a depender de um ecossistema digital em obsolescência. No longo prazo, só poderá ser viável a preservação via emulação, que será mais satisfatória se feita enquanto os *softwares* estão mais acessíveis à memória da *web*. Deve ser destacado que os *softwares* e as dependências de emulação deverão passar pelas mesmas tratativas antes citadas para os *softwares* de leitura, testando-se periodicamente as versões da ferramenta utilizada para emulação, assim como a execução da preservação dos executáveis e suas dependências de *software*.

Em uma política de preservação de *softwares*, não só devemos nos preocupar com os *softwares* de leitura dos documentos, mas também com os *softwares* usados para verificação de autenticidade, encapsulamento, *backups*, gestão, recuperação, etc. Podemos afirmar que *softwares* da atividade fim e *softwares* da atividade meio devem estar protegidos pela política de preservação, e devem passar por testes periódicos. Isso pode parecer óbvio, mas na prática *softwares* auxiliares dentro do processo de preservação muitas vezes ficam relegados ao esquecimento, por teoricamente terem menos relevância.

Em síntese, a preservação do *software* implica a preservação de todo um ecossistema conceitual dentro do processo de uma política de preservação digital, e, em contraponto, a preservação do *hardware* é a preservação desse mesmo ecossistema, porém físico, em que ambos coexistem e dependem um do outro em uma perspectiva histórica. O próprio processo de preservação em si já é algo deveras interessante, quiçá de relevância histórica, e os *softwares* preservados podem ser alvo de estudos futuros.

Há também a questão particular dos sistemas gerenciadores que também são *softwares,* mas que possuem suas particularidades dignas de uma discussão própria sobre o assunto. Ainda podemos usar todos os argumentos supracitados para defender o não uso de sistemas gerenciadores como única forma de acesso ao acervo, mas para isso devemos compreender por que os usuários do corpo técnico de uma hipotética instituição usariam esse tipo de solução.

A realidade da maioria das instituições que, de alguma forma, possuem serviços bibliotecários ou arquivísticos é de não contar com profissionais com expertise tecnológica, como enfatizaram Arellano e Andrade (2006) ao analisar a dissertação de mestrado de Boeres (2004):

> Na dissertação de mestrado de Boeres (2004), foi verificada a carência de conhecimento por parte dos respondentes quanto ao assunto preservação digital. Esta conclusão veio das respostas obtidas por meio do instrumento de pesquisa adotado naquela dissertação, que foi o questionário, enviado a todas as universidades federais brasileiras que possuem curso de doutorado. Naquela dissertação também foi apontado um dos entraves para a preservação está na falta de pessoal preparado, ou seja, atualizado e com conhecimento técnico para levar adiante os procedimentos da preservação digital. Também, constataram-se, dificuldades quanto ao aparato tecnológico, como falta de equipamento apropriado para este fim e falta de verba para comprar tal equipamento.

Uma política de preservação digital deve, no momento de sua elaboração, definir um corpo técnico responsável por ela, que normalmente é constituído por colaboradores da própria instituição. Estes não devem necessariamente passar por longos processos de atualização técnica sobre tecnologia da informação, já que sua função na equipe não é essa. Porém, é interessante que esses profissionais compreendam ao menos a teoria por trás de cada processo. Tal treinamento não exclui a necessidade de alguns profissionais que tenham conhecimento tecnológico para executar e/ou gerir os processos que exigem maior expertise, e que compreendam tanto as questões bibliotecárias e/ou arquivísticas quanto as da tecnologia da informação. Assim se garante certa autonomia ao corpo técnico (e, consequentemente, à instituição) no processo de desenvolvimento da política que enseja constituir e manter. Os profissionais do corpo técnico que necessitam de um conhecimento mais refinado sobre tecnologia da informação precisam de recursos e disposição para se manterem constantemente atualizados sobre o assunto. Esses gastos com atualização profissional devem estar previstos no orçamento destinado à política de preservação digital, pois, como Boeres (2004) afirmou:

> Com a rapidez com que a TI se desenvolve o profissional tem que estar lendo e fazendo cursos frequentemente, de modo a não apenas acompanhar as mudanças que estão ocorrendo, mas poder tirar delas o melhor proveito profissional, respondendo rápida e pontualmente às demandas dos usuários e se destacando num mercado que está cada vez mais competitivo, onde o profissional defasado não tem lugar.

Sistemas gerenciadores podem e devem ser usados quando facilitarem o acesso aos documentos preservados da instituição por seus usuários, sejam eles do corpo técnico ou os usuários finais, mas certamente não devem ser a única forma de acesso prevista na política de preservação, pois o próprio *software* de

gestão está fadado à obsolescência tecnológica. Se esse *software* centralizar demasiados processos da política, quando não se tornar mais operacional exigirá a reestruturação e migração de toda a política de preservação da instituição, pondo em risco todo o trabalho feito durante anos. Ademais, a migração de toda uma plataforma geralmente é onerosa, trabalhosa e passível de perdas de dados ou adulteração de metadados, se os documentos não forem devidamente manipulados. Para mitigar essa situação, a política de preservação digital deve prever quais são os objetivos da plataforma de gerenciamento, tendo em vista não depender dela. Além disso, deve definir como será formado o corpo técnico que manterá tal política de preservação no nível operacional e gerencial, especificando a quantidade de pessoal com conhecimentos técnicos suficiente para a execução das atividades em que tais profissionais se fazem necessários.

4.4 - A IMPORTÂNCIA DA GESTÃO DE FORMATOS E SUPORTES

É de suma importância, em uma política de preservação digital, a gestão dos formatos que rumam para os repositórios da instituição. A política de preservação deve antecipar quais serão os formatos de documentos que receberá. Isso não implica assumir um pensamento de caráter excludente, mas sim dinamizar as tratativas. Para que isso seja feito de forma coerente, deve-se compreender a natureza de cada formato de documento digital que a política abrange, tendo em vista que, em dado momento, novos formatos aparecerão para ser preservados, os quais necessitarão de novos procedimentos. O CONARQ - Conselho Nacional de Arquivos (2011) propõe a classificação dos formatos de documentos digitais em quatro categorias:

> Os formatos de arquivo podem ser: **1. aberto**, quando as especificações são públicas (p. ex.: .xml, .html, .odf, .rtf, .txt e .png); **2. fechado**, quando as especificações não são divulgadas pelo proprietário (p. ex.: .doc); **3. proprietário**, quando as especificações são definidas por uma organização que mantém seus direitos, sendo seu uso gratuito ou não (p. ex.: .pdf, .jpeg, .doc e .gif); **4. padronizado**, quando as especificações são produzidas por um organismo de normalização, sendo os formatos abertos e não proprietários (p. ex.: .xml, .pdf/A).

As mesmas questões discutidas sobre a recomendação do uso de *softwares* livres nos processos previstos na política de preservação digital se refletem na recomendação do uso de formatos de documentos digitais abertos, pois estes possuem maior confiabilidade de acesso em longo prazo. É frequente, entretanto, que as instituições, em seus processos cotidianos, usem ferramentas que geram documentos em formato proprietário ou fechado, que necessitariam, então, ser convertidos para formatos abertos. Conforme citado anteriormente, o CONARQ - Conselho Nacional de Arquivos identifica dois tipos de documentos como abertos, sendo o segundo tipo também classificado como padronizado. Este tipo de formato de documento é importante para nossa discussão, pois, enquanto os formatos de documentos categoricamente abertos estão sujeitos às variações do suporte dado pela comunidade que mantém o *software* de leitura, os formatos padronizados são mantidos por organizações internacionais que visam à produção de normas para que dado formato digital seja o mais confiável possível visando, a princípio, a preservação da informação. É aconselhável, portanto, a conversão de documentos de quaisquer outros formatos para formatos padronizados. Entretanto, a lista de formatos sob essa classificação é pequena, se compararmos com a quantidade de diferentes formatos de documentos digitais existentes. Inclusive, na maioria dos casos, quando tais formatos são passíveis de conversão, frequentemente o são apenas para formatos abertos. A política de preservação deve definir quando a conversão se faz necessária, para qual formato, e qual ferramenta deve ser usada para tal. Esta ferramenta deve ser confiável e devidamente atualizada quando necessário; porém, seus executáveis e dependências de *software* devem ser preservados, como discutido anteriormente, visando à conversão de documentos em formatos fora de uso em algum momento futuro, o que pode ser importante para instituições de memória.

O guia do INTERPARES 2 (Duranti e Preston, 2008) recomenda que prestemos atenção em alguns detalhes quando estamos decidindo se um formato de documento pode ser seguro para fazer parte do repositório, seja ele resultado de conversão ou não. Conforme Duranti e Preston (2008, em tradução livre):

> Escolha formatos não compactados, amplamente utilizados, não proprietários, independentes de plataforma e com especificações disponíveis gratuitamente, sempre que possível. Estes são freqüentemente chamados de "formatos abertos", o que significa que suas especificações são publicadas e disponibilizadas gratuitamente. No entanto, também pode significar que o formato está isento de taxas de patentes ou royalties ou a possibilidade de tais taxas serem aplicadas no futuro e / ou que são amplamente adotadas. Deve-se notar que os formatos “abertos” não são necessariamente os mesmos produzidos pelo software de código aberto, pois este último é definido como um software o qual o código é disponibilizado gratuitamente e pode ser modificado. O software de código aberto nem sempre produz formatos não proprietários. Diferencia-se formatos de arquivo e formatos de invólucro (ou contêiner) e formatos de tagueamento, como arquivos com tags XML, e garanta que a versão, codificação e outras características sejam claras e totalmente documentadas. Para arquivos XML, verifique se os arquivos são bem formados e válidos e acompanhados pelos DTDs (Definição de Tipo de Documento) ou esquemas relevantes. Se não for conveniente seguir esta recomendação, consulte um arquivo que aceite materiais digitais e escolha entre os formatos recomendados para preservação a longo prazo. Você não deve comprimir seus documentos digitais, se possível, pois isso pode levar a problemas para a preservação a longo prazo. Se você precisar compactá-los, escolha técnicas de compactação sem perdas que estejam em conformidade com os padrões internacionais aceitos.

Para que a preservação de formatos e *softwares* em longo prazo se faça real, a conversão de formatos deve acontecer em ciclos periódicos. De tempos em tempos, discussões internas devem identificar as tendências futuras de quando determinado suporte ou formato dentro do processo de preservação virá a se tornar obsoleto. Como já mencionado nos tópicos anteriores, com um bom controle de preservação dos documentos originais e dos *softwares* que os abrem (de preferência livres), as soluções de emulação podem vir a se tornar uma boa opção para dar sobrevida a formatos ou *softwares* que estejam entrando em obsolescência.

A emulação pode ajudar a instituição a ganhar tempo para a tomada de decisões de migração, ou mesmo para evitar migrações de formato ou suporte desnecessárias. A decisão de migração deve ser muito bem definida e avaliada pelo corpo técnico, que deve refletir bem sobre quais soluções de migração preservarão os arquivos no repositório e quais soluções disponibilizarão os conteúdos para os usuários. Nem sempre a mesma solução de *software* ou fluxo de trabalho irá resolver ambas as situações, sendo talvez necessária a criação de dois fluxos de conversão de formatos. Dependendo da infraestrutura da organização e por razões de segurança, pode ser recomendável ter uma versão

dos documentos digitais exclusiva para preservação, que deve ficar no repositório, e outra para ser acessada e utilizada. Isto pode tornar o acesso pelos usuários mais rápido, já que, em casos específicos, nos formatos de uso corrente pode-se perder um pouco de resolução gráfica e ganhar desempenho na abertura do documento. Soluções de emulação exigem, entretanto, equipes com maior expertise técnica, o que não é a realidade na maioria das instituições.

Podemos citar o *site* Internet Archive (Archive.org) como exemplo de emulação, para entendermos melhor as possibilidades de aplicação dessa técnica aos fluxos de preservação digital. O *site* possui emuladores que carregam sistemas operacionais, jogos e programas de 40 anos atrás, prática que seria muito interessante para a preservação digital como um todo, o que exige uma maior pesquisa sobre sua metodologia interna de preservação. Essa organização é composta por uma extensa comunidade *on-line*, e o *site* é administrado pela ALA (American Library Association), renomada instituição bibliotecária, mostrando que há projetos robustos que mesclam a ciência bibliotecária com a tecnologia da informação. Seu acervo *on-line* possui uma grande densidade de dados: segundo informações do *site* coletadas no período de elaboração deste trabalho, eles hospedavam então mais de 45 Petabytes de dados de diversas categorias. Entretanto, como se trata de uma biblioteca *on-line*, o *site* também recebe material cultural da comunidade que o rodeia, a internet. Navegando pelos diferentes tipos de mídia, percebe-se que há uma infinidade de *uploads* de usuários com diversos temas, o que indica um fraco controle ou classificação do que é hospedado pela instituição. É compreensível que isso ocorra: nenhuma instituição daria conta de tal enxurrada de dados e, a despeito desse fato, o projeto do *site* é de grande valia para o universo bibliotecário. A análise deste capítulo nos leva a ponderar que a equipe e a instituição como um todo devem estar dispostas a criar e manter processos, às vezes complexos, para preservar os formatos de documentos digitais, se realmente estão visando sua preservação a longo prazo. Esta não é uma tarefa simples, sendo necessário que os profissionais da informação criem uma relação mais próxima com a tecnologia e também com as comunidades de desenvolvedores dos *softwares* livres utilizados na política de preservação. Tal aproximação pode ser uma forma barata de mitigar os riscos, porém exige empenho

de cada profissional, pois quanto mais demorar uma solução de migração quando esta se fizer necessária, maiores serão os custos para migrações retroativas do repositório e maior o risco de falhas no processo, sendo a migração de formatos e suportes uma das mais sensíveis.

4.5 - AMBIENTES DISTINTOS PARA MELHOR SEGURANÇA

Uma política de preservação digital não deve tentar resolver lacunas de outras políticas de uma instituição, como, por exemplo, as políticas de segurança da informação. Entretanto, ela deve abordar a infraestrutura que suportará os fluxos de trabalho e o processo de preservação em si. É importante frisar que um acervo digital possui as mesmas necessidades de segurança de qualquer outro ambiente computacional. Como todos os acervos físicos, estes ambientes não somente estão expostos a agentes de degradação física intrínsecos ao próprio material constituinte, mas também a agentes externos, como danos elétricos por sobrecargas na rede, componentes eletrônicos defeituosos em algum dispositivo, vírus de computador ou até mesmo ataques *hackers*. Existem diversas formas de se pensar uma infraestrutura física para os fins que estamos a discutir; entretanto, essa infraestrutura deve abranger determinados requisitos básicos para obtermos os menores riscos de perda de informação dentro de um acervo.

Uma importante possibilidade de contaminação de documentos nato-digitais são os vírus. Alguns deles podem eliminar acervos inteiros e, se tais acervos não possuírem *backup*, os danos informacionais podem ser irreversíveis. A capacidade de destruição informacional desses vírus é tamanha que, nos dias de hoje, já é possível encontrar vírus que não destroem conteúdos, mas os encriptam com senha – são os *ransomwares*. Os responsáveis pelo vírus, então, pedem um resgate pela descriptografia dos dados. Este exemplo mostra o quanto a informação digital se tornou importante e ao mesmo tempo vulnerável a ataques, que podem acontecer de um ponto a outro do mundo.

Devemos então entender o porquê de alocar os documentos e seus *backups* em locais fisicamente separados, como sugere Innarelli (2015). Sejam quais forem os problemas que os acervos podem enfrentar, muitos deles podem ser evitados

simplesmente se as infraestruturas de uso e de preservação não estiverem conectadas à mesma rede de energia, à mesma rede de internet, se o fluxo de trabalho acontecer em computadores separados ou desconectados da rede externa, etc. Quanto mais descentralizada for a infraestrutura que permeia os fluxos de trabalho da política de preservação do acesso aos usuários e da localização dos *backups*, mais segura ela será.

A depender da especificidade da instituição, seu acervo pode ser alimentado por documentos que circulam na rede de trabalho, que chegam em suportes físicos ou por ambos. Os fluxos de preservação desses documentos devem ser feitos em servidores ou computadores separados de outras redes, a internet ou a intranet, que não façam parte do processo de *input* de documentos no acervo. Isto cria um caminho de mão única de acesso a esses documentos no acervo. Os profissionais que irão manusear esses conteúdos devem, preferencialmente, usar computadores que estejam atualizados com os *patches* de segurança, podendo ser usadas também máquinas virtuais para o acesso desses *inputs* de documentos. Em caso de contaminação digital, tais máquinas podem ser facilmente isoladas. Elas também podem possuir rotinas de verificação ou *resets* periódicos, para sanitizar qualquer contaminação. Dentro desses ambientes serão feitos todos os tratamentos informacionais cabíveis: indexação, catalogação, gestão de metadados, conversão, documentações processuais etc. Após os devidos tratamentos, o *output* de preservação deve ser gravado primeiro em mídias físicas, como discos rígidos, que também devem passar por um esquema de indexação e ser acondicionados de maneira apropriada em ambiente controlado. Este procedimento é importante pois os discos rígidos desligados mantêm a integridade da informação por muito mais tempo do que os conectados. O acervo deve, assim, contar com cópias de segurança. Quanto menos manuseadas e melhor acondicionadas, menor será sua degradação.

É importante que haja uma política de verificação de integridade, que pode usar a tecnologia de cálculo de *Hash*. O cálculo de *Hash*, explicado de maneira simples, é uma forma de se ler os *bits* de determinado arquivo e, por meio de cálculos matemáticos, atribuir um código único para tal arquivo. Caso algo mude em seu conteúdo, sua estrutura de *bits* será alterada e, em uma próxima

verificação, o *hash* acusará seu corrompimento ou alteração indevida. Os *hashes* podem ser importantes metadados a serem registrados, para que haja sempre uma possibilidade de verificação de confiabilidade dos arquivos, o que será melhor abordado no item 4.10. Sob a ótica de se ter ambientes distintos para melhor segurança, estes ambientes devem possuir políticas bem fundamentadas para a gestão de suportes de armazenamento, usando-se de normas técnicas que serão dispostas no item seguinte.

4.5.1 - GESTÃO DE SUPORTES DE ARMAZENAMENTO

Uma política de preservação digital não deve ser construída sobre a crença de que, a partir do momento que o documento digital está guardado, ele está preservado. A preservação de fato requer checagens periódicas da qualidade dos processos e da integridade dos documentos e suportes, por via de auditoria. Para auxiliar os profissionais da informação na tomada de decisão e formação de políticas eficientes, há diversas normas técnicas estabelecidas que, se seguidas corretamente, podem tornar a preservação dos documentos digitais mais eficaz. Para preservar os conteúdos que portam, os suportes necessitam de tratativas bem fundamentadas e periodicamente executadas. Innarelli (2015), ao recuperar seu trabalho de 2007, diz o seguinte sobre a fragilidade dos suportes digitais:

> O suporte de armazenamento digital, como já observado, é considerado frágil frente às tecnologias convencionais e sua degradação acontece com rapidez, este é considerado um problema crítico, quando é tomado como base o fundamento de que a informação digital, diferentemente da analógica, pode ser perdida simplesmente por um dano de parte do suporte, seja ele mínimo. No documento — analógico é possível perceber a degradação do suporte mesmo antes da perda da documentação, enquanto no digital, esta perda se torna silenciosa, pois o documento pode estar disponível em um dia e no outro ele pode desaparecer sem que tenhamos percebido a degradação do suporte (INNARELLI, 2007, p. 60).

Considerado o apresentado, é coerente afirmar que a preservação de documentos digitais só é possível mediante a sinergia entre bons processos e boa infraestrutura. No quesito suportes, a efemeridade de suas capacidades de funcionamento em longo prazo nos põe a refletir que bons processos são

fundamentais para tornar possível a preservação digital. Por conta disso, com o auxílio da publicação da política de preservação do Arquivo Nacional (2016), citaremos algumas normas técnicas relevantes para o armazenamento e a transferência de documentos digitais e para a segurança digital, disponíveis no *site* da ABNT:

Tabela 1 - Normas ABNT e ISO relevantes para preservação digital

Organização	Norma	Descrição
NBR/ISO	NBR ISO/IEC27001	Tecnologia da Informação - Técnicas de segurança - Sistemas de gestão de segurança da informação – Requisitos.
NBR/ISO	NBR ISO/IEC27002	Tecnologia da informação - Técnicas de segurança - Código de prática para a gestão da segurança da informação (Conteúdo técnico idêntico ao da ABNT NBR ISO/IEC 17799).
NBR	NBR 15247	2004 - Unidades de armazenagem segura - Salas-cofre e cofres para hardware - Classificação e métodos de ensaio de resistência ao fogo.
NBR	NBR 15472	2007 – Modelo de referência para um sistema aberto de arquivamento de informação (SAAI).
NBR/ISO/ABNT	NBR ISO/ABNT 11515	2007 - Guia de práticas para segurança física, relativas ao armazenamento de dados.
ISO	ISO 14721	2012 - Space data and information transfer systems – Open archival information system – Reference model.
ISO	ISO 23081-1	2017 - Informação e documentação - Processos de gestão de documentos de arquivo - Metadados para documentos de arquivo. Parte 1: Princípios.
ISO	ISO 16363	2012 - Space data and information transfer systems – Audit and certification of trustworthy digital repositories.

Fonte: Elaborada pelo autor com base nos *sites* da ABNT.

Conforme Sousa (2014), a segurança da informação é imprescindível na elaboração de uma política de gestão da informação. Cada vez mais, ela é um assunto que deve ser discutido não só nas organizações privadas, mas também nas organizações públicas, que estão se tornando alvo de ataques *hackers*, tanto por conta do impacto de suas informações quanto pela informatização precária de certos processos. Sem o devido investimento em segurança e educação preventiva, os sistemas de informação da administração pública são alvos fáceis para cibercriminosos. Neste contexto, as normas ISO 27001 e 27002 - entre outras da série 27000 - são importantes para que determinada organização atenda aos devidos padrões que assegurem seu âmbito digital. Existem certificações profissionais e institucionais para esta série de normas, sendo a 27001 a que tange os requisitos básicos de segurança e a 27002 a de controle e vigilância da segurança digital da própria organização. A respeito da segurança física, há duas normas que visam proteger acervos digitais. A NBR 15247 (ABNT, 2004) determina métodos para criar condições ambientais específicas para salas cofre e cofres, além de regras para evitar danos por incêndios. Ela também descreve práticas de testes de impacto para cofres específicos para guarda de dispositivos de *hardware*. A NBR 11515 (ABNT, 2007) determina boas práticas para o armazenamento de documentos digitais em suportes, para uso operacional, *backups* e transporte, em situações comuns ou mesmo emergenciais.

As NBR 15472 e ISO 14721 tratam da criação de processos visando ao desenvolvimento de arquivos digitais abertos. Estes devem priorizar o uso de ferramentas e formatos de documentos digitais do tipo abertos, com a premissa de que, em longo prazo, essas modalidades de arquivos digitais são muito mais consistentes em preservar seus acervos, por usarem os mesmos padrões que outros muitos arquivos no mundo. Cria-se, assim, uma metodologia que visa à recuperação da informação teoricamente por qualquer outro arquivo, em qualquer época, em qualquer lugar do globo.

A ISO 23081-1 (ABNT, 2017) descreve os princípios básicos do uso de metadados para a gestão de documentos de um arquivo. As políticas de preservação, a princípio, devem possuir informações correlatas aos processos que

lidam com esses documentos, em quais sistemas eles residem e quais são as organizações responsáveis pelo conteúdo desses documentos.

Por último, a ISO 16363 (ABNT, 2012) descreve práticas de acesso confiável a documentos contidos em um repositório digital. Esta ISO é passível de certificação e é usada como base para a definição de processos de auditoria para comprovar a eficácia dos processos de digitalização.

Sobre a premissa de “Não confiar cegamente nos suportes de armazenamento”, o próximo tópico abordará a importância das cópias de segurança de forma cíclica, e a relevância que essa ciclicidade tem para a preservação digital, levando em consideração a efemeridade das tecnologias de *hardware* para armazenamento existentes até então.

4.5.2 - *BACKUP* E CÓPIAS DE SEGURANÇA NA PRESERVAÇÃO DIGITAL

Backups e cópias de segurança são centrais para a preservação digital, como afirma Fialho Jr (2007):

> As cópias de segurança em computadores são instrumentos importantes para compensar – ou tentar sanar – problemas advindos de hardware, como, por exemplo, uma pane no disco rígido, ou de software, como a invasão do sistema por hackers, através de vírus, perda acidental de arquivos, conflitos no sistema operacional etc. Por isso, a cópia de segurança é a melhor forma de prevenção e recuperação das informações, já que os dados podem voltar fielmente para o disco, quando for necessário.

Apesar de imprescindível, a aplicação de técnicas de replicação digital está além das práticas do senso comum, pois um acervo digital está sujeito às mesmas lógicas de manutenção e *backup* com que são mantidos servidores em outras situações. A esse respeito, Barros, Castro e Arellano (2018) enfatizam tal necessidade, afirmando que:

> Os esforços praticados perante as tentativas de preservação digital visam fortalecer o movimento e desenvolver novas formas de se trabalhar na temática. A atividade de preservar objetos digitais está muito além das técnicas de migração, autenticação, backup, etc. Na verdade, a temática envolve fatores macros que vão desde a conscientização até a adoção de políticas estratégicas para preservação digital.

As políticas de *backup* compreendidas pela maioria dos profissionais da área de tecnologia da informação estão firmadas conceitualmente em práticas de preservação rápida dos dados, que em muitos casos estão sendo modificados e transferidos a todo momento, não existindo uma preocupação quanto à preservação dos metadados ou de seu conteúdo em longo prazo. Recordando Castells (2010), que afirma que a computação em rede seria um produto formado por inúmeras variáveis históricas, geopolíticas e sociais, talvez não seja de todo errado afirmar que a causa da efemeridade da informação em meio digital se deve ao fato de ela ser vista como matéria-prima dentro da lógica capitalista. Assim sendo, ela se tornaria descartável depois de sua devida exploração, como qualquer outra matéria-prima. Tal concepção corroboraria a construção de todo um domínio conceitual de técnicas de *backup* que, muitas vezes, não enxergam a importância de se fazer guarda em cofres ou da preservação dos metadados dos documentos digitais.

Como os métodos de *backup* serão executados por profissionais de TI, estes devem estar cientes dos procedimentos exigidos para se fazer *backups* e cópias de segurança que preservem os metadados, protejam os documentos digitais de subscrição e impeçam acessos indevidos ou descuidados. Como reforça Innarelli (2015), os profissionais de TI não entendem a importância dos metadados no processo de preservação, o que exige que as equipes envolvidas nesse processo sejam multidisciplinares, somando os conhecimentos de profissionais da ciência da informação com os da tecnologia. Além disso, é de suma importância que o processo de geração dos discos de *backup* para guarda em ambiente de acesso restrito seja feito de maneira periódica pelos profissionais, sendo papel do arquivista arquivar devidamente os suportes. Innarelli (2015) ressalta a importância de se separar o conceito de preservação do conceito de cópia de segurança:

> Cabe ressaltar que as cópias de segurança, apesar de influenciarem diretamente na preservação digital, não representam a política de preservação digital, pois sua função é garantir a recuperação dos dados existentes nos sistemas informatizados, permitindo assim a preservação digital no presente e não a preservação para o futuro.

Uma tecnologia que pode ser útil para realizar cópias de segurança é a tecnologia RAID. RAID é um protocolo de espelhamento de conteúdo digital armazenado dentro de discos, bastante comum em políticas de *backup*. Dentre os diversos modos RAID, citaremos os dois mais relevantes, conforme Morimoto (2005).

RAID 1:

Protocolo para espelhamento. Nele, o conteúdo gravado em um disco rígido é automaticamente copiado para outro, sem interação com o usuário, exigindo apenas configurações preliminares dos profissionais de infraestrutura de TI. Dessa maneira, caso um dos discos seja removido para ser usado em outro processo ou apresente defeito, pode ser substituído por outro que receberá automaticamente os dados.

Esse modo de RAID pode ser útil quando dois discos com o mesmo conteúdo tenham que ir para processos distintos, sendo divididos em cópia de trabalho e cópia de arquivamento. No esquema abaixo, são representados dois sistemas de armazenamento, sendo que em ambos os dados são armazenados de maneira idêntica:

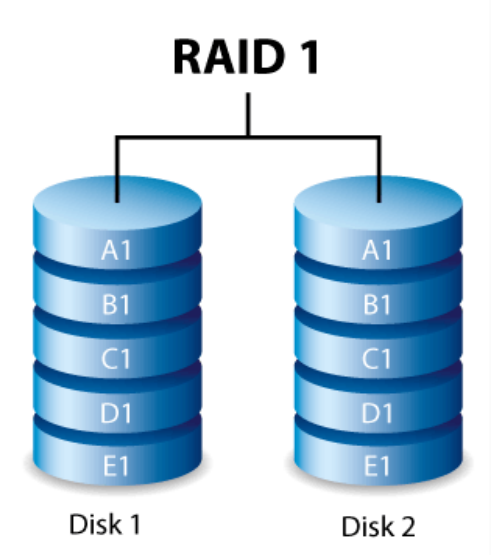


Figura 2 - Representação do protocolo RAID 1

Fonte: Manual para modos RAID - Seagate[5].

[5] Disponível em: <https://www.seagate.com/br/pt/manuals/network-storage/business-storage-nas-os/raid-modes/>. Acesso em: 19 jan 2020.

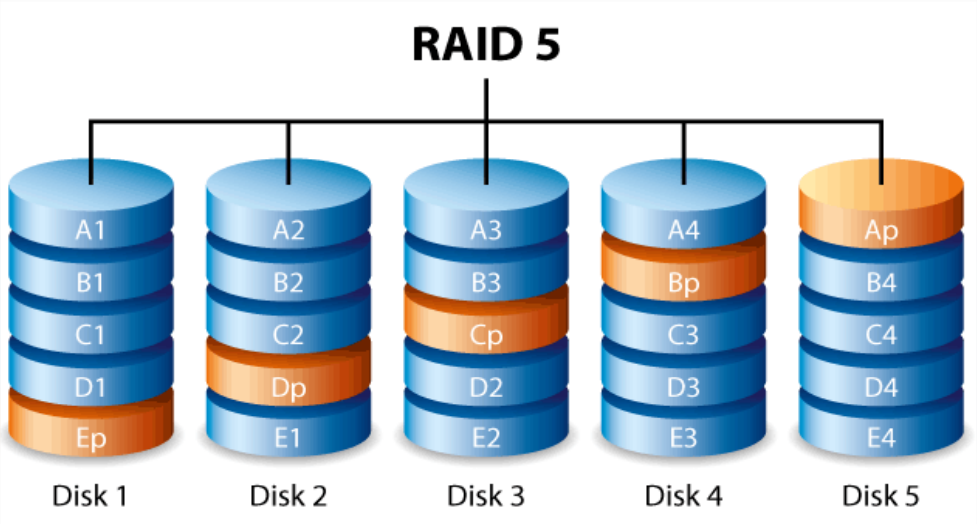
O dado, quando gravado no disco 1 pelo sistema operacional, automaticamente é gravado no disco dois pela controladora do sistema RAID. Essa ação independe do sistema operacional; portanto, não tem como ser adulterada por outros processos.

RAID 5:

Diferentemente do que ocorre no RAID 1, no RAID 5 a gravação é dividida entre no mínimo dois discos, enquanto um terceiro é usado para gravar *bits* de paridade. Este *bit* é gerado a partir de um cálculo feito com cada conjunto de 4 *bits* gravados de cada arquivo salvo em um dos dois primeiros discos. Caso haja uma degradação em algum desses conjuntos de 4 *bits*, o sistema consegue reconstruir aquele conjunto, usando o *bit* de paridade salvo no terceiro disco, e mantendo o arquivo digital integro.

O RAID 5 geralmente é usado em servidores com soma grande de discos. São necessários no mínimo três discos para seu funcionamento. O esquema abaixo apresenta 5 discos, divididos em 4 volumes de dados e um quinto de paridade. Os 4 volumes de dados, assim como o volume de paridade, ficam salvos em discos diferentes. Assim, caso haja perda de algum volume ou disco, é possível a reconstrução dos dados, já que seu volume de paridade está em outro disco.

Figura 3 – Representação do protocolo RAID 5



Fonte: Manual para modos RAID - Seagate[6] [7].

[6] Disponível em: <https://www.seagate.com/br/pt/manuals/network-storage/business-storage-nas-os/raid-modes/>. Acesso em: 19 jan. 2020.

[7] Para mais informações: <https://www.hardware.com.br/termos/raid-5>. Acesso em: 19 jan. 2020.

Este protocolo traz vantagens para o arquivamento de documentos digitais, pois como a paridade dos arquivos é controlada e mantida por uma controladora independente do sistema operacional, dá maior segurança ao processo de preservação. Existem outras técnicas de *backup*, mais conhecidas no universo da TI, que, com apoio do corpo técnico, podem ser úteis, desde que se tenha cuidado com a preservação dos metadados e que sejam feitas também rotinas de validação de conteúdo. Uma tecnologia que poderia ser usada para que *backups* de conteúdos sejam feitos com mais segurança, dentro do escopo de preservação digital, é a tecnologia de *backups* de conteúdos encapsulados. Independentemente da técnica de *backup* utilizada, esses volumes encapsulados preservam os metadados dos documentos, possibilitando a transferência segura e, consequentemente, o uso de técnicas de *backup* mais comuns. No encapsulamento, as tratativas de preservação da autenticidade são feitas antes da movimentação dos documentos, mitigando riscos de alteração.

De acordo com Durbano (2019), as técnicas de *backup* mais comuns são: completo, incremental e diferencial. No *backup* completo, são feitas cópias do sistema todo, fixado num ponto do tempo, independente de se os dados contidos no sistema sofreram qualquer alteração.

No *backup* incremental, por outro lado, identifica-se quando há novos documentos ou documentos modificados, sendo estes os únicos a serem copiados. Apesar de parecer uma boa ideia quando se pensa na relação custo-benefício, o ideal é que, quando se faz o uso de *backups* incrementais, também se façam *backups* completos. Isto porque *backups* incrementais podem corromper arquivos íntegros no *output*, por copiarem arquivos com erro na origem, substituindo-os no *backup*.

No *backup* diferencial, é feito inicialmente um *backup* completo. Define-se, então, um período de tempo que se iniciará com esse *backup* e abrangerá dois ou mais *backups* completos posteriores. Então, entre os dias que passam entre os *backups* completos, são feitos *backups* dos arquivos alterados, sempre comparados com o *backup* completo anterior. Sendo assim os *backups* intermediários não usam outros *backups* intermediários como base de comparação, somente os *backups* completos. Essa técnica é boa se pensarmos que, se tivermos

um *backup* completo íntegro, ele é a base para que o sistema identifique alterações e faça *backup* de conteúdo alterado, e também para que se tenha outra fonte de substituição. Isto pode ser interessante para identificar documentos corrompidos e tornar a tratativa desses documentos mais ágil para a identificação de erros no sistema ou na infraestrutura.

Com as técnicas de *backup* supracitadas como norte para pesquisa mais aprofundada, os profissionais da informação podem se inteirar melhor sobre *backups*, podendo assim desenvolver técnicas ou até mesmos *softwares* de *backup* que realmente valorizem as técnicas de preservação digital.

4.6 - LIXO DIGITAL E SUAS POSSIBILIDADES DE RECICLAGEM

Da mesma forma que os acervos físicos, os acervos digitais exigem práticas periódicas de averiguação para se determinar se os documentos que os compõem realmente têm relevância para serem mantidos. Por conta de os acervos digitais possuírem capacidade de armazenamento muitas vezes maior que a capacidade de produção, acumula-se muito lixo digital. E, diferentemente do que ocorre com os acervos físicos, os documentos digitais esquecidos no repositório muitas vezes não criam uma sensação de incômodo, de “estorvo”. Porém, caso não sejam tomadas medidas para mitigar a preservação de lixo digital, gastos desnecessários com *hardware* para armazenamento podem vir a ocorrer com o passar do tempo, além do esforço perdido da mão de obra das equipes de preservação com volumes de lixo digital.

Innarelli (2015) comenta sobre a relevância do uso de TTDs, Tabelas de Temporalidade de Documentos:

> Desta forma concluímos a discussão sobre as idades do ciclo vital e introduzimos a discussão de que a relação entre o ciclo vital, o valor (primário ou secundário) e a idade do documento arquivístico deve ser determinada por um instrumento fundamental para gestão e preservação dos documentos arquivísticos, a chamada Tabela de Temporalidade de Documentos (TTD), a qual estabelece entre outras coisas, os prazos de arquivamento dos documentos e sua destinação final: eliminação ou guarda permanente. A TTD é um instrumento que evoluiu ao longo do tempo, tendo como um dos seus precursores Schellenberg (2006), que a apresenta como instrumento de destinação.

O tempo de guarda citado por Innarelli pode ser renovado e, caso o documento tenha alguma relevância histórica, informativa ou probatória, vai para a guarda permanente.

Além de evitar o acúmulo de lixo digital nos acervos, a política de preservação digital deve definir de maneira inteligente quais tipos de documentos devem ser preservados, quais formatos, por quais razões e por quanto tempo, levando em consideração seu próprio contexto. Como aponta Faria Filho (2000):

> Preservar não significa guardar tudo, mas avaliar a documentação, descartando o desnecessário e criando condições mínimas de sobrevivência do suporte físico (materialidade) e da informação do documento.

Tomemos como exemplo o caso da Pinacoteca de São Paulo, que, segundo Luz e Maringele (2018), decidiu preservar todos os seus *e-mails* trocados. Alguns podem achar que se trata de uma guarda excessiva de dados no acervo, enquanto outros acreditarão se tratar de uma boa prática do ponto de vista de *compliance*. É claro que, em longo prazo, apenas razões ligadas a *compliance* não seriam suficientes para manter a guarda de *e-mails* muito antigos; portanto, cada caso precisa ser avaliado com cuidado, já que o documento digital é diversas vezes difuso, enredado em uma miríade de objetivos e formatos.

Ainda sobre formatos de documentos digitais, em uma organização podem existir documentos nato-digitais ou digitalizados. Os documentos digitalizados são cópias impressas digitalmente que, a princípio, não permitem edição, o que pode dificultar buscas pelo conteúdo do arquivo no sistema e a interatividade do usuário. Esses arquivos estão mais suscetíveis a serem considerados (por vezes equivocadamente) como lixo digital, pois, se não forem devidamente tratados com uma boa gestão de metadados, seus conteúdos não serão recuperáveis de maneira automatizada. Além da gestão de metadados, outra estratégia que pode ser útil é o uso de OCR (*Optical Character Recognition*), ou, em tradução livre, Reconhecimento Óptico de Caracteres. Cunha e Cavalcanti (2008) advertem que o reconhecimento óptico artificial não é perfeito, pois a máquina pode confundir os caracteres de acordo com a fonte usada; entretanto, os algoritmos de

reconhecimento têm se tornado cada vez mais precisos, ampliando a capacidade de se trabalhar com documentos digitalizados em meio digital.

As chamadas técnicas de Data Analytics adicionam ainda mais complexidade à discussão sobre a melhor forma de determinar se um documento (ou grupo de documentos) é de fato lixo digital. Tais técnicas analisam grandes quantidades de documentos ou dados para gerar conhecimento, oferecer *insights* ou alimentar inteligências artificiais, a partir de combinações variáveis de dados, quantificando valores que, aos olhos humanos, seriam praticamente impossíveis. A possibilidade de analisar grandes quantidades de documentos ou dados em conjunto e deles extrair inteligência torna ainda mais complexa a classificação de qualquer documento como lixo digital, indicando que o próprio conceito de “lixo digital” deve ser relativizado.

A realidade naturalmente irá variar de instituição para instituição e de acordo com o momento histórico. Não podemos afirmar que tudo poderá ser de fato útil em dado momento, ignorando as circunstâncias contextuais específicas da instituição. Esta discussão necessitaria de uma abordagem mais extensa, na qual este trabalho não irá se aprofundar. Qualquer afirmação sobre melhores práticas relacionadas ao lixo digital poderia induzir a alguma forma de sanitarismo digital que empobreceria análises de Data Analytics futuras. Isto posto, a relevância, a validade e a temporalidade do documento digital no acervo devem ser discutidas com muita cautela pelo corpo técnico de cada política de preservação digital.

4.7 - MÉTODOS PARA PRESERVAR A AUTENTICIDADE DO DOCUMENTO DIGITAL

O documento digital pode ser considerado autêntico quando se puder verificar que ele não sofreu adulterações, como nos esclarece Rondineli (2005):

> A autenticidade de um documento está diretamente ligada ao modo, à forma e ao status de transmissão desse documento, bem como às condições de sua preservação e custódia. Isso quer dizer que o conceito de autenticidade refere-se à adoção de métodos que garantam que o documento não foi adulterado após a sua criação e que, portanto, continua sendo tão fidedigno quanto era no momento em que foi criado.

Seguindo esse pressuposto, podemos apontar duas práticas principais de autenticação que se complementam. Uma delas é oriunda da prática bibliotecária e documentalista, que é a gestão e preservação dos metadados dos documentos digitais. A outra é oriunda da ciência da computação, que é o cálculo de *Hash* do documento digital. Comentaremos brevemente cada uma dessas práticas, iniciando com a preservação dos metadados.

4.7.1 - GESTÃO DE METADADOS

A gestão dos metadados de um documento digital começa no processo de entrada de tais metadados no serviço de preservação, entrada essa que deve ter uma metodologia bem definida. O tratamento dos documentos deve ser particularmente cuidadoso na preservação dos metadados, que estão sob o risco inicial advindo dos sistemas operacionais ou da manipulação descuidada, os quais podem, por exemplo, alterar a data de criação ou do último acesso do documento. A perda dessa informação pode ser problemática pois, dependendo da política de preservação da instituição, a alteração de determinados metadados pode acarretar a invalidação de dado documento.

Existem modelos de metadados que podem ser usados para a classificação e a catalogação de documentos digitais, assim como para viabilizar o registro de outras informações relevantes sobre o documento ou sobre o próprio processo de preservação. Há diversos modelos para a construção desses documentos de metadados: OAIS (Open Archival Information System), EAD (Encoded Archival Description), Dublin Core, EDM (Europeana Data Model), RDF Schemma, e muitos outros. A escolha do modelo de registro dos metadados deve levar em consideração alguns pontos: se é um modelo aberto, se possui o suporte de uma reconhecida comunidade internacional, se possui interoperabilidade, se é o mais condizente com os propósitos da instituição, etc.

O documento de registro de metadados é muito semelhante à representação descritiva de fichas catalográficas, podendo os metadados de um documento ser registrados em um outro documento digital. Este é construído de forma padronizada, sendo composto de campos tais como título, autor, ano de publicação,

mas podendo incluir também registros sobre conversão, emulação entre outros. Os documentos de metadados geralmente são salvos em formato XML ou outros formatos de linguagem de tagueamento, desde que padronizados.

Além das escolhas já citadas de modelo e formato de registro, é importante atentar para a inserção de dados. Esta deve, antes de mais nada, ser feita por pessoal capacitado não só por possuir conhecimento bibliotecário ou arquivístico de classificação e catalogação, mas também por dominar bem o uso do modelo de metadados em questão, além de estar ciente de seus objetivos com a preservação digital. Tal cuidado é necessário pois, dependendo do tamanho do acervo digital e de seus métodos de recuperação, documentos com metadados incorretos podem ser um grande desafio para serem recuperados novamente.

Um outro fator muito importante que integra a gestão de metadados é a interoperabilidade. Através desta, quando um acervo bibliográfico possuir um determinado item, e desde que ele esteja devidamente descrito num sistema interoperável, outra biblioteca pode usar os mesmos metadados já produzidos e publicados por outra instituição, processo que facilita e otimiza a classificação e catalogação de documentos em meio digital e em rede, segundo Marcondes (2016). Tal processo se assemelha ao que é feito pela organização responsável pelo DOI, que visa gerar identificadores únicos para documentos eletrônicos, gravando seus metadados e os vinculando a registros para conferência internacional, segundo Brito (2016).

A gestão dos metadados é uma questão de suma importância para a preservação digital, como define o CONARQ – Conselho Nacional de Arquivos (2005):

> A preservação da informação em formato digital não se limita ao domínio tecnológico, envolve também questões administrativas, legais, políticas, econômico-financeiras e, sobretudo, de descrição dessa informação através de estruturas de metadados que viabilizem o gerenciamento da preservação digital e o acesso no futuro.

Os gestores da política de preservação digital devem definir se determinado campo no arranjo de metadados é obrigatório ou não, se ele deve ser registrado na entrada do documento ou ao longo de sua estadia no acervo, entre outros. O CONARQ - Conselho Nacional de Arquivos (2011) - traz algumas normas para

referência. Elas fazem uso do modelo Dublin Core e podem ser vistas na tabela 2, a seguir:

Tabela 2 - Normas ISO relevantes para gestão de metadados

Organização	Norma	Descrição
ISO	ISO 23081-1	Informação e documentação - Processos de gestão de documentos de arquivo - Metadados para documentos de arquivo Parte 1: Princípios
ISO	ISO 15836	O documento define 15 elementos de metadados para a descrição de documentos.

Fonte: Elaborada pelo autor com base no site da ABNT.

Cabe ressaltar que todas as características sobre a gestão de metadados deve ser documentada na política de preservação, segundo o CONARQ - Conselho Nacional de Arquivos (2011). Uma prática interessante, citada por Flores e Santos (2015), é que, no arquivamento dos documentos digitais, seus respectivos arquivos de metadados sejam encapsulados no mesmo arquivo. Tal procedimento, além de proteger os documentos originais de manuseio equivocado, permite também que sejam protegidos os arquivos de metadados, além de unir no mesmo arquivamento os documentos que visam descrever. Silva Junior e Mota (2012) citam a importância do encapsulamento para a proteção dos metadados:

> O encapsulamento é a estratégia de preservar o conteúdo informacional com todos os metadados, de modo que possibilite, no futuro, o desenvolvimento de conversores, visualizadores ou emuladores.

Para Flores (2015), o encapsulamento pode também ser a melhor saída para a preservação digital quando o documento em questão não possui uma forma de conversão para formatos abertos, resguardando-o para soluções futuras:

> Os ciclos de obsolescência das tecnologias estão cada vez mais curtos, por isso a recuperação dos documentos torna-se mais improvável. Considerando-se a necessidade de se desenvolver um emulador ou conversor, além do custo elevado para a implementação das estratégias de preservação de tecnologia em longo prazo, pode-se dizer que o encapsulamento será, em um primeiro momento, o procedimento mais viável.

Mesmo havendo a proteção dos metadados por meio do encapsulamento, é importante que haja também um controle da autenticidade do volume encapsulado pelo cálculo periódico de *Hash*. Isso porque, ainda que os documentos digitais sejam bem conservados, seus suportes estarão em contínua degradação e seus dados passíveis de serem corrompidos. Essa situação deve ser o mais rápido possível identificada pelo corpo técnico para que sejam tomadas as devidas providências de substituição dos dados por *backups* íntegros e substituição dos suportes.

4.7.2 - CÁLCULO DE *HASH*

Além de manter os metadados já normalmente registrados relacionados ao conteúdo e aos processos administrativos, podemos adicionar o cálculo de *Hash* para aumentar ainda mais a presunção de autenticidade de um documento digital. Como dito anteriormente, o cálculo do *Hash* é um algoritmo que calcula, com base nos *bits* do conteúdo do documento digital, um valor correspondente. Quando o conteúdo do documento digital é alterado, consequentemente seus *bits* são alterados, e portanto seu valor de *Hash* também é modificado. Assim, a comparação do *Hash* armazenado nos metadados com o cálculo feito a qualquer momento serve para atestar - ou não - a autenticidade do documento.

Existem diversos algoritmos para cálculo do *Hash*, tais como o MD5, SHA1 e SHA2, entre outros. O CONARQ define que, quando usado o cálculo de *Hash* como forma de autenticação, certos metadados relativos ao *Hash* devem ser considerados: “identificação do cálculo, data do cálculo, agente responsável pelo cálculo e detalhes do cálculo” (2011, p. 122). Essas informações são importantes se pensarmos que os algoritmos de cálculo também sofrem obsolescência, sendo substituídos por algoritmos mais rápidos e/ou mais precisos. Como é necessária a periódica reanálise de *Hash* dos documentos digitais de um acervo, não é de todo errado imaginarmos que, em longo prazo, o cálculo deverá passar por uma reverificação no modelo já usado, e um novo cálculo deverá ser feito com base no modelo que se deseja implementar como substituto. Naturalmente, tanto a

alteração do algoritmo quanto os novos resultados de cálculo deverão ser devidamente documentados.

A presunção de autenticidade do documento digital só pode ser devidamente alcançada pela soma de gestão e preservação de seus metadados e um processo contínuo de verificação da integridade do documento por cálculo de *Hash*. A constatação da integridade pode ser feita de maneira automatizada, através de rotinas programadas dentro de um sistema, e também de maneira manual. Enquanto tecnologias de *machine learning* para classificação e catalogação de informação ainda não são acessíveis, certamente alguns metadados podem ser registrados de forma automatizada. Entretanto, aquilo que se pode automatizar deve ser feito de maneira cautelosa, pois a inserção, mesmo que estatisticamente baixa, de dados incorretos ou imprecisos pode resultar em uma recuperação falha, o que vai contra tudo o que representa uma política eficiente de preservação digital.

CONCLUSÃO

Este trabalho teve como objetivo reunir tanto a teoria da bibliografia especializada quanto a apresentação de modelos para a execução prática, no que tange ao campo tecnológico, de políticas de preservação digital. Para os interessados no assunto, esperamos que este levantamento possa ser interessante, ainda que sucinto, visto que depende de nós identificarmos os caminhos a trilhar em terras ainda desconhecidas, da mesma forma que outras gerações de profissionais da informação, em séculos passados, tiveram de se debruçar sobre suas mesas em busca de maneiras de preservar a informação de suas sociedades, o que nunca foi uma missão fácil. Atualmente, estamos apenas iniciando novos ciclos de questionamentos e aprendizagem, condizentes com a complexidade de nossa época.

Podemos dizer que, com o tempo, toda organização deverá ter, em algum nível, a aplicação de políticas de preservação digital, seja em toda a organização, de forma setorial ou apenas em relação a documentos específicos. Isto dependerá da relevância que dada organização enxergar em sua produção, sendo tendência aumentar a abrangência e a complexidade das políticas com o tempo, pois, em dado momento, as organizações necessitarão com mais frequência possuir formas de averiguar a autenticidade, a confiabilidade e a integridade dos documentos digitais de seus acervos. Vislumbramos, em um futuro não muito distante, uma sociedade cujas instituições de gestão públicas estarão cada vez mais instrumentalizadas tecnologicamente e exigirão garantias maiores de que a informação contida em meio digital seja de fato confiável quando assim precisarem.

No momento, os suportes que contêm informação digital não são desenvolvidos para durar em longo prazo, por razão da lógica capitalista de consumo, na qual a obsolescência programada é uma estratégia bem difundida da indústria eletrônica. É desafiador para bibliotecários e arquivistas executarem sua árdua missão sem conhecimento mais avançado em tecnologia da informação e sem o auxílio de profissionais da tecnologia da informação para colaborar com soluções computacionais robustas e inteligentes para substituir atividades intensivas pelo poder computacional. Entretanto, depende dos profissionais da

ciência da informação dominar os conhecimentos necessários para serem protagonistas dessa realidade.

As políticas de preservação digital só serão de fato efetivas com recursos suficientes, seja de capital humano ou financeiro, proporcionais à importância do assunto. Isso não necessariamente significa que a preservação digital será para sempre uma atividade financeiramente dispendiosa, pois isso dependerá do tamanho e da complexidade do acervo, sendo essas variáveis teoricamente proporcionais aos recursos que dada instituição possui, mas ela certamente exigirá pessoal qualificado e focado no objetivo de preservar, sendo essa qualificação um dos principais abismos entre teoria e execução, além do financiamento. A despeito de haver extensa bibliografia sobre o assunto, como cada organização possui suas especificidades, torna-se quase impossível prever quais serão, na prática, os processos de trabalho de uma política de preservação digital em uma organização. Cada caso terá suas especificidades e caberá a pessoal qualificado criar a melhor política possível. Ademais, acreditamos que as políticas de preservação digital nas organizações só se tornarão populares após a criação de leis específicas para tal. Um possível primeiro passo talvez tenha sido dado com a criação da LGPD – Lei Geral de Proteção de Dados (2018), na qual o dado digital toma relevância social, tornando-se um objeto informacional específico, diferente se comparado ao mesmo conteúdo em meio analógico, pois tal lei ressalta o quanto a informação em meio digital pode ser usada de maneiras muito mais diversas, e nem sempre muito claras.

A bibliografia abordada neste trabalho permitiu identificar quais são alguns dos caminhos possíveis para o desenvolvimento de políticas de preservação digital. Foi identificado que poucos trabalhos acadêmicos sobre o assunto citavam ferramentas específicas. O que a princípio aparentava ser uma forma superficial de abordagem provou-se, no decorrer do trabalho, ser uma forma de não induzir os interessados no assunto a se perder na busca por *softwares* que podem ficar datados rapidamente. Em vez disso, colabora para a busca de um melhor entendimento de quais os preceitos para a escolha de metodologias ou *softwares* pelos profissionais levando em consideração seu contexto técnico. A literatura abordada neste trabalho mostrou-se rica, mesmo as produções datadas da década

passada, já que, no que tange à preservação digital, mesmo com o avanço da tecnologia, muitos preceitos se mantém ainda válidos para 2020, pois estão intrinsecamente relacionados com questões técnicas do comportamento dos dispositivos digitais até onde conhecemos.

Esperamos que este trabalho estimule as discussões de políticas de preservação digital e quiçá sua execução, pois, sem essas políticas, muito de nossa época cairá, segundo Silva (2015), em uma amnésia coletiva digital. Então, não só as riquezas de nosso tempo serão perdidas, mas também nossos erros históricos, sendo ambas fontes de informação indispensáveis para a construção de uma sociedade melhor. Cabe a nós, profissionais da informação, e à sociedade civil à qual pertencemos, defendermos que haja mais interesse do poder público, das organizações privadas, e da sociedade como um todo, em preservar a memória de nosso tempo. Vivenciamos um momento histórico de transição, no qual a tecnologia alterou substancialmente a forma com que os seres humanos se comunicam, retém e difundem a informação. Recai, assim, sobre nós e as novas gerações a responsabilidade de darmos os primeiros passos em direção à construção de uma cultura de preservação digital, principalmente em nosso campo laboral.

REFERÊNCIAS

ABNT. Associação Brasileira de Normas Técnicas. Disponível em: <http://www.abnt.org.br/>. Acesso em: 29 dez. 2019.

ALLENTOFT, Morten Erik et al. The half-life of DNA in bone: measuring decay kinetics in 158 dated fossils. **Proceedings of the Royal Society B: Biological Sciences**. 2012. Disponível em: <https://royalsocietypublishing.org/doi/10.1098/rspb.2012.1745>. Acesso em: 06 out. 2019.

ARELLANO, Miguel Ángel Márdero. 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/article/view/1043/1113>. Acesso em: 22 fev. 2020.

ARELLANO, Miguel Ángel Márdero; ANDRADE, Ricardo Sodré. Preservação digital e os profissionais da informação. **DataGramaZero**, v. 7, n. 5, out. 2006. Disponível em: <http://www.brapci.inf.br/index.php/res/v/5978>. Acesso em: 30 ago. 2019.

ARQUIVO NACIONAL. **Política de Preservação Digital**. dez. 2016. Disponível em: <http://www.siga.arquivonacional.gov.br/images/an_digital/and_politica_preservacao_digital_v2.pdf>. Acesso em: 22 dez. 2019.

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. **ABNT NBR 11515:2007**: Guia de práticas para segurança física relativas ao armazenamento de dados. 2007. Disponível em: <https://www.abntcatalogo.com.br/norma.aspx?ID=003199>. Acesso em: 22 fev. 2020.

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. **ABNT NBR 15247:2004**: Unidades de armazenagem segura - Salas cofre e cofres para hardware - Classificação e método de ensaio de resistência ao fogo. 2004. Disponível em: <https://www.abntcatalogo.com.br/norma.aspx?ID=141>. Acesso em: 22 fev. 2020.

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. **ABNT NBR 15472:2007**: Sistemas espaciais de dados e informações - Modelo de referência para um sistema

aberto de arquivamento de informação (SAAI). 2007. Disponível em: <https://www.abntcatalogo.com.br/norma.aspx?ID=138>. Acesso em: 22 fev. 2020.

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. **ABNT ISO 14721:2012**: Space data and information transfer systems -- Open archival information system (OAIS) -- Reference model. 2012. Disponível em: <https://www.abntcatalogo.com.br/norma.aspx?ID=92010>. Acesso em: 22 fev. 2020.

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. **ABNT ISO 15836-1:2017**: Information and documentation — The Dublin Core metadata element set — Part 1: Core elements. 2017. Disponível em: <https://www.abntcatalogo.com.br/norma.aspx?ID=371207>. Acesso em: 22 fev. 2020.

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. **ABNT ISO 16363:2012**: Space data and information transfer systems -- Audit and certification of trustworthy digital repositories. 2012. Disponível em: <https://www.abntcatalogo.com.br/norma.aspx?ID=90122>. Acesso em: 22 fev. 2020.

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. **ABNT ISO 23081-1:2017**: Information and documentation — Records management processes — Metadata for records — Part 1: Principles. 2017. Disponível em: <https://www.abntcatalogo.com.br/norma.aspx?ID=381660>. Acesso em: 22. fev. 2020.

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. **ABNT ISO/IEC 27001:2013**: Information technology -- Security techniques -- Information security management systems -- Requirements. 2013. Disponível em: <https://www.abntcatalogo.com.br/norma.aspx?ID=304866>. Acesso em: 22 fev. 2020.

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. **ABNT ISO/IEC 27002:2013**: Information technology -- Security techniques -- Code of practice for information security controls. 2013. Disponível em: <https://www.abntcatalogo.com.br/norma.aspx?ID=304865>. Acesso em: 22 fev. 2020.

BARROS, Diego Bil; CASTRO, Jetur Lima de; ARELLANO, Miguel Ángel Márdero. Mapeamento das Revistas do Portal de Periódicos da Universidade Federal do Pará: uma

abordagem sobre a importância da elaboração de políticas e estratégias de preservação digital. **Informação & Informação**, v. 23, n. 3, p. 38-64, dez. 2018. Disponível em: <http://www.uel.br/revistas/uel/index.php/informacao/article/view/27503>. Acesso em: 22 fev. 2020.

BEAGRIE, Neil; GREENSTEIN, Daniel. **A strategic policy framework for creating and preserving digital collections**: a report to the Digital Archiving Working Group. London: British Library Research and Innovation Centre, 1998. Disponível em: <ukoln.ac.uk/services/elib/papers/supporting/pdf/framework.pdf>. Acesso em: 03 nov. 2019.

BELLOTTO, H. L. Arquivos permanentes: tratamento documental. Rio de Janeiro: Ed. da FGV, 2006.

BOERES, Sonia Araújo de Assis. **Política de preservação da informação digital em bibliotecas universitárias brasileiras**. 2004. 180 f. Dissertação (Mestrado em Ciência da Informação e Documentação) – Universidade de Brasília, Brasília. Disponível em: <https://repositorio.unb.br/handle/10482/1693>. Acesso em: 03 nov. 2019.

BRASIL. Lei nº 12.527, de 18 de novembro de 2011. Regula o acesso a informações previsto no inciso XXXIII do art. 5o, no inciso II do § 3o do art. 37 e no § 2o do art. 216 da Constituição Federal; altera a Lei no 8.112, de 11 de dezembro de 1990; revoga a Lei no 11.111, de 5 de maio de 2005, e dispositivos da Lei no 8.159, de 8 de janeiro de 1991; e dá outras providências. **Diário Oficial da União**, Brasília, DF, 18 nov. 2011. Disponível em: <http://www.planalto.gov.br/ccivil_03/_ato2011-2014/2011/lei/l12527.htm>. Acesso em: 22. fev. 2020.

BRASIL. Lei Nº 13.709, de 14 de agosto de 2018. Lei Geral de Proteção de Dados Pessoais (LGPD). **Diário Oficial da União**, Brasília, DF, 15 ago. 2018. Disponível em: <http://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/L13709.htm>. Acesso em: 22 fev. 2020.

BRITO, Ronnie Fagundes de et al. **Guia do Usuário do Digital Object Identifier**. Brasília: IBICT, 2016. Disponível em:

<https://www.abecbrasil.org.br/arquivos/Guia_usuario_DOI-online3.pdf>. Acesso em: 25 ago. 2019.

BULLOCK, Alisson. **Preservation of digital information**: issues and current status. 1999. Disponível em: <http://epe.lac-bac.gc.ca/100/202/301/netnotes/netnotes-h/notes60.htm>. Acesso em: 23 fev. 2020.

CASTELLS, Manuel. **A sociedade em rede**. 10 ed. São Paulo: Paz e Terra, 2010.

CHIAVENATO, Idalberto. **Introdução à teoria geral da administração: uma visão abrangente da moderna administração das organizações**. Rio de Janeiro: Elsevier, 2003.

CONSELHO NACIONAL DE ARQUIVOS (CONARQ). **Carta para a Preservação do Patrimônio Arquivístico Digital**. Brasília: Conselho Nacional de Arquivos**,** 2005. Disponível em: <http://www.conarq.gov.br/images/publicacoes_textos/Carta_preservacao.pdf>. Acesso em: 25 ago. 2019.

CONSELHO NACIONAL DE ARQUIVOS (CONARQ). **Modelo de Requisitos para Sistemas Informatizados de Gestão Arquivística de Documentos**. Rio de Janeiro: Conselho Nacional de Arquivos, 2011. Disponível em: <http://www.siga.arquivonacional.gov.br/images/publicacoes/e-arq.pdf>. Acesso em: 25 ago. 2019.

CUNHA, Murilo Bastos da; CAVALCANTI, Cordélia Robalinho de Oliveira. **Dicionário de Biblioteconomia e Arquivologia.** Brasília: Briquet de Lemos, 2008. Disponível em: <https://repositorio.unb.br/handle/10482/34113>. Acesso em: 19 jan. 2020.

ARQUIVO NACIONAL. **Dicionário brasileiro de terminologia arquivística**. Rio de Janeiro: Arquivo Nacional, 2005. Disponível em: <http://www.arquivonacional.gov.br/images/pdf/Dicion_Term_Arquiv.pdf>. Acesso em: 07 mai. 2020

DURANTI, Luciana; PRESTON, Randy. **International Research on Permanent Authentic Records in Electronic Systems (InterPARES) 2**: Experiential, Interactive and Dynamic Records. Italy: Associazione Nazionale Archivistica Italiana, 2008. Disponível em: <http://www.interpares.org/ip2/display_file.cfm?doc=ip2_book_complete.pdf>. Acesso em: 29 ago. 2019.

DURBANO, Vinicius. **Saiba diferenciar os 4 tipos de backup que você pode utilizar**. ECOIT, 2019. Disponível em: <https://ecoit.com.br/tipos-de-backup/>. Acesso em: 19 jan. 2020.

FARIA FILHO, Luciano Mendes de. **Arquivos, fontes e novas tecnologias**: questões para a história da educação. Campinas: Autores Associados, 2000.

FERREIRA, Flávia Catarino Conceição; SILVA, Rubens Ribeiro Gonçalves da. Preservação, Salvaguarda da Informação Pública e Repositórios Digitais: noções em pauta., **Páginas a&b**, S.3, n. especial, p. 17-29, 2018. Disponível em: <http://ojs.letras.up.pt/index.php/paginasaeb/article/view/3953>. Acesso em: 22. fev. 2020.

FERREIRA, Miguel. **Introdução à preservação digital**: conceitos, estratégias e actuais consensos. Portugal: Escola de Engenharia da Universidade do Minho, 2006. Disponível em: <http://repositorium.sdum.uminho.pt/bitstream/1822/5820/1/livro.pdf>. Acesso em: 03 nov. 2019.

FERREIRA, Miguel; SARAIVA, Ricardo; RODRIGUES, Eloy. **Estado da arte em preservação digital**. 2012. Disponível em: <http://hdl.handle.net/1822/17049>. Acesso em: 09 fev. 2020.

FIALHO JUNIOR, Mozart. **Guia essencial do backup**. São Paulo: Digerati Books, 2007.

FLORES, Daniel; SANTOS, Henrique Machado dos. As estratégias de emulação como fundamento para a preservação de objetos digitais interativos: a garantia de acesso fidedigno em longo prazo. **Informação Arquivística**, Rio de Janeiro, v.3, n.1, p.95-116, jan./jun., 2014. Disponível em:

<http://www.aaerj.org.br/ojs/index.php/informacaoarquivistica/article/view/79/34>. Acesso em: 12 jan. 2020.

FLORES, Daniel; SANTOS, Henrique Machado dos. Preservação de documentos arquivísticos digitais: reflexões sobre as estratégias de encapsulamento. **Liinc em Revista**, Rio de Janeiro, v. 11, n. 1, p. 167-180, maio/2015. Disponível em: <http://revista.ibict.br/liinc/article/view/3610>. Acesso em: 12 jan. 2020.

INNARELLI, Humberto Celeste. Os dez mandamentos da preservação digital. In: SANTOS, V. B.; INNARELLI, H.C.; SOUSA, R. T. B. In: **Arquivística**: temas contemporâneos. Brasília: SENAC, 2007.

INNARELLI, Humberto Celeste. **Gestão de preservação de documentos arquivísticos digitais**: proposta de um modelo conceitual. 2015. 348f. Tese (Doutorado em Ciências da Informação) – Escola de Comunicações e Artes, Universidade de São Paulo, São Paulo. Disponível em: <https://www.teses.usp.br/teses/disponiveis/27/27151/tde-27052015-101628/pt-br.php>. Acesso em: 22 fev. 2020.

INNARELLI, Humberto Celeste. Introdução aos dez mandamentos da preservação digital. **Sínteses: Revista Eletrônica do SimTec**, n. 2, 12 set. 2016. Disponível em: <https://econtents.bc.unicamp.br/inpec/index.php/simtec/article/view/8483>. Acesso em: 27 out. 2019.

INTERNET ARCHIVE. Archive.org: Historical Software Collection. 2019. Disponível em: <https://archive.org/details/historicalsoftware>. Acesso em: 08 dez. 2019.

LITERIS. Descomplicando a Gestão do Conhecimento. **Literis Treinamento Online**. 2015. Disponível em: <https://literis.com.br/blog/descomplicando-a-gestao-do-conhecimento/> Acesso em: 19 jan 2020.

LOGAN, Robert K. **Que é informação?**: A propagação da organização na biosfera, na simbolosfera, na tecnosfera e na econosfera. Rio de Janeiro: Editora PUC-Rio, 2012.

LOGAN, Robert K. What Is Information? Why Is It Relativistic and What Is Its Relationship to Materiality, Meaning and Organization. **Information,** v. 3, p. 68–91. 2012. Disponível em: <https://www.mdpi.com/2078-2489/3/1/68>. Acesso em: 29 set. 2019.

LUZ, Charlley dos Santos. Curadoria Digital, Custódia Arquivística: relações possíveis. **Páginas a&b**, S.3, n. 10, p. 92-103, 2018. Disponível em: <https://ojs.letras.up.pt/index.php/paginasaeb/article/view/4775>. Acesso em: 22 fev. 2020.

LUZ, Charlley dos Santos; MARINGELI, Isabel Cristina Ayres da Silva. Política de preservação digital: caso Pinacoteca de São Paulo. **Perspectivas em Ciência da Informação**, v. 23, n. 2, p. 189-200, jul. 2018.
Disponível em: <http://portaldeperiodicos.eci.ufmg.br/index.php/pci/article/view/2995>. Acesso em: 01 set. 2019.

MACKAY, Donald MacCrimmon. **Information, Mechanism and Meaning**.
Massachusetts: MIT Press, 1969.

MARCONDES, Carlos Henrique. Interoperabilidade entre acervos digitais de arquivos, bibliotecas e museus: potencialidades das tecnologias de dados abertos interligados. **Perspectivas em Ciência da Informação**, v. 21, n. 2, p. 61-83, jun. 2016. Disponível em: <http://portaldeperiodicos.eci.ufmg.br/index.php/pci/article/view/2735>. Acesso em: 23 jun. 2019.

Modos RAID. **Seagate Manuals**. Disponível em: <https://www.seagate.com/br/pt/manuals/network-storage/business-storage-nas-os/raid-modes/>. Acesso em: 19 jan. 2020.

MORIMOTO, Carlos E. **RAID 5**. Hardware. 26 jun. 2005. Disponível em: <https://www.hardware.com.br/termos/raid-5>. Acesso em: 19 jan. 2020.

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

SANTOS, Henrique Machado dos; FLORES, Daniel. Políticas de preservação digital para documentos arquivísticos. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 20, n. 4, p. 197-217, dez. 2015. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1413-99362015000400197&lng=en&nrm=iso>. Acesso em: 20 jun. 2019.

SCHÄFER, Murilo Billig; CONSTANTE, Sônia Elisabete. Políticas e estratégias para a preservação da informação digital. **Ponto de Acesso**, Salvador, v. 6, n. 3, p. 108-140, dez. 2012. Disponível em: <https://portalseer.ufba.br/index.php/revistaici/article/view/6449/4817>. Acesso em: 25 ago. 2019.

SHANNON, Claude. A mathematical theory of communication. **The Bell System Technical Journal**, v. 27, p. 379-423 & 623-656, jul. /out. 1948. Disponível em: <http://www.math.harvard.edu/~ctm/home/text/others/shannon/entropy/entropy.pdf>. Acesso em: 29 set. 2019.

SILVA, Fabíola Rubim. Preservação digital: um diagnóstico da literatura especializada brasileira. **Biblionline**, João Pessoa, v. 11, n. 2, p. 57-72, 2015. Disponível em: <http://www.brapci.inf.br/index.php/res/download/50927>. Acesso em: 23 jun. 2019.

SILVA, William; FLORES, Daniel. Política arquivística de preservação digital: um estudo sobre sua aplicabilidade em instituições públicas federais. **Perspectivas em Ciência da Informação**, v. 23, n. 3, p. 144-166, set. 2018. Disponível em: <http://portaldeperiodicos.eci.ufmg.br/index.php/pci/article/view/3187>. Acesso em: 07 jul. 2019.

SILVA JÚNIOR, Laerte Pereira da; MOTA, Valéria Gameleira da. Políticas de preservação digital no Brasil: características e implementações. **Ciência da Informação**, v. 41, n. 1, p. 51-64, jan./abr., 2012. Disponível em: <http://www.brapci.inf.br/index.php/article/view/0000015692/2a2b0400527992a857365f913df9c7bb/>. Acesso em: 22. fev. 2020

SOUSA, Maciel. Gestão da Informação: do modelo de segurança e preservação ao repositório confiável. **Páginas a&b**, S.3, n. 1, p. 91-119, 2014. Disponível em: <ojs.letras.up.pt/index.php/paginasaeb/article/view/572>. Acesso em: 22 fev. 2020.

UNESCO/NLA. **Guidelines for the preservation of digital heritage**. National Library of Australia**,** mar. 2003**.** Disponível em: <https://unesdoc.unesco.org/ark:/48223/pf0000130071>.Acesso em: 25 ago. 2019.