

PÁGINAS WEB: PRESERVAÇÃO DIGITAL E O PAPEL DO 'INTERNET ARCHIVE' PARA A MEMÓRIA DIGITAL/VIRTUAL¹

Douglas de Souza Caldas

RESUMO: O tema deste artigo é a preservação digital de informação *online* a partir de uma análise sobre a ferramenta “WayBack Machine” (WM) do site *web.archive.org*. O problema da pesquisa pode ser definido com a seguinte pergunta: O Wayback Machine contribui para as discussões sobre memória virtual no que dispõem as teorias arquivísticas? O objetivo geral foi o de pesquisar as características arquivísticas do Wayback Machine do Internet Archive, no tocante à sua relação com a teoria arquivística no contexto digital. Os objetivos específicos foram: problematizar o conceito de documento de arquivo no contexto digital *online*; e analisar as contribuições da página ‘archive.org’ relacionando-a aos conceitos de preservação da memória digital/virtual. Metodologicamente, trata-se de uma pesquisa básica, qualitativa e exploratória, com procedimentos técnicos de pesquisa bibliográfica e documental. Os resultados apurados pela pesquisa demonstram que, embora o *WayBack Machine* não possa ser considerado um repositório digital confiável por não possuir a cadeia de custódia da informação *online*, ele possui relevância para a preservação do patrimônio virtual. Além disso, o Internet Archive potencializa a constituição da memória virtual, uma vez que cria coleções que salvaguardam cópias das páginas publicadas na *web*.

Palavras-chave: preservação digital, ciberespaço, páginas da web

ABSTRACT: The theme of this article is the digital preservation of online information based on an analysis of the “WayBack Machine” (WM) tool from *web.archive.org*. The research problem can be defined with the following question: Does the WayBack Machine contribute to discussions about virtual memory in terms of archival theories? The general objective was to research the archival characteristics of the Internet Archive WayBack Machine, regarding its relationship with archival theory in the digital context. The specific objectives were to problematize the concept of archival document in the online digital context; and to analyze the contributions of the page 'archive.org' relating it to the concepts of digital/virtual memory preservation. Methodologically, it is a basic, qualitative, and exploratory research, with technical procedures of bibliographic and documental research. The results obtained by the research demonstrate that, although the *WayBack Machine* cannot be considered a reliable digital repository because it does not have the chain of custody of online information, it has relevance for the preservation of virtual heritage. In addition, the Internet Archive enhances the constitution of virtual memory, as it creates collections that safeguard copies of pages published on the web.

Keywords: digital preservation; cyberspace; web pages

1 INTRODUÇÃO

Levando em conta o contexto informacional atual, percebe-se que muitos dos serviços essenciais são administrados através da *internet*. Não somente isso, mas muita informação é administrada e compartilhada através de *sites* e redes sociais. Tendo isso em mente, este artigo apresenta resultados de uma pesquisa sobre a preservação da informação digital e de que forma pode-se garantir sua acessibilidade e permanência na rede mundial de computadores.

¹ Trabalho de Conclusão de Curso. Graduação em Arquivologia. Universidade Federal de Santa Catarina. 2022. Orientação: Prof. Cezar Karpinski, Dr.

Assim, o tema da pesquisa é a preservação digital de informação *online*. Já o objeto de foi a ferramenta “WayBack Machine” (WM) do site *web.archive.org* e suas potencialidades como repositório de páginas da *web*. O contexto digital atual, produz e dissemina muita informação através de *links* (endereços *web*) e páginas (documentos apresentados pelos navegadores). Entretanto, não se percebem políticas que padronizam ou se preocupam com o tratamento desse tipo de informação, de sua preservação e como realizar a custódia das páginas *web*, pensando-os como um documento arquivístico.

Neste sentido, considerando as especificidades do objeto e das várias possibilidades de investigação, a pesquisa se deteve na seguinte pergunta: O Wayback Machine contribui para as discussões sobre memória virtual no que dispõem as teorias arquivísticas? Para tanto, o objetivo geral foi o de pesquisar as características arquivísticas do Wayback Machine do Internet Archive, no tocante a sua relação com a teoria arquivística do contexto digital. Entende-se que essa reflexão pode contribuir para as discussões acerca do tratamento que deve ser dado às páginas *web* ou, em última instância, se essa tipologia documental se constitui como arquivística, potencializando o campo de análise para a memória virtual.

Os objetivos específicos foram: entender o conceito de documento de arquivo no contexto digital *online*; e analisar as contribuições da página ‘archive.org’ relacionando-a aos conceitos de preservação da memória digital/virtual. No primeiro objetivo, questiona-se esse ‘novo’ paradigma digital, uma vez que grande parte do conteúdo informacional é consumido no ambiente *online*. Já o segundo se estabelece como forma de compreender uma ferramenta que se coloca como salvaguarda da memória no ambiente *online*.

A bibliografia da área de Arquivologia tem indicado que ferramentas de recuperação de páginas da *web* têm um papel importante para a constituição social da memória digital e para a preservação da informação no meio digital. No Brasil, poucas pesquisas têm se debruçado sobre o assunto no âmbito arquivístico, demonstrando a relevância dessa temática em território nacional, uma vez que este país é um grande consumidor de informação em meio digital. Junte-se a isso a preocupação com as notícias falsas, especialmente em 2022, ano eleitoral no Brasil, onde os olhos dos cidadãos se voltarão às campanhas políticas, muitas das quais feitas total ou parcialmente *online*. Nesse sentido, investigações sobre a preservação digital de informações na *web* são de relevância para a sociedade.

Um exemplo de pesquisa sobre preservação de páginas de *internet* ocorre na Catalunha, Espanha. Trata-se do projeto *Padcat* que faz arquivamento de documentação produzida por perceberem o problema de não salvaguardar o que se produz no ambiente *web*.

No sistema democrático espanhol, as administrações públicas podem renovar seus líderes políticos a cada quatro anos. A explosão das TIC na política transformou a atividade dos eleitos e dos representantes políticos e sociais porque o uso das tecnologias pelos partidos políticos acrescentou suas estratégias de comunicação, baseadas essencialmente nos impactos na mídia tradicional. A utilização de páginas pessoais e blogs, a presença generalizada nas redes sociais, a utilização de vídeos e aplicações multimídia, etc. levou a uma situação na qual uma parte importante das ações de comunicação política durante uma campanha eleitoral se resolvem na internet, exclusivamente em formato digital: faz parte da filosofia 2.0 aplicada à política. (LLUECA et. al, 2006, p. 310311, tradução nossa)

Esses dispositivos de captura de páginas, tanto o *web archive* quanto o *padcat* trazem a reflexão sobre o arquivamento de páginas da internet, demonstrando assim que, apesar de cada um ter seu escopo (o *padcat* se preocupa com campanhas eleitorais e o Wayback Machine com entradas de páginas da internet em geral), são respostas ao arquivamento de páginas que não ocorre atualmente, mesmo que este seja entendido como um documento de arquivo, já que páginas web são registros de atividades em rede e que devem ser preservados (FLORES, 2021)

Pessoalmente, a pesquisa se tornou instigante por meio de questionamentos advindos de leituras e seminários que foram presenciados pelo autor. Primeiramente, na rede de *microblogs Twitter*, onde o professor e editor do Intercept Brasil, Leandro Demori, problematiza a linguagem burocrática utilizada pelos governos. Em uma postagem de 16 de agosto de 2021, Demori publicou um *post* com o seguinte título “Como transformar duas frases em um textão em 5 passos” (título na íntegra) disponibilizado diretamente na página do Governo Federal (BRASIL, 2021), onde, devido a sua exposição e repercussão midiática, o conteúdo foi removido instantes após sua publicação. No entanto, Demori, utilizando da ferramenta Wayback Machine, recuperou a informação em questão, trazendo-a como crítica em seu microblog pessoal, reacendendo a discussão sobre como transformar simples frases em grandes textos. Apesar de a página ter sido removida do site ‘original’, o WayBack Machine já havia salvado uma cópia, mostrando o poder que esta ferramenta tem para a sociedade, a fim de fiscalizar informações públicas, por exemplo.

Um segundo momento de reflexão sobre o assunto adveio de um seminário sobre preservação de páginas da *web* como documento de arquivo, ministrado pelo prof. Daniel Flores. Nesta primeira escuta, foi possível problematizar alguns conceitos e práticas voltadas à preservação de documento arquivístico digital, sua disponibilização na *web* e as relações deste

tema com o Wayback Machine. Essas primeiras incursões instigaram o autor na busca por compreensões teóricas acerca das funções da ferramenta e o contexto arquivístico dos documentos digitais publicados na *web*.

Este artigo está estruturado de forma objetiva e simplificada, contendo esta introdução, o referencial teórico com material bibliográfico a respeito do tema, os procedimentos metodológicos que permitiram a análise e elaboração do material, encerrando com os resultados obtidos da pesquisa, ou seja, o que essa pesquisa elucidou através de análise.

2 REFERENCIAL TEÓRICO

O advento das tecnologias de comunicação e informação, mais especificamente as que se desenvolvem ao longo do Século XXI, vem moldando a forma como a sociedade produz e consome informação. Essa realidade também se aplica, especificamente, no âmbito dos documentos de arquivo. Isso porque os documentos, principalmente nos últimos dez anos, são produzidos ou reformatados para o meio digital, mudando a dinâmica de acesso à informação.

Como consequência, o estudo dessa documentação também apresenta singularidades nos últimos anos, em especial as discussões sobre a memória no campo do digital e virtual. Este debate leva em consideração que o meio principal de consumo e disseminação de conteúdo informacional, e mais precisamente documentos, se dá através de computadores conectados à internet. Esse fato chama a atenção sobre o fenômeno da memória virtual, que começou a ocorrer nos anos 1990, em países mais desenvolvidos, e nos anos 2000 para o resto do mundo. Neste aspecto, Dodebei (2009) chama a atenção à evolução tecnológica ocorrida nos suportes informacionais:

Passados cinco séculos, a escrita digital que ocorre no mundo virtualizado da rede mundial de computadores pode ser compreendida por dois objetos: os blogs e os websites. De acordo com Lopes (2003), Blog é uma abreviatura de weblog e significa uma página na web que se propõe a ser um diário pessoal com acesso público. Os blogs são, assim, publicações frequentes, organizadas cronologicamente, de considerações pessoais e links para a web. Diferem-se dos websites por sua estrutura conversacional. Enquanto nos blogs as matérias são “postadas” (posts) por data e esses posts constituem toda a sua essência, os websites harmonizam uma rede de informações fixas, sobre um tema específico, que pode ter como autoria uma pessoa física ou uma pessoa jurídica (DODEBEI, 2009, p. 139).

Esses dois objetos que a autora cita e que existem há pelo menos duas décadas, demonstram o fluxo informacional que ocorre *online*, uma vez que “a informação circula em

rede ou vias de comunicação, sempre transformando outras e a si mesma, num processo de criação contínua, como ocorre com nossa memória individual” (DODEBEI, 2009, p. 139).

Contudo, essa informação não está sendo tratada, problematizada e, por que não dizer, patrimonializada. Durante todo esse tempo em rede, não se constituíram políticas ou sistemas que pudessem fazer um trabalho de tratamento e recuperação dessa informação de forma a disponibilizá-la, assim “[...] porque a evolução da tecnologia e a cultura digital tem sido tão rápida que os governos e instituições não puderam elaborar estratégias de preservação oportunas e bem fundamentadas” (LLUECA et. al., 2006, p. 311, tradução nossa).

A complexidade que tange o documento de página da web também diz respeito ao conceito de memória virtual. Lévy (2003) explica que a virtualização rompeu com o espaço-tempo, onde “quando uma pessoa, uma coletividade, um ato, uma informação se virtualizam, eles se tornam ‘não presentes’, se desterritorializam. Uma espécie de desengate os separa do espaço físico ou geográfico ordinários e da temporalidade do relógio e do calendário” (LÉVY, 2003, p. 21). Essa ocorrência se deve ao fato de que, no âmbito virtual, a informação não fica detenta a um único espaço, tempo ou suporte, ela circula e se altera conforme persiste no ambiente virtual.

Cabe colocar que não se tem um conceito definido e completo sobre a memória virtual, ou seja, é uma discussão em andamento. Embora, na arquivologia já se tenha definido o que é um documento digital, sendo natodigital ou digitalizado, sua criação, sua guarda, ou seja sua cadeia de custódia fazem desse tipo de documento, um documento de arquivo. Contudo

A diferença entre digital e virtual está diretamente vinculada ao processo, no caso do primeiro atributo, e ao meio ou ambiente, no caso do segundo. Dessa forma, podem existir objetos digitalizados que habitam tanto o mundo concreto quanto o mundo virtual, mas este é habitado apenas por objetos digitais. (DODEBEI, 2008, p. 28)

E isso se dá porque o documento virtual, ele existe e embora não se conceitue sua ideia de acumulação, como nos documentos de suporte analógico, pode-se ver o patrimônio virtual como coleções, pois

Textos, imagens e sons organizados, como num recorte enciclopédico, podem ser acessados em tempo real por um número cada vez mais amplo de internautas, que se apropriam, reformatam e devolvem ao ciberespaço novas informações. O sentido de coleção pode estar presente no ciberespaço, e não sabemos ainda se o atributo da acumulação lhe será inerente. (DODEBEI, 2008, p. 30)

Esse conceito apresentado por Dodebei, ressalta o tema desta pesquisa, uma vez que, independentemente do contexto arquivístico dos documentos disponibilizados *online*, o

WayBack Machine tem potencializado a preservação de um tipo específico de patrimônio, o virtual. Isso porque captura e salva, em seu repositório, as *páginas web*.

O conceito de patrimônio digital, na área de Arquivologia, agregou várias perspectivas presentes nas ciências sociais e humanas do século XX e início do século XXI. Em contrapartida o patrimônio virtual, mesmo englobando o conceito de patrimônio digital, não está inserido num espaço-tempo específico. Contudo, o acesso a esse meio, cada dia mais usual no ciberespaço, insere seus usuários num contexto distinto, pois, ao mesmo tempo em que não se encontra nem no espaço nem em um tempo específico, está em todos os lugares e em qualquer tempo, para o acesso de quem possua conexão com a internet.

3 PROCEDIMENTOS METODOLÓGICOS

De acordo com a bibliografia levantada, para obter os resultados esperados, foi realizada uma pesquisa básica, na qual se “objetiva gerar conhecimentos novos úteis para o avanço da ciência sem aplicação prática prevista” (MENEZES, 2009, p. 15). No tocante à abordagem do problema, a pesquisa é qualitativa, que “considera que há uma relação dinâmica entre o mundo real e o sujeito [...]” (MENEZES, 2009, p. 16).

Quanto aos objetivos, a pesquisa é exploratória, pois “visa proporcionar maior familiaridade com o problema vistas a torná-lo explícito ou a construir hipóteses” (MENEZES, 2009, p. 16). Do ponto de vista dos procedimentos técnicos, a pesquisa utilizada é a documental “[...] elaborada a partir de materiais que não receberam tratamento analítico” (MENEZES, 2009, p. 17) uma vez que entre os documentos levantados há documentação que não foi publicada, como postagens na rede de microblogs *Twitter* e um vídeo de um seminário postado na rede social de vídeos *YouTube*. A pesquisa ainda é bibliográfica pois foi “[...] elaborada a partir de material já publicado, constituído principalmente de livros, artigos de periódicos e, atualmente, com material disponibilizado na Internet.” (MENEZES, 2009, p. 17).

Como procedimentos técnicos, foram utilizadas as pesquisas bibliográfica e documental. Para a primeira, a fonte de informação foi o artigo científico, desde que tivesse relação específica com o tema, coletados das seguintes bases de dados: Base de Dados Referenciais de Artigos de Periódicos em Ciência da Informação (BRAPCI) ; Biblioteca Eletrônica Científica Online ou Scientific Electronic Library Online (SCIELO); GOOGLE SCHOLAR; e Web of Science (WoS). A WoS foi a que mais contribuiu no levantamento de fontes. Já a pesquisa documental se desenvolveu sobre a ferramenta WM da página web Internet

Archive, identificando suas especificidades, convergências ou divergências ao que explica a literatura acerca do patrimônio digital e memória virtual.

O instrumento técnico utilizado foi o fichamento de fontes de informação que poderiam de alguma forma contribuir para o desenvolvimento da pesquisa, para isso, foi elaborado um documento no Software Word, da Microsoft, cujo arquivo foi armazenado em nuvem no Google Drive, para que da mesma forma que fosse levantado esses fichamentos, o orientador pudesse verificar o andamento da pesquisa.

Por se tratar de uma temática atual, transversal e interdisciplinar, a preservação de páginas *web* tem ganhado destaque em fóruns de comunicação informal na *internet* como em blogs, sites de notícias e redes sociais, especialmente em palestras disponíveis no Youtube. Neste aspecto, para esta pesquisa foi relevante um vídeo de um seminário postado na página do YouTube do Professor Doutor Daniel Flores (FLORES, 2021) e uma postagem na rede de microblogs Twitter, na conta do jornalista e professor, Leandro Demori, do *The Intercept* Brasil (DEMORI, 2021)

Com a finalidade de sistematizar os procedimentos metodológicos da pesquisa, apresenta-se o Quadro 1, destacando os objetivos, metodologias e procedimentos técnicos da pesquisa.

Quadro 1. Relação entre objetivos, metodologia e procedimentos técnicos da pesquisa

OBJETIVOS	METODOLOGIA	MATERIAL UTILIZADO
<p>Pesquisar as características arquivísticas do Wayback Machine do Internet Archive no tocante à sua relação com a teoria arquivística no contexto digital.</p>	<p>Pesquisa Bibliográfica em bases de dados. BRAPCI, SCIELO, GOOGLE SCHOLAR e WEB OF SCIENCE Pesquisa documental através de vídeos no Youtube e postagem no Twiter</p>	<p>BRASIL. Como transformar duas frases em um textão. Brasília, 2021. Disponível em https://web.archive.org/web/20210816164901/https://www.gov.br/pt-br/guia-de-edicao-de-servicos-do-gov.br/como-transformar-duas-frases-em-um-textao-em-5-passos. Acesso em: 30 jun 2022.</p> <p>DEMORI, Leandro. (@demori) “Acabaram de tirar do ar. A versão arquivada está aqui”. 16 de agosto de 2021. 15:57. Tweet. Acesso em: 24 fev 2022 Disponível em: https://twitter.com/demori/status/1427336179230388228</p> <p>FLORES, Daniel. Preservação de páginas web e redes sociais em cadeia de custódia: identificação seleção arquivamento. Videoconferência, son., color (29 min.). Brasília: Conselho Nacional de Justiça, 2021. Disponível em: https://www.youtube.com/watch?v=pOaQUo3PH-A. Acesso em: 25 fev. 2022.</p>
<p>Entender o conceito de documento de arquivo digital no âmbito das páginas de internet;</p>	<p>Pesquisa Bibliográfica em bases de dados. BRAPCI, SCIELO, GOOGLE SCHOLAR e WEB OF SCIENCE Pesquisa documental no em vídeo no Youtube</p>	<p>BELLOTTO, H. L. A diplomática como chave da teoria arquivística. Archeion Online, v. 3, n. 2, 2015. Disponível em: http://hdl.handle.net/20.500.11959/brapci/14906. Acesso em: 27 fev. 2022.</p> <p>DODEBEI, V. L. Digital virtual: o patrimônio do Século XXI. In: DODEBEI, V. L.; ABREU, R. (Orgs.). E o patrimônio? Rio de Janeiro: Contracapa : Programa de Pós-Graduação em Memória Social da Universidade do Estado do Rio de Janeiro, 2008. p.11-32.</p> <p>DODEBEI, V. L. Cultura digital: novo sentido e significado de documento para a memória social?. DataGramZero, v. 12, n. 2, 2011. Disponível em: http://hdl.handle.net/20.500.11959/brapci/7335. Acesso em: 30 jun. 2022.</p> <p>YEO, Geoffrey. Falando sobre diferença: Percepções sobre unicidade e identidade nos arquivos. In: Arq. & Adm., Rio de Janeiro, v.10, n.1, jan./jun. 2011</p> <p>ISOGLIO, Antonela; VIGNA, Diego Germán. Desafios teóricos y metodológicos en torno al documento de archivo como objeto digital en la red. Investig. bibl., Ciudad de México, v. 35, n. 87, p. 129-150, jun. 2021. Disponible en http://www.scielo.org.mx/scielo.php?script=sci_arttext&id=S0187-358X2021000200129&lng=es&nrm=iso. Acesso em: 14 nov. 2021.</p> <p>LÉVY, Pierre. Cibercultura. Paulo Neves (Trad.) São Paulo: Ed. 34, 2005</p>

<p>Analisar as contribuições do ‘archive.org’ para a preservação da memória digital.</p>	<p>Pesquisa Bibliográfica em bases de dados. BRAPCI, SCIELO, GOOGLE SCHOLAR e WEB OF SCIENCE</p>	<p>INTERNET ARCHIVE. Internet Archive: Digital Library of Free and Borrowable Books, Movies, Music & WayBack Machine. 2022. Disponível em: https://www.archive.org. Acesso em: 20 ABR 2022</p> <p>MURPHY, Jamie; HASHIM, Noor Hazarina; O’CONNOR, Peter. Take me back: validating the wayback machine. Journal of Computer-Mediated Communication, v. 13, n. 1, p. 60-75, 2007. https://academic-oup-com.ez46.periodicos.capes.gov.br/jcmc/article/13/1/60/4583038</p>
--	--	--

Fonte: O autor (2022)

4 RESULTADOS

A partir dos dados coletados pela pesquisa é possível realizar algumas reflexões a respeito do objeto de estudo, o Wayback Machine, com o propósito de entender essa ferramenta como um repositório virtual, contribuindo, dessa forma, com a preservação do patrimônio digital virtual.

Para que sejam criadas estratégias de preservação, é importante entender o aspecto do documento de página da web, como assim cita Flores (2021):

o arquivamento web ou arquivamento da internet é o processo de coleta e preservação de porções da internet em um arquivo, como um site de arquivamento para futuros pesquisadores, historiadores e ao público interessado em visualizar versões de páginas antigas ou não mais existentes” (FLORES, 2021, 7 min)

No entanto, há que se considerar os aspectos específicos do documento de arquivo neste processo, pois, para que um documento disponível *online* possa ser considerado um documento de arquivo, deverá ser mantida sua cadeia de custódia. Só assim, segundo o mesmo autor, se preserva, junto com o documento, as características arquivísticas que devem obedecer aos princípios da Arquivologia “pois demonstram a origem, a organicidade e a autenticidade” (FLORES, 2021, 10 min 55 seg) dos documentos.

Ressalta-se que, segundo Flores (2021), os documentos web são complexos e específicos (2021, 8 min e 45 seg) e, na conjuntura atual, o Wayback Machine não pode ser considerado um repositório digital arquivístico confiável (RDC-Arq), por não possuir os registros ou metadados da cadeia de custódia do conteúdo documental disponível na *web*.

Entretanto, tendo em vista que há um tratamento documental por conta dos procedimentos de captura e acesso, a comunidade arquivística deve refletir sobre o que se

espera deste tipo de ferramenta. Afinal, o WayBack Machine vem se consolidando como um importante repositório de salvaguarda da memória virtual dos documentos disponibilizados na *web*.

De certa forma, esse debate implica no que já alertava Dodebei (2011, p.4) de que a sociedade precisa rever o conceito de documento, já que no contexto de memória, “o documento não pode representar ao mesmo tempo, tanto a memória instituída quanto a memória em criação” (2011, p. 4), caso que ocorre com o congelamento de páginas web no Wayback Machine: assim que a página é subida em rede e o sistema do Internet Archive o percebe, faz sua cópia. Ou seja, ao mesmo tempo que é criada, é instituída.

Criado em 1996, o projeto “internet archive” funciona hoje como um acervo documental de informações *online*. Na sua página oficial, no *menu* “about” (sobre, em inglês), consta um pouco de seu histórico.

Começamos em 1996 arquivando a própria Internet, um meio que estava começando a crescer em uso. Assim como os jornais, o conteúdo publicado na web era efêmero – mas, diferentemente dos jornais, ninguém o salvava. Hoje temos mais de 25 anos de história da web acessíveis através do Wayback Machine e trabalhamos com mais de 750 bibliotecas e outros parceiros por meio de nosso programa Archive-It para identificar páginas importantes da web. (INTERNET ARCHIVE, 2022, tradução nossa)

O site criado por Brewster Kahle, engenheiro da computação, armazena desde sua criação, copiando as páginas da web em um banco de dados para indexação, como uma captura de tela. O artigo de Isoglio e Vigna (2021) traz importante explicação sobre o que é o Internet Archive e a ferramenta WayBack Machine. Segundo os autores, o projeto

Foi concebido inicialmente como uma iniciativa para que todos os objetos digitais dedicados à produção cultural – sites, filmes, livros, obras musicais, até softwares – que navegam na Web não desaparecessem sem deixar rastros. O método para atingir esse objetivo baseia-se na geração de cópias de segurança – na forma de capturas de tela, como uma fotografia – de cada documento publicado, com base em seu Uniform Resource Locator (URL, na sigla em inglês). Isso é feito por um mecanismo de busca criado por Kahle, chamado Alexa, que copia cada objeto digital em um banco de dados para indexação. Desta forma, o Internet Archive contempla uma arquitetura de milhares de coleções, cada uma correspondendo a um domínio web, que se nutrem de capturas a cada modificação do código HTML que as páginas sofrem. [...] Esses objetos digitais foram organizados no banco de dados de acordo com cada domínio original e rotulados com a referência de tempo de sua captura. Levando em conta, então, que esses objetos são produzidos e circulam online [...]. O que cada captura permite é “congelar” o conteúdo dos documentos em uma entidade fixa e preservada. (ISOGLIO; VIGNA, 2021, p. 140-141, tradução nossa)

Como supracitado, o sistema captura uma imagem, ou seja, uma cópia do site, e armazena no banco de dados, tornando-o acessível online. Nesse aspecto, são importantes as constatações de Flores (2021). Para esse autor, “as páginas web são documentos arquivísticos” (FLORES, 2021, 3 min), porém, no que diz respeito ao Wayback Machine do Internet Archive, esta ferramenta funciona mais como uma coleção (FLORES, 2021, 6 min 05 seg).

Entretanto é sabido que o conceito de unicidade, leva em consideração que uma duplicata, ou seja, a cópia de um original, na ausência deste, pode ser considerada fonte de informação, porque isso depende do contexto e da relação que esse documento faz com o fundo (YEO, 2011, p. 6). Assim, mesmo que o WayBack Machine não seja (ou não funcione como) um repositório digital de páginas da web, ele, de certa forma, cumpre um papel arquivístico importante, preservando informações institucionais, de forma organizada e conjunta.

Levando em consideração que o projeto se preocupa com a preservação dessas páginas web e que muitas dessas páginas já não existem mais, a hipótese da pesquisa é a de ser possível considerar o Internet Archive um banco de dados arquivístico.

Ainda sobre o WM, pode-se dizer que, por meio desta ferramenta,

Os usuários podem visualizar a versão original de cada site, bem como as datas e o conteúdo das atualizações posteriores. Para acessar sites arquivados, os usuários digitam a URL do site desejado na caixa de endereço na página inicial do WM. O WM então retorna a data de criação do site original, número e data das atualizações do site e links para sites arquivados.” (MURPHY, O’CONNOR, 2007, p. 63, tradução nossa)

Tal discurso mostra que a ferramenta possui características capazes de ser analisadas por meio da diplomática contemporânea. Isso porque, uma análise tipológica dos documentos arquivados pela ferramenta pode averiguar a “autenticidade dos documentos por meio do reconhecimento abalizado dos caracteres internos, intermediários e externos que os caracterizam e identificam [...]” (BELLOTO, 2015, p. 4-5). Tais características estão presentes em todo seu conteúdo, como o *endereço web* (o qual direcionou o usuário até o local), a formatação da escrita (linguagem coloquial ou formal), as imagens que compõem a estrutura desta página web, elementos que possibilitam a identificação deste tipo de documento digital.

Além de informações disponibilizadas em páginas de internet, o Wayback Machine, a partir do ano de 2000, abrange também o conteúdo de um meio de informação considerado, pelos próprios responsáveis pelo projeto, como “efêmero”, a televisão. Para tanto, criaram um arquivo (no sentido coletivo da palavra, conjunto de documentos), fazendo a captura televisiva

de notícias do 11 de setembro de 2001, contribuindo assim, desde 2009, para que o público use o acervo como referência citável e compartilhável.

Entretanto, Isoglio e Vigna (2021) levantam uma provocação acerca do projeto no que diz respeito ao seu alinhamento com a teoria arquivística, dizendo que:

[...] o documento arquivado não é a *página web* que se apresenta dinamicamente ao usuário, mas um objeto digital estático, cuja constituição não depende do acesso às fontes originais para informações atualizadas. Da mesma forma, a temporalidade do documento é estabelecida quando se realiza o processo de captura da informação. Além disso, os objetos digitais na rede não têm uma existência diferenciada anterior ao seu arquivamento: é o próprio processo de captura que os transforma em documentos de arquivo [...] (ISOGLIO; VIGNA, 2021, p. 141-142, tradução nossa)

Da citação acima emerge uma importante reflexão sobre o aspecto do documento em meio digital, apresentando a especificidade do documento de página da web: ele não possui uma fase de criação do documento, a partir do momento que ele é capturado, ele automaticamente já é documento de arquivo.

Além disso, se deve levar em conta a problemática em torno das páginas da *web* no contexto atual. As informações inseridas em sites de notícias, as agências de comunicações governamentais, não possuem uma ferramenta que cumpram com o requisito de ser um RDC-Arq, fazendo com que o WM seja a única opção no sentido de salvar as informações online.

Por fim, este artigo é uma iniciativa de problematizar a realidade das páginas da internet no contexto informacional brasileiro. O Brasil está vivenciando na atualidade um momento onde grande parte da documentação, principalmente pública/governamental, está sendo produzida nato-digital e muitas informações de utilidade pública estão sendo inseridas nas *páginas web*, trazendo o questionamento do tratamento que as instituições estão dando para esse tipo de documento, num momento onde a informação está desconstruída, o que faz necessário um plano para a criação de sistemas para a captura desse tipo de informação que atualmente é efêmera.

CONSIDERAÇÕES FINAIS

O presente artigo se propôs a discutir as potencialidades do WayBack Machine como ferramenta para preservação do patrimônio digital produzido *online*: as páginas da *web*. Para isso, foram realizadas pesquisas sobre documentos arquivísticos digitais e sobre memória

virtual e digital. A iniciativa da pesquisa se deu pelo interesse em estudar aspectos da memória no ambiente *web*, uma vez que, no âmbito da área de Arquivologia, o autor não teve contato com muitas publicações sobre a salvaguarda de páginas da internet. Para isso, foi realizada uma pesquisa exploratória e documental, analisando o site onde está hospedado a ferramenta WayBack Machine, o *archive.org*. Assim, procurou-se entender as contribuições dessa ferramenta para a Arquivologia e a memória digital e virtual.

Entende-se que a pesquisa atingiu os resultados esperados. Quanto ao objetivo de pesquisar as características do WayBack Machine no tocante a área de Arquivologia, foi possível perceber que a ferramenta não é um repositório arquivístico. No entanto, analisado sob aspectos das teorias de memória virtual, a ferramenta cumpre seu papel no sentido de preservar o patrimônio no ambiente virtual da internet. Independentemente de seu valor arquivístico, a informação dos arquivos de páginas *web* é preservada e serve como prova em alguns casos específicos, como foi verificado acima.

Já no que diz respeito ao objetivo específico de problematizar o conceito de documento de arquivo digital no âmbito das páginas de internet, entende-se que a pesquisa obteve resultados substanciais. Mostrou-se que não se pode definir ainda com exatidão se uma página da *web* é um documento de arquivo no sentido estrito de seu conceito, pois, para tanto, seria necessário o registro de sua cadeia de custódia, assegurando, assim, sua autenticidade, rastreabilidade etc. Entretanto, o levantamento da pesquisa para responder esse objetivo, traz um questionamento sobre o próprio objeto, o documento de página da *web*, pois mesmo que a ferramenta WayBack Machine não possua um mecanismo de rastreamento da cadeia de custódia, ele cria uma cópia da página publicada, tornando o próprio site da *web* um fundo/coleção. Essa organização classificada das cópias das páginas do *site*, podem torná-lo um fundo documental.

Ainda no que tange o objetivo geral de analisar as contribuições do ‘*archive.org*’ para a preservação da memória digital, a pesquisa apontou que a ferramenta pode contribuir para a preservação da memória virtual. Isso porque, conforme apresentado no artigo, se preserva o patrimônio virtual da *web* que é efêmero. Verificou-se que não há salvaguarda dessas informações de forma a assegurar sua autenticidade, mas a coleção de objetos digitais salvos pode ser vista como uma ferramenta de preservação da memória virtual, desde que exista o acesso ao conteúdo preservado no portal/plataforma.

Cumpridos os objetivos, pode-se afirmar que o Wayback Machine contribui para as discussões sobre memória virtual no que dispõem as teorias arquivísticas se confirmou de forma

parcial. Isso porque a ferramenta não se adequa ao que preconiza a teoria arquivística voltada à preservação de documentos arquivísticos digitais, mas possibilita a problematização do próprio objeto arquivístico disponibilizado na *web*. Contudo, é importante destacar que o WayBack Machine poderia contribuir neste campo se fosse analisado do ponto de vista do princípio da ordem original, onde uma cópia possa cumprir sua função informativa uma vez que o original se perde. Ao analisar a função do WayBack Machine, que é salvar cópias de páginas que não se encontram mais disponíveis, essa cópia age com a função de comprovar a atividade da página. Apesar disso, é relevante destacar que por não possuir registro da cadeia de custódia em seus metadados, automaticamente, é entendido que as cópias de páginas salvas não são consideradas documentos de arquivo para os autores que serviram de base para a pesquisa.

Por fim, essa pesquisa é inicial e espera-se que cumpra seu papel apenas de provocar a discussão a respeito da ferramenta WayBack Machine e das páginas *web*, pois entende-se que é necessária uma discussão mais aprofundada dos conceitos da memória virtual focado às páginas da *web*, tratando-as como documentos arquivísticos. Além disso, com a discussão inicial provocada por este trabalho, espera-se futuramente que ela possa ser aprofundada e revista, para que objetos de pesquisa como esse ganhem maior visibilidade na área de Arquivologia.

REFERÊNCIAS

BELLOTTO, H. L. A diplomática como chave da teoria arquivística. **Archeion Online**, v. 3, n. 2, 2015. Disponível em: <http://hdl.handle.net/20.500.11959/brapci/14906>. Acesso em: 27 fev. 2022.

BRASIL. Como transformar duas frases em um textão. Brasília, 2021. Disponível em <https://web.archive.org/web/20210816164901/https://www.gov.br/pt-br/guia-de-edicao-de-servicos-do-gov.br/como-transformar-duas-frases-em-um-textao-em-5-passos>. Acesso em: 30 jun 2022.

DEMORI, Leandro. (@demori) “Acabaram de tirar do ar. A versão arquivada está aqui”. 16 de agosto de 2021. 15:57. Tweet. Acesso em: 24 fev 2022 Disponível em: <https://twitter.com/demori/status/1427336179230388228>

DODEBEI, V. L. Digital virtual: o patrimônio do Século XXI. In: DODEBEI, V. L.; ABREU, R. (Orgs.). E o patrimônio? Rio de Janeiro: Contracapa : Programa de Pós-Graduação em Memória Social da Universidade do Estado do Rio de Janeiro, 2008. p.11-32.

DODEBEI, V. L. Cultura digital: novo sentido e significado de documento para a memória social?. **DataGramZero**, v. 12, n. 2, 2011. Disponível em: <http://hdl.handle.net/20.500.11959/brapci/7335>. Acesso em: 30 jun. 2022.

FLORES, Daniel. **Preservação de páginas web e redes sociais em cadeia de custódia: identificação seleção arquivamento**. Videoconferência, son., color (29 min.). Brasília: Conselho Nacional de Justiça, 2021. Disponível em: <https://www.youtube.com/watch?v=pOAqUo3PH-A>. Acesso em: 25 fev. 2022.

INTERNET ARCHIVE. Internet Archive: Digital Library of Free and Borrowable Books, Movies, Music & WayBack Machine. 2022. Disponível em: <https://www.archive.org>. Acesso em: 20 ABR 2022

ISOGLIO, Antonela; VIGNA, Diego Germán. Desafíos teóricos y metodológicos en torno al documento de archivo como objeto digital en la red. **Investig. bibl**, Ciudad de México , v. 35, n. 87, p. 129-150, jun. 2021 . Disponible en http://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S0187-358X2021000200129&lng=es&nrm=iso. Acesso em: 14 nov. 2021.

LÉVY, Pierre. **Cibercultura**. Paulo Neves (Trad.) São Paulo: Ed. 34, 2005.

LLUECA, Ciro et al. A ritmo de tweet: archivando elecciones 2.0. **Profesional de la Información**, v. 20, n. 3, p. 309-314, 2011.

MENEZES, Estera. **Pesquisa Bibliográfica**. UFSC. Florianópolis. 2009. 86 p.

YEO, Geoffrey. Falando sobre diferença: Percepções sobre unicidade e identidade nos arquivos. In: **Arq. & Adm.**, Rio de Janeiro, v.10, n.1, jan./jun. 2011

MURPHY, Jamie; HASHIM, Noor Hazarina; O'CONNOR, Peter. Take me back: validating the wayback machine. **Journal of Computer-Mediated Communication**, v. 13, n. 1, p. 60-75, 2007. <https://academic-oup-com.ez46.periodicos.capes.gov.br/jcmc/article/13/1/60/4583038>