



XXI ENANCIB

Encontro Nacional de Pesquisa em Ciência da Informação

50 anos de Ciência da Informação no Brasil:
diversidade, saberes e transformação social

Rio de Janeiro • 25 a 29 de outubro de 2021

XXI Encontro Nacional de Pesquisa em Ciência da Informação – XXI ENANCIB

GT-8 – Informação e Tecnologia

A IMPORTÂNCIA DOS FORMATOS DE ARQUIVO E DOS CODECS NA PRESERVAÇÃO DOS DOCUMENTOS DIGITAIS

THE IMPORTANCE OF FILE FORMATS AND CODECS IN THE PRESERVATION OF DIGITAL DOCUMENTS

Êda Maria Bastos de Moura – Universidade Federal Fluminense (UFF)

Linair Maria Campos – Universidade Federal Fluminense (UFF)

Modalidade: Trabalho Completo

Resumo: O trabalho tem por finalidade contribuir para uma maior conscientização dos profissionais de informação sobre a importância do contexto tecnológico relacionado aos documentos digitais e sua preservação. Objetivo: explicitar diferentes aspectos relacionados aos formatos de arquivo e os seus tipos, bem como abordar a importância dos *codecs* aos documentos audiovisuais e como tais tecnologias, sem a devida atenção quanto às suas vulnerabilidades, podem comprometer a preservação dos documentos digitais em longo prazo. Metodologia: revisão bibliográfica de natureza qualitativa e exploratória. Conclusão: os formatos de arquivo e os *codecs* apresentam desafios para a preservação digital, desafios esses que devem ser compreendidos pelos profissionais responsáveis pela sua guarda, sob risco de que as ações tomadas para a sua preservação a longo prazo possam ser feitas de maneira equivocada, por desconhecimento dos aspectos que são inerentes a esse tipo de documento, e o contexto tecnológico ligado ao seu nível lógico. Resultado: explicitação, de forma sistemática, de aspectos básicos relacionados ao contexto tecnológico do documento digital no que tange os formatos de arquivo e os *codecs*, que impactam na preservação do documento digital.

Palavras-Chave: Formatos de arquivo; *Codecs*; Documento Digital; Preservação Digital.

Abstract: *The work aims to contribute to a greater awareness of information professionals about the importance of the technological context related to digital documents and their preservation. Objective: explaining different aspects related to file formats and their types, as well as addressing the importance of codecs to audiovisual documents and how such technologies, without due attention to their vulnerabilities, can compromise the preservation of digital documents in the long term. Methodology: bibliographical review of a qualitative and exploratory nature. Conclusion: file formats and codecs present challenges for digital preservation, challenges that must be understood by professionals who are responsible for their storage, at the risk that the actions taken for their long-term preservation can be done in a way mistaken, due to lack of knowledge of the aspects that are inherent to this type of document, and its technological context linked to its logical level. Result: systematic explanation of basic aspects related to the technological context of the digital document regarding file formats and codecs, which impact the preservation of the digital document.*

Keywords: *File Formats; Codecs; Digital Document; Digital Preservation.*

1 INTRODUÇÃO

A sociedade contemporânea tem passado por profundos avanços tecnológicos que surgem vertiginosamente. Os computadores, a internet e a evolução tecnológica têm contribuído para que as informações possam estar acessíveis em várias dimensões da vida histórico-social, científica, cultural e institucional de uma nação. Segundo Innarelli (2011, p. 83) “a construção e uso da informação recebida, processada e transmitida são colocadas como pilares para construção do conhecimento e a evolução da cultura humana”, A informação em ambiente digital tem relevância significativa para que a sociedade evolua em conhecimento. De acordo com Grácio, Fadel e Valetim (2013, p. 1) “a informação digital tem contribuído para o desenvolvimento social e cultural, pois propicia a diferentes comunidades, a possibilidade de acesso à informação e ao conhecimento”, entretanto, embora essas inovações tecnológicas sejam vantajosas na rotina de todas as atividades humanas, elas apresentam grandes desafios na preservação dos conteúdos digitais em longo prazo, pois os documentos digitais podem ficar em constante vulnerabilidade tecnológica.

O documento digital depende da tecnologia para ser decifrado e lido. E isso representa uma das especificidades inerentes a ele próprio. De acordo com Flores e Santos (2015, p. 48), “a especificidade dos documentos digitais em seus contextos tecnológicos afeta diretamente o acesso e o uso de seus conteúdos”. Assim, os documentos digitais apresentam uma fragilidade intrínseca que os colocam em permanentes riscos, o que torna a preservação dos conteúdos de natureza digital um dos grandes desafios da contemporaneidade (SAYÃO, 2010). Além das especificidades dos documentos em ambiente digital, as mudanças frequentes da tecnologia (suas diversidades de suportes e formatos) têm contribuído para que ocorram as obsolescências tecnológicas e as falhas tecnológicas, o que comprometeria a preservação dos acervos digitais. Tal problema se justifica pela possível incompatibilidade no que tange os novos suportes e os *hardwares* necessários para acessá-los e os formatos de arquivo e os *softwares* necessários para decodificá-los. De acordo com Avancini (2012, p. 1) “a própria evolução tecnológica pode impor dificuldades à preservação da documentação digital, já que os formatos usados para gravar os documentos são rapidamente substituídos por novos”. Os formatos de arquivos são infraestruturas tecnológicas indispensáveis para a existência dos documentos digitais. Os documentos audiovisuais, por sua vez, apresentam uma dependência adicional de *software* para

compactação e codificação/decodificação de dados, os *codecs*. Embora a questão dos formatos de arquivo seja abordada na literatura, não é possível encontrar, até onde pode ser percebida, uma visão abrangente dos aspectos que estão envolvidos no contexto tecnológico dos documentos digitais, em especial, no que diz respeito aos *codecs*.

Dessa forma, o problema do trabalho está centrado no fato que o desconhecimento sobre os aspectos do contexto tecnológico e sobre os formatos de arquivo podem prejudicar ações de preservação digital. Diante do exposto, o objetivo do trabalho é explicitar diferentes aspectos relacionados aos formatos de arquivo e seus tipos, bem como abordar a importância dos *codecs* em relação aos documentos audiovisuais que podem influir na preservação digital em longo prazo. Nesse sentido, o trabalho se justifica em função da importância dessa compreensão para que os profissionais, que trabalham na guarda dos arquivos em meio digital, estejam cientes dos aspectos tecnológicos ligados à parte lógica do arquivo digital, o que vem a ser um desafio talvez maior do que o cuidado com a sua parte física – em especial, considerando que boa parte desses profissionais não necessariamente possui em sua formação determinados conhecimentos ligados à área da tecnologia da informação. Além disso, a contribuição do trabalho para Ciência da Informação se dá por discutirmos como um dos objetos de estudo dos mais importantes da Ciência da Informação, o documento, pode estar em constante vulnerabilidade no ambiente digital.

No que tange à metodologia adotada: quanto aos objetivos à pesquisa é de caráter exploratório; quanto à abordagem, é qualitativa; quanto aos seus procedimentos técnicos, a coleta de dados foi feita por meio de pesquisa bibliográfica, para a qual foram utilizadas as bases de dados dos portais da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), da Base de Dados Referenciais de Artigos de Periódicos em Ciência da Informação (BRAPCI), da Biblioteca Digital Brasileira de Teses e Dissertações (BDTD) e a Base de Dados dos Encontros Nacionais de Pesquisa e Pós-Graduação em Ciência da Informação e Biblioteconomia (BENANCIB). As escolhas para a utilização das bases de dados acima citadas foram feitas pelo reconhecimento junto à comunidade científica, por suas pesquisas desenvolvidas com padrões, critérios e qualidade e pelas publicações reconhecidas por seus pares, nacionalmente e internacionalmente, por meio de revistas e livros.

Como método usado para a pesquisa bibliográfica, adotou-se a análise de conteúdo de Bardin (1995, p. 19) que “é uma técnica de investigação que tem por finalidade a descrição objetiva, sistemática e quantitativa do conteúdo manifesto da comunicação”. A busca foi feita mediante o uso de palavras-chaves e para o entendimento preliminar do material utilizado foi adotada a técnica de leitura flutuante conforme Bardin (1995). Nela foram selecionados os artigos considerados como relevantes para a pesquisa. Quanto às palavras-chaves, foram utilizadas as seguintes: formatos de arquivo, *codecs*, documento digital e preservação digital. Após a seleção do material foi feita a sua exploração com a definição de categorias de análise e, por fim, o tratamento dos resultados e sua interpretação.

O trabalho está estruturado como segue: na seção 2, será abordado o documento digital e suas características, nas subseções 2.1 e 2.2, respectivamente, discutem-se os formatos de arquivos e os *codecs*. Na seção 3, será abordada a importância da conscientização acerca da vulnerabilidade tecnológica do documento digital. Na seção 4, apresentar-se-ão as considerações finais.

2 O DOCUMENTO DIGITAL E SUAS CARACTERÍSTICAS

Durante muitas décadas, pensou-se em documento com algo rígido, fixo, imutável e palpável. Alimentaram-se tais concepções, em grosso modo, antes do advento da tecnologia de informação. Os registros deixados pela humanidade no decorrer de sua evolução, da história das civilizações no percurso dos tempos ficaram, de certa forma, na memória da coletividade tais como os documentos escritos em argila, pergaminho, papiro e papel. De acordo com Moura e Campos (2020, p. 4) “um documento existe a partir de alguma forma de registro e em algum tipo de suporte”. Contudo, no universo dos documentos digitais, há uma natureza dinâmica e de variabilidade tecnológica na forma e no conteúdo dos documentos digitais que devem ser levadas em conta (RONDINELLI, 2013).

O Conselho Nacional de Arquivos (2014, p. 20) define que documento digital é uma “informação registrada, codificada em dígitos binários, acessível e interpretável por meio de sistema computacional”. No dicionário *InterPARES 3*, o documento digital é compreendido como “um componente digital ou grupo de componentes digitais, salvo e tratado e gerenciado como um documento” (INTERPARES 3, 2012, n.p), isto é, uma unidade e/ou item

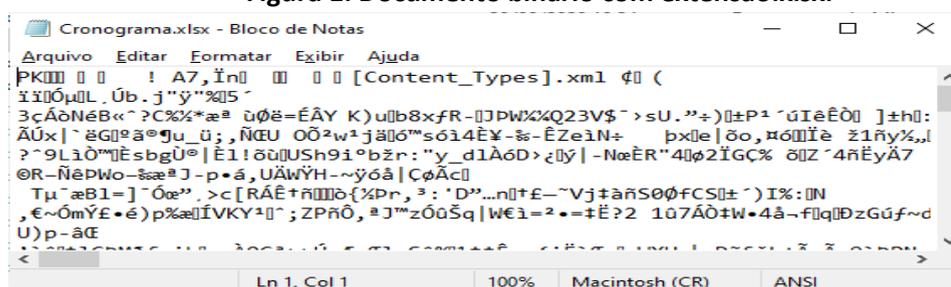
documental digital. Há também de ser explicitar que no contexto tecnológico digital o documento digital possui diversos tipos que de acordo com Yamaoka e Gauthier (2012, p. 214) podem ser “textos, fotos, diagramas, filmes, áudios e programas executáveis”.

Um documento digital para ser decifrado e lido e depende de uma infraestrutura complexa de um sistema computacional. Em relação à dependência tecnológica, Bodê (2016, p. 4) enfatiza que “para o uso de documentos digitais passa a ser imprescindível a presença de *software*, além de equipamentos para sua leitura (computadores e seus periféricos)”. Dessa forma, o conteúdo do arquivo, que corresponde ao seu nível intelectual, possui aspectos lógicos, que são mais ligados aos *softwares* usados para a sua leitura, e aspectos físicos, que dizem respeito ao suporte onde o arquivo está materializado e armazenado, bem como do *hardware* necessário para acessar e ler o conteúdo registrado nesse suporte (FERREIRA, 2006). Outro enfoque relevante é a natureza binária do documento digital, que é a linguagem única e exclusiva computacional inerente ao documento digital. Rondinelli (2013) enfatiza que a linguagem binária é uma característica exclusiva do documento digital, que pode ser dividida em documentos digitais, quanto à forma de codificação do conteúdo; em documentos digitais textuais puros e documentos digitais binários. Os chamados documentos textuais puros são estruturados de maneira diferente dos documentos binários (UNIVERSIDADE DE AMSTERDAM, 2020). Cabe dar ênfase que, a rigor, todos os arquivos em meio digital são codificados em código binário. Entretanto, essa é a nomenclatura usual à área de Tecnologia de Informação (TI) utilizada para diferenciar os dois tipos de formatos de arquivos pela codificação de seu conteúdo (MOURA, 2021).

Os documentos digitais, cujos formatos de arquivos são tipos textuais puros, podem ser lidos e interpretados com editores de texto tipo *Notepad ou NoteTab* (UNIVERSIDADE DE AMSTERDAM, 2020), não oferecem maiores desafios à preservação digital quanto à questão da obsolescência de *software*. Entretanto, o mesmo já não ocorre com os binários: nesse tipo de arquivo os conteúdos podem ser diversos, tais como: sons, imagens, vídeos, etc. Além disso, pode estar comprimido e organizado de diferentes formas, o que implica na necessidade de se conhecer os detalhes de sua especificação de formato para que o mesmo possa ser apresentado de forma adequada. Os arquivos textuais basicamente podem conter textos sem uma formatação ou estrutura preliminares ou podem conter, dentro dos próprios textos, instruções. Em suma, ao analisar as extensões desses arquivos observa-se, por

exemplo, que a extensão .txt. contém, de maneira geral, somente texto. Já os arquivos .html contêm, juntamente com o conteúdo textual, os dados sobre sua formatação. Os arquivos textuais tipo.xml contêm texto, dados, estrutura e identificação do conteúdo do texto. Seja qual for o tipo de arquivo textual, ele contém apenas texto, visualizado em um formato de codificação, o qual tem o papel de fornecer uma equivalência numérica para cada caractere, uma vez que o computador armazena qualquer conteúdo em código binário. O formato de codificação influencia apenas nos tipos de caracteres que serão permitidos expressar, sendo disponíveis de forma pública. Por exemplo, no formato de codificação ASCII, a letra A equivale ao número binário 00100001, mas não podemos representar caracteres em outro idioma, como o chinês, o que já é possível com o formato de codificação UTF-8. Já nos documentos binários, o conteúdo se acha codificado de outra maneira. Na figura 1 ilustramos essa diferença, ao abrirmos um documento em um formato binário no bloco de notas do *Windows*. Como podemos perceber não se trata apenas de caracteres simples codificados de acordo com um esquema padrão de codificação.

Figura 1: Documento binário com extensão.xlsx.



Fonte: elaboração própria.

Por outro lado, os documentos audiovisuais, além de serem binários, possuem uma camada de infraestrutura tecnológica mais complexa, envolvida na compressão e codificação de seu conteúdo, que pode muitas vezes não ser perceptível nas atividades rotineiras de uma instituição, mas que podem comprometer preservação dos documentos digitais em longo prazo. Esses aspectos serão detalhados a seguir.

2.1 FORMATOS DE ARQUIVO

Para que os documentos digitais sejam preservados em longo prazo é preciso saber como os seus conteúdos estão alocados dentro de um sistema, pois esses documentos são codificados em uma cadeia de *bits*, e armazenados em arquivos digitais e como tal

necessitam de especificações que são cruciais para sua decifração. De acordo com a *UC Curation Center* (2012, p. 3) “um formato é o conjunto de regras sintáticas e semânticas que governam o mapeamento entre informações e *bits* que representam essa informação”.

O formato de um documento digital pode representar tecnologicamente um conjunto de *bits* e seus conteúdos podem ser em texto, som ou imagem. De acordo com Bodê (2008, p. 54) “um formato de arquivo qualquer especifica como um determinado conteúdo está estruturado”. A especificação do formato de um arquivo é um fator determinante para que haja identificação exata do tipo de documento e a configuração (forma) na qual ele se apresenta. É a trilha digital em *bits*.

Uma especificação de formato de arquivo – normalmente chamada formato de arquivo simplesmente – é a explicação, normalmente registrada num documento formal, da disposição dos *bits* de um arquivo digital e a função desses *bits* ou grupos de *bits* (BODÊ, 2008, p. 22).

No âmbito das especificações técnicas e formatos de arquivos para a preservação digital, o Projeto *InterPARES* (2007) tratou acerca das especificações e dos padrões de formatos de arquivos abertos e fechados para preservação digital. Sobre os padrões abertos que são importantes para a preservação digital, o Projeto *InterPARES* ressalta que os padrões abertos, sejam eles formatos de arquivos ou *softwares*, são aqueles que possuem suas especificações publicadas e disponibilizadas gratuitamente. Podendo significar que são amplamente utilizados e/ou livres de patentes (INTERPARES 2, 2007). Quanto à propriedade (direito legal sobre o padrão), o padrão do formato pode ser definido como proprietário ou não proprietário. Nesse sentido, pode-se definir um formato proprietário como aquele que uma empresa possui e controla, sendo que sua especificação pode sofrer mudanças ao critério da empresa que o possui (GLOSSÁRIO OPEN DATA HARDBOOK, 2021). Uma vez que o direito à propriedade assim lhe dá respaldo, cabe destacar que o fato de o formato de arquivo ser proprietário não significa que ele não pode ser aberto, uma vez que disponibilizar publicamente a especificação do formato também é direito da empresa que o detém. É o caso do formato PDF, é de propriedade da ADOBE, mas cuja especificação está disponível publicamente sem restrições de uso. Já os formatos não proprietários são os que não têm um dono, e, dessa forma, são sempre abertos.

Os padrões abertos podem propiciar maior longevidade dos documentos digitais. As especificações dos padrões fechados podem promover entraves em termos de preservação

de documentos digitais, pois não estão disponíveis publicamente ou estão sujeitos a restrições de uso. Vale ressaltar que, em relação aos formatos de arquivos, eles podem ser divididos em padrões de direito e de padrões de fato. Os formatos de arquivos com padrões de direito são chancelados por órgãos oficiais e por instituições nacionais, como a Associação Brasileira de Normas Técnicas (ABNT), pelas multinacionais, como o Comitê Europeu de Normatização (CEN), e pelas instituições internacionais, como a Organização Internacional de Padronização (ISO), (SCHAFER, 2013). Como exemplo de formatos de arquivo que são padrões de direito podemos citar o PDF/A e o ODF *Open Document Format* (CONARQ, 2010). Já os formatos de arquivo de fato, esses não têm legitimação de órgão de padronização, mas são frequentemente utilizados por pessoas e instituições, dentre os formatos desse arquivo tem-se o PDF, DOC, TIFF e o ZIP (SCHAFER, 2013). Os formatos de arquivos de padrões abertos são os mais indicados para preservação dos documentos digitais. A escolha de tais padrões faz parte das políticas de preservação utilizadas pelas instituições, um dos procedimentos adotados por tais políticas é a escolha dos formatos de arquivos recomendados para a preservação digital (FLORES; SANTOS, 2015). Sendo que, preferencialmente, recomendam-se os formatos de arquivos de direito. A escolha de padrões abertos de acordo com Flores e Santos (2018, p. 44) compreende que as “licenças de uso claramente definidas simplificará o processo de preservação no que tange a utilização de plataformas de *hardware* e *software*, além de reduzir os custos com licenças”. Isso significa que os padrões abertos ou fechados estão baseados em especificações técnicas imprescindíveis ao funcionamento dos mesmos. As especificações estão ligadas diretamente à preservação dos documentos digitais, embora tais especificações sejam de natureza técnica, são elas que explicam para os desenvolvedores de *software* como as sequências de *bits* estão interligadas. E isso é importante também no acesso e recuperação da informação contida em um documento digital.

No contexto dos formatos de arquivo, existem instituições que se preocupam com a questão da obsolescência tecnológica e, para tanto, desenvolveram recomendações de formatos que podem ter uma maior durabilidade na preservação dos documentos digitais. Uma das instituições a propor tais recomendações é o Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT), órgão subordinado ao Ministério Ciência, Tecnologia e Inovação (MCTI). Desde 2002, o IBICT em parceria com Programa *LOCKSS*, da Universidade de

Stanford, desenvolve o programa chamado Rede Cariniana (IBICT, 2017) que é uma rede de serviços de preservação digital em longo prazo de publicações eletrônicas brasileiras com base na análise da infraestrutura tecnológica, padronização e a segurança tecnológica da informação. De acordo com a Rede Cariniana, os formatos de arquivo tipo *Office Open XML*, *Office Word* e *Office Power Point* com extensão (.docx.) têm adequação média para preservação digital. Já os formatos de arquivos tipo A-1 com extensão (.pdf), o PNG com extensão (.png) e o *QuickTime Movie* com extensão(.mov) possuem alta adequação para preservação dos documentos digitais em longo prazo (REDE CARINIANA, 2017). Assim, um formato adequado para preservação tem características específicas como: isento de direitos legais que restrinjam a sua utilização; uma norma emitida por um organismo oficial; documentado abertamente através de uma especificação pública e livre; amplamente usado; lido e escrito por múltiplas plataformas de *software*, incluindo código aberto; não comprimido ou comprimido sem perdas de informação (ARQUIVO P.T, 2017). Por outro lado, há formatos que podem vir a comprometer a preservação dos documentos em ambientes digitais. Basicamente, surgem como antagônicos às funcionalidades acima apresentadas (P.T, 2017), formato proprietário e especificação fechada, de pouco uso, lidos em poucas plataformas de *software*, código-fechado, comprimido com perdas de informação. Portanto, a escolha dos formatos de arquivos deve levar em conta aspectos tecnológicos que apontem para o uso preferencial de padrões abertos, preferencialmente de direito, que favorecem a maior longevidade dos acervos digitais.

2.2 OS CODECS

Para que um documento audiovisual, especificamente, possa ser acessado e visto, ele depende de uma tecnologia capaz de interpretar os dados denominada *codec*. O *codec* é a implementação de uma técnica de compactação e descompactação e é usado com a finalidade de reduzir grandes arquivos de vídeo, de modo que estes possam ser transferidos com relativa facilidade por meio da internet (GRECH, 2003). Para que um documento de áudio ou vídeo seja decifrado, interpretado e acessível é fundamental o uso dos *codecs* (ROBERTSON, 2010). Os *codecs* costumam ser implementados em *software*, dessa forma, caso o *software* do *codec* fique obsoleto, os arquivos que por ele foram

compactados/codificados podem deixar de ser exibidos ou mesmos perdidos, se não houver tecnologia compatível.

Um caso interessante de ser abordado em relação à obsolescência dos *codecs* é o da empresa norte-americana *Sirius Publishing*, criadora do *codec MotionPixels*, e sua obsolescência vertiginosa no período de 1995, período marcado pela passagem do VCD para o DVD. A empresa *Sirius Publishing* foi criadora também do *MovieCD*, um disco digital que só era possível ser lido em computadores por meio de um *codec* próprio chamado *MotionPixels* (LA CARRETTA, 2012). A dependência de um *player* virtual para que o *MovieCD* pudesse ser executado gerou um condicionante que provocou desinteresse por parte dos usuários, pois acarretou na impossibilidade de atualização de formato, tornando-o obsoleto juntamente com o desaparecimento de seu *codec*. Por essas razões, o *MovieCD* foi a primeira mídia da história dos acervos audiovisuais da história a desaparecer por completo. Nada restou do seu *codec*, da página do seu desenvolvedor e do seu produtor. Os poucos filmes que sobraram não podem ser reproduzidos por computadores atuais e nem por qualquer outro (LA CARRETTA, 2012).

Assim, diante do que foi exposto, observa-se a dependência tecnológica dos documentos audiovisuais por conta dos *codecs*. Imaginar que muitos documentos dessa natureza, com valores científicos, históricos, culturais, jurídicos e administrativos, podem ser perdidos por tais circunstâncias reforça ainda mais a necessidade de propor uma maior conscientização da vulnerabilidade desses documentos em ambiente digital.

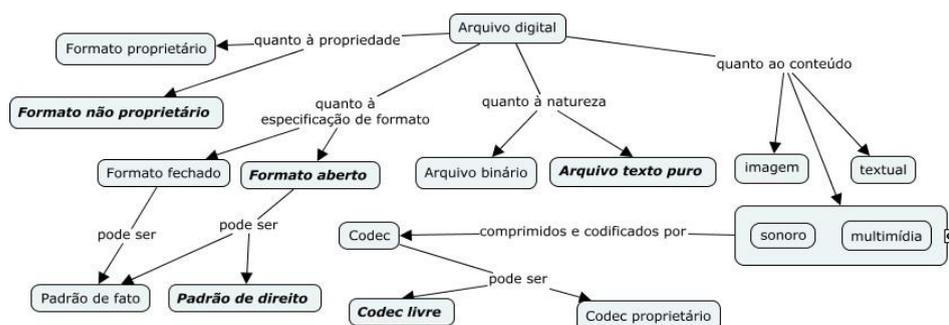
3 ASPECTOS TECNOLÓGICOS RELACIONADOS AOS FORMATOS DE ARQUIVO E A PRESERVAÇÃO DIGITAL

A Câmara Técnica dos Documentos Eletrônicos do Arquivo Nacional (CTDE), por meio do seu Glossário (2016, p. 34), define a noção de preservação digital como sendo um “conjunto de ações gerenciais e técnicas exigidas para superar as mudanças tecnológicas e a fragilidade dos suportes, garantindo o acesso e a interpretação de documentos digitais pelo tempo que for necessário”. Não obstante, os profissionais que trabalham na salvaguarda desses acervos muitas vezes não têm noção da complexidade tecnológica envolvida ou/e não tomam consciência de tais implicações no contexto da preservação digital. Vale destacar que a preservação digital é composta por três grupos de elementos essenciais: organizacional,

legal e técnico (GRÁCIO *et al.*, 2020). Por conta de tal complexidade administrativa é salutar observar os procedimentos corretos e as características fundamentais do documento digital e sua infraestrutura tecnológica. Atentar para as características básicas dos documentos digitais quanto à sua natureza de dependência tecnológica, ou seja, a necessidade de processamento das informações por meio de sistemas computacionais para serem decifrados e lidos envolvendo *software* e *hardware*; a natureza binária; a fragilidade ante as infraestruturas tecnológicas complexas que podem, inclusive, falhar; a vulnerabilidade por conta da rápida obsolescência tecnológica; a natureza dinâmica diferente dos documentos não digitais, em especial, no que diz respeito à rápida disseminação da informação em diversos tipos e formatos de arquivo, dentre outras.

Os documentos digitais em nossa sociedade atual requerem uma atenção especial por conta de sua complexidade tecnológica, em particular, dos formatos de arquivos e os *codecs* que podem ter problemas como as falhas ou obsolescências tecnológicas. Tais complexidades podem comprometer significativamente o acesso a tais acervos, ou os mesmos podem ser perdidos sem que haja recuperação, comprometendo dessa forma o acesso às informações. Nesse cenário, se propõe, conforme explicitado na Figura 2, um conjunto de aspectos a serem observados de forma atenta ao se tratar com a preservação dos documentos digital. Em negrito e *italico* observam-se os aspectos mais favoráveis para a preservação digital, embora, no que diz respeito à natureza do arquivo, entenda-se que na maioria das vezes não será possível usar arquivos de natureza texto puro, devido ao seu conteúdo. Cabe destacar, ainda, que existem outros aspectos da preservação digital, ligados à parte física do arquivo digital, que envolvem o *hardware* e, de forma específica, as mídias digitais, porém esses aspectos estão fora do escopo do presente trabalho.

Figura 2 – Aspectos do contexto lógico do documento digital



Fonte: elaboração própria.

Conforme se pode sintetizar, um documento digital, do ponto de vista da natureza do conteúdo, pode ser binário ou textual puro. Do ponto de vista do tipo do conteúdo, pode conter texto, imagem, som ou multimídia. Documentos digitais de som ou vídeo possuem dependência de *codecs* (*software*). Cada documento digital possui uma especificação de formato. Da perspectiva da especificação do formato, os arquivos digitais podem ter especificação de formato aberta – sendo denominados formatos abertos, podendo ser “de fato” e “de direito” – ou fechada – sendo denominados formatos fechados que podem ser apenas “de fato”.

Do ponto de vista da propriedade do arquivo, eles podem ser proprietários ou não proprietários. Lembrando que um formato de arquivo proprietário pode ser aberto ou fechado, enquanto os formatos não proprietários são sempre abertos. Quanto aos *codecs*, eles são *softwares* necessários para a compressão/descompressão e codificação do conteúdo de arquivos de som e/ou vídeo, e como *software* pode ser *software* livre, cuja fonte está disponível publicamente, ou podem ser proprietários, sem que seu código-fonte esteja disponível ao público. O uso de *codecs* livres e de formatos abertos e não proprietários é a opção mais aconselhada para a preservação dos documentos digitais.

Observando o mapa conceitual, Figura 2 exposta acima, podemos evidenciar que a relação de vulnerabilidade de um documento digital vai além da aplicação de uma mera reparação de natureza técnica. Ela expõe a complexidade inerente da própria natureza peculiar do documento digital como, por exemplo, em relação à condição a binária, aos formatos de arquivos e aos *codecs* formando o contexto tecnológico que deve ser considerado em sua plenitude. A falta de compreensão desse contexto pode acarretar decisões equivocadas na preservação dos acervos digitais.

Nesse sentido, das contribuições acerca do trabalho pode-se destacar uma visão geral do contexto lógico do documento digital para o profissional da informação ter conscientização acerca das escolhas dos formatos de arquivos, bem como da questão da dependência dos *codecs* para conteúdo sonoro e multimídia para serem adotados na salvaguarda dos documentos digitais dentro do âmbito das políticas de preservação. O profissional que trabalha na salvaguarda dos acervos digitais também necessita compreender essas funcionalidades para que não haja, inclusive, o manuseio tecnológico

inadequado, o que pode resultar em perdas dos documentos em ambiente digital. É necessário, portanto, um conhecimento mínimo desse contexto para trabalhar na preservação desses acervos. A conscientização dos aspectos tecnológicos que envolvem os documentos digitais constitui uns dos requisitos fundamentais que os profissionais precisam ter no âmbito das políticas de preservação de documentos digitais em longo prazo.

4 CONSIDERAÇÕES FINAIS

Observam-se, nas últimas décadas, profundas transformações por conta dos avanços tecnológicos. É inegável, por conseguinte, afirmar que a produção desses documentos digitais inevitavelmente depende de um computador, de um *software* e demais infraestruturas tecnológicas para que o documento digital seja interpretado. Com o advento da internet, houve uma modificação na forma de comunicar, de produzir, de armazenar, de transmitir e de acessar tais documentos nas instituições. Nesse sentido, o trabalho teve como proposta explicitar a natureza complexa do documento digital, em especial dos formatos de arquivos e os *codecs*. A necessidade de uma maior conscientização acerca da natureza dos documentos digitais no contexto tecnológico, dos manuseios adequados dessas tecnologias e a manutenção constante dos acervos digitais, perpassa por toda gestão de documentos e das políticas de preservação digital de uma instituição. O profissional da informação, nesse sentido, tem papel fundamental na salvaguarda desses acervos tecnológicos para as futuras gerações. É necessário, igualmente, ter uma conscientização de que a preservação digital não é algo à parte da gestão, envolve todos os aspectos institucionais. Muitos profissionais podem não estar cientes que a não observância dessas vulnerabilidades tecnológicas podem propiciar perdas imensuráveis ao documento digital. Nesse sentido, propõem-se um maior debate acerca do envolvimento dos profissionais da informação sobre o tema, bem como aquisição de conhecimentos mais aprofundados sobre preservação digital, sobre os formatos de arquivos e sobre os *codecs* por meio da qualificação profissional, pois muitas vezes esses profissionais tem contato com os acervos digitais e sua complexidade e necessitam de uma formação mais específica para lidar com os mesmos.

REFERÊNCIAS

ARQUIVO P.T(portal). **Formatos adequados para preservação**. Arquivo.PT, 2017. Disponível em: <https://arquivo.pt/wayback/20170805182209/http://sobre.arquivo.pt/pt/recomendacoes/formatos-adequados-para-preservacao/>. Acesso em: 21 jun. 2021.

AVANCINI, Marta. A importância da preservação dos acervos digitais: a migração das publicações em papel para os meios digitais é um fato comprovado. **Revista Eletrônica de Jornalismo Científico**, 2012. Disponível em: <http://www.dicyt.com/viewNews.php?newsId=25362>. Acesso em: 24 jun. 2021.

CONSELHO NACIONAL DE ARQUIVOS (CONARQ). **Recomendações para digitalização dos documentos arquivísticos permanentes**. 2010. Disponível em: http://conarq.gov.br/images/publicacoes_textos/Recomendacoes_digitalizacao_completa.pdf. Acesso em: 18 jun.2021.

CONSELHO NACIONAL DE ARQUIVOS (CONARQ). **Glossário**. Documentos Arquivísticos Digitais. Arquivo Nacional. Rio de Janeiro, 2014. 48 p.

CONSELHO NACIONAL DE ARQUIVOS (CONARQ). **Glossário**. Documentos Arquivísticos Digitais. Arquivo Nacional. Rio de Janeiro, 2016. Disponível em: https://www.gov.br/conarq/pt-br/assuntos/camaras-tecnicas-setoriais-inativas/camara-tecnica-de-documentos-eletronicos-ctde/2016_CTDE_Glossario_V7.pdf. Acesso em: 12 jun. 2021.

BARDIN, Laurence. **Análise de conteúdo**. Lisboa: Edições 70, 1995. 229 p.

BODÊ, Ernesto Carlos. Documento digital e preservação digital: algumas considerações conceituais. **Revista Ibero-Americana de Ciência da Informação**, v. 9, n. 2, p. 503-516, 2016. Disponível em: [10.26512/rici.v9.n2.2016.2425](https://doi.org/10.26512/rici.v9.n2.2016.2425). Acesso em: 21 jun. 2021.

BODÊ, Ernesto Carlos. **Preservação de documentos digitais**: o papel de formatos de arquivos. 2008. 153 f. Dissertação (Mestrado em Ciência da Informação), Departamento de Ciência da Informação e Documentação, Universidade de Brasília, Brasília, 2008. Disponível em: <https://repositorio.unb.br/handle/10482/2034>. Acesso em: 18 jun. 2021.

FERREIRA, Miguel. **Introdução à Preservação Digital**: conceitos, estratégias e actuais consensos. Guimarães, Portugal: Escola de Engenharia da Universidade do Minho, 2006. 88 p. Disponível em: <https://repositorium.sdum.uminho.pt/bitstream/1822/5820/1/livro.pdf>. Acesso em: 06 jun. 2021.

FLORES, Daniel; SANTOS, Henrique Machado dos. As vulnerabilidades dos documentos digitais: obsolescência tecnológica e ausência de políticas e práticas de preservação digital. **Biblios**, n. 59, p. 45-54, 2015.

FLORES, Daniel; SANTOS, Henrique Machado dos. A obsolescência do conhecimento em preservação digital. **Ciência da Informação em Revista**, v. 5, n. 1, p. 41-58, 2018.

GLOSSÁRIO OPEN DATA HANDBOOK. Disponível em: <https://opendatahandbook.org/glossary/en/terms/proprietary/>. Acesso em: 26 jun. 2021.

GRÁCIO, José Carlos Abbud; FADEL, Bárbara; VALETIM, Marta Ligia Pomim. Preservação digital nas instituições de ensino superior: aspectos organizacionais, legais e técnicos.

Perspectivas em Ciência da Informação, Minas Gerais, n. 3. p. 1, jul./set. 2013.

GRÁCIO, José Carlos Abbud; TROITIÑO, Sonia; MADIO, Telma Campanha de Carvalho; BREGA, José Remo Ferreira; MORAES, Maria Blassioli. Modelo para elaboração de políticas de preservação digital de documentos de arquivo por instituições de ensino superior: o caso da Unesp. **Revista Eletrônica de Comunicação, Informação e Inovação em saúde**. Disponível em: <https://doi.org/10.29397/reciis.v14i3.2111>. Acesso em: 19 ago. 2021.

GRECH, Victor. Codecs e descompressores de vídeo, **Pediatric. Cardiol**, 2003. Disponível em: <https://europepmc.org/articles/PMC3232547?sessionid=11B93D431391BC0CF04E2EDFBCBC780C>. Acesso em: 20 jun. 2021.

INNARELLI, Humberto Celeste. Preservação digital: a influência da gestão dos documentos digitais na preservação da informação e da cultura. **Revista Digital de Biblioteconomia & Ciência da Informação**, v. 9, n. 1, p. 72-87, 2011. Disponível em:

[10.20396/rdbci.v8i2.1934](https://doi.org/10.20396/rdbci.v8i2.1934). Acesso em: 08 jun. 2021.

INTERPARES 2 PROJECT. **Diretrizes do Produtor**: a elaboração e a manutenção de materiais digitais: diretrizes para indivíduos. TEAM Brasil. Tradução: Arquivo Nacional e Câmara dos Deputados, 2007. Disponível em: http://www.interpares.org/ip2/display_file.cfm?doc=ip2_creator_guidelines_booklet--portuguese.pdf. Acesso em: 22 jun. 2021.

INTERPARES 3 PROJECT. **Brazil Dictionary**. TEAM Brasil. Tradução: Arquivo Nacional, 2012. Disponível em: http://www.interpares.org/ip3/ip3_terminology_db.cfm?team=4&status=dictionary. Acesso em: 22 jun. 2021.

INSTITUTO BRASILEIRO DE INFORMAÇÃO EM CIÊNCIA E TECNOLOGIA (IBICT). **Rede Cariniana**, Instituto Brasileiro em Ciência e Tecnologia, 2017.

LA CARRETTA, Enrique López da Cunha Pereira Marcelo. **Prosumidores Nostálgicos**: reflexões sobre a obsolescência das mídias audiovisuais. 2012. 233 f. Tese (Doutorado) – Escola de Belas Artes, Universidade Federal de Minas Gerais, Belo Horizonte, 2012.

MOURA. Êda Maria Bastos de; CAMPOS, Linair Maria. A preservação dos documentos históricos em ambientes digitais. *Revista Brasileira de Preservação Digital*. 2020. Disponível em: <https://econtents.bc.unicamp.br/inpec/index.php/rebpred/article/view/13858/9820>. Acesso em: 05 jun. 2021.

MOURA, Êda Maria Bastos de. O contexto tecnológico da preservação dos documentos digitais: os prejuízos e as propostas de ações institucionais na salvaguarda do patrimônio documental digital. 2021. 157 f. Dissertação (Mestrado em Ciência da Informação) – Universidade Federal Fluminense, 2021. Disponível em: <https://app.uff.br/riuff/handle/1/22199>. Acesso em: 05 jun. 2021.

ROBERTSON, Mark R. **Video File Container Formats, Compression and Codecs**. Tubular Insights, 2010. Disponível em: <https://tubularinsights.com/file-formats-containers-compression/>. Acesso em: 14 jan. 2021.

RONDINELLI, Rosely Curi. **O documento arquivístico ante a realidade digital**: uma revisão conceitual necessária. Rio de Janeiro: FGV, 2013. 280 p.

SAYÃO, Luis Fernando. Uma outra face dos metadados: informações para a gestão da preservação digital. **Revista Eletrônica de Biblioteconomia e Ciência da Informação**, Florianópolis, v. 15, n. 30, p. 1-13, 2010. Disponível em: <https://periodicos.ufsc.br/index.php/eb/article/view/1518-2924.2010v15n30p1/19527>. Acesso em: 20 jun. 2021.

SCHAFER, Murilo Billig. Digitalização de documentos: implicações no acesso às informações arquivísticas. Dissertação (Mestrado)-Universidade de Santa Maria, Centro de Ciências Sociais e Humanas, Programa Profissionalizante de Patrimônio Cultural. RS, 2013.

UC CURATION CENTER. Unified Digital Format Registry (UDFR): **Final Report**. Digital Library University of California, 2012. 26 p. Disponível em: <http://udfr.org/project/UDFR-final-report.pdf>. Acesso em: 06 jun. 2021.

UNIVERSITY AMSTERDAM. **Digital Humanities Workbench**. 2020. Disponível em: https://www2.fgw.vu.nl/dighum/digital_data/digital_text/dt-introduction.php. Acesso em: 18 jun. 2021.

YAMAOKA, E. J.; GAUTHIER, F. A. O. Ontologia de dependência tecnológica de documentos digitais: instrumento de apoio à preservação digital. **Revista Eletrônica de Biblioteconomia e Ciência da Informação**, v. 17, n. esp., p. 211-226, 2012. Disponível em: [10.5007/1518-2924.2012v17nesp2p211](https://periodicos.ufsc.br/index.php/eb/article/view/10.5007/1518-2924.2012v17nesp2p211). Acesso em: 18 jun. 2021.