

Arquivamento da web: sistematização de razões e justificativas para arquivar sites¹

Danilo Formenton^I

^IUniversidade Federal de São Carlos, São Carlos, SP, Brasil;
 formenton.danilo@gmail.com; <https://orcid.org/0000-0001-6531-4295>

Luciana de Souza Gracioso^{II}

^{II}Universidade Federal de São Carlos, São Carlos, SP, Brasil;
 lugracioso@yahoo.com.br; <https://orcid.org/0000-0002-6320-4946>

Resumo: O objetivo deste estudo é apresentar razões para se arquivar sites por meio de casos de uso de arquivos da web e de arquivamento da web. Caracteriza-se como uma pesquisa exploratória que adota o método bibliográfico e documental a partir de uma revisão de literatura sobre o assunto. Aplicou-se a análise de conteúdo da revisão de produções científicas atuais nacionais e internacionais da Ciência da Informação, assim como, de sites oficiais, políticas e diretrizes de arquivos da web no mundo. Concluiu-se que há motivações mais abrangentes para que instituições desenvolvam os arquivos da web, incluindo-se, dentre tais motivações, a urgência de se preservar conteúdos web, devido ao risco de perda desse conteúdo, salvaguardando, dessa maneira, o patrimônio digital produzido por nações, e os inúmeros casos de uso dos arquivos da web. Complementarmente, no entanto, reconhece-se, efetivamente, a questão da legislação, que pode obrigar e estimular essas iniciativas, conferindo o direito e a proteção desses conteúdos às instituições, delegando-lhes as funções de captura, abrigo e arquivamento dos materiais web, seja no papel de um depositário legal, visando: preservação da memória social; ou, por conformidade normativa, e apoio em processos judiciais. Entender os *insights* providos pelos arquivos da web justificam, em certa medida, os futuros investimentos financeiros, políticos, científicos, etc., na preservação digital e arquivamento da web, sobretudo, em países que não detêm um arquivo da web nacional, como no caso do Brasil.

Palavras-chave: arquivamento da web; preservação digital; websites; justificativas; Ciência da Informação

1 Introdução

Nas últimas décadas, a preservação digital se tornou uma temática de pesquisa do campo da Ciência da Informação e suas áreas afins (Arquivologia, Biblioteconomia e Museologia). É um desafio emergente, coletivo, complexo e

inevitável que está cada vez mais presente nas publicações nacionais e internacionais do campo, demandando abordagens e análises inter/multidisciplinares, assim como soluções sustentáveis, integradas e colaborativas.

Para maior clareza do conceito de preservação digital, utilizamos a definição de Duranti (2010, p. 157, tradução nossa), que o entende como sendo o “[...] conjunto de princípios, políticas, regras e estratégias destinadas a prolongar a existência do objeto digital, mantendo-o em condições adequadas para uso [...]”, complementando-se pela função de proteger “[...] a identidade e integridade do objeto, ou seja, sua autenticidade.” (Duranti, 2010, p. 157, tradução nossa).

Estes objetos digitais, nascidos digitais ou digitalizados, são todos os tipos de conteúdo em meio digital – tais como, textos, imagens, vídeos, áudios, jogos, sites, mídias sociais, e-mails, etc. –, dos quais a preservação digital pode agir, constituindo-os enquanto “[...] itens na forma digital que requerem um computador para dar suporte à sua existência e apresentação visual.” (Pinheiro; Ferrez, 2014, p. 163), considerando-se ainda que, para Baucom (2019, p. 5, tradução nossa) tais objetos são compostos por “[...] cadeias de uns e zeros, que requerem componentes específicos de software e hardware para permanecerem acessíveis aos usuários.”

Além da emulação tecnológica, ou da migração de dados, ou da adoção de padrões de metadados, etc., uma das estratégias de preservação digital abrange a manutenção e o arquivamento do conteúdo de websites, segundo Formenton e Gracioso (2020, 2022). Cabendo mencionar que, sendo um tema recente e carente de investigações, de políticas públicas, de iniciativas oficiais e sistematizadas, no Brasil (Luz, 2022; Rockembach; Pavão, 2018), o arquivamento da web (*web archiving*) integra, no momento, o processo de seleção e captura, armazenamento, preservação e fornecimento de acesso de conteúdo da web, em longo prazo.

Sobre a preservação de páginas da web enquanto documentos digitais reconhecidos como documentos arquivísticos² (Conselho Nacional de Arquivos, 2020), Flores (2021, p. 13), aponta que o documento arquivístico digital se

define como algo complexo e específico. Segundo o autor para que tal documento seja usado “[...] de fonte de prova, evidências, testemunho, memória, patrimônio, garantia de direitos e exercício pleno da cidadania [...]”, “[...] este original (nativo digital) ou um representante digital (digitalização)” deve ser mantido:

- a) autêntico - com controle de sua transmissão, preservação e custódia, isto é, dotado de componentes de identidade e integridade;
- b) confiável (possuir completude na forma e no controle de produção) - considerando-o em uma cadeia de custódia digital eficiente.

No tocante à presunção de autenticidade, cabe considerar a Resolução n. 37, de 19 de dezembro de 2012, do Conselho Nacional de Arquivos (CONARq) (CONARq, 2012), que aprova as diretrizes para a presunção de autenticidade de documentos arquivísticos digitais, citada por Flores (2021), que aponta que essa presunção no contexto de documentos arquivísticos “[...] sempre fez parte do processo tradicional de avaliação desses documentos e é fortemente apoiada na análise de sua forma e de seu conteúdo [...]”, dois elementos que “[...] nos documentos não digitais estão inextricavelmente ligados ao suporte – isto é, forma, conteúdo e suporte são inseparáveis.” (Conselho Nacional de Arquivos, 2012³, p. 1 *apud* Flores, 2021, p. 20).

Complementando-se ainda que tal presunção se baseia:

[...] na confirmação da existência de uma cadeia de custódia ininterrupta⁴, desde o momento da produção do documento até a sua transferência para a instituição arquivística responsável pela sua preservação no longo prazo. Caso essa cadeia de custódia seja interrompida, o tempo em que os documentos não estiveram sob a proteção do seu produtor ou sucessor pode causar muitas dúvidas sobre a sua autenticidade (Conselho Nacional de Arquivos, 2012⁵, p. 1 *apud* Flores, 2021, p. 20).

Ainda quanto a autenticidade de documentos arquivísticos digitais, Hirtle (2000, p. 10, tradução nossa) complementa as presentes considerações, observando que “[...] um verdadeiro arquivo é um corpo orgânico de evidências contextualmente baseado, não uma coleção de informações diversas.” e que a sua existência está condicionada a existência de “[...] uma cadeia de custódia

ininterrupta desde o órgão que o criou até aos arquivos.” (Hirtle, 2000, p. 12, tradução nossa).

Neste sentido, Terrada (2022, p. 71), parafraseando as considerações de Flores (2021), – assim como a palestra *Preservação de Páginas web e Redes Sociais em Cadeia de Custódia: Identificação, Seleção e Arquivamento* (2021) e Hirtle (2000) –, refletem o mesmo princípio, ou seja, de que o produtor do documento necessita “[...] prover um ambiente seguro desde a produção do documento até o momento de transferência para instituição arquivística responsável pela salvaguarda e preservação em longo prazo.” (Terrada, 2022, p. 71).

Visando evitar a perda permanente de sites – em razão da dinamicidade da Internet – e promover a preservação digital dos seus conteúdos, diversas iniciativas de arquivos da web vêm surgindo no mundo, apresentando-se por diferentes formações e propósitos de arquivamento, com abordagens em âmbito: (1) global (por exemplo, o *Internet Archive* (Internet Archive, 2014c) uma organização sem fins lucrativos, iniciada em 1996, que fornece acesso gratuito a uma biblioteca digital com milhões de páginas web, ebooks, imagens, etc. de várias partes do mundo); ou, (2) nacional, regional e local (por exemplo, o *Arquivo.pt* em Portugal (Arquivo.pt, 2008), o arquivo da web da Catalunha, da Biblioteca da Catalunha na Espanha (Biblioteca de Catalunya, c2011), e os arquivos da web das Bibliotecas da Universidade de Columbia, nos Estados Unidos, (Columbia University Libraries, c2021)).

Aliás, para apoiar a preservação e o desenvolvimento da web, consórcios internacionais foram estabelecidos com a finalidade de definir padrões e diretrizes que orientem sua expansão e arquivamento de seu conteúdo ao longo do tempo, tais como: o *World Wide web Consortium* (W3C) (W3C, c2024), o *International Internet Preservation Consortium* (IIPC) (IIPC, c2024), ou, ainda, a *Digital Preservation Coalition* (DPC) (DPC, c2024) (Formenton; Gracioso, 2020; Rockembach, 2018) e o *Internet Engineering Task Force* (IETF) (IETF, [2022]).

No Brasil, nos últimos anos, surgiram algumas iniciativas de grupos de pesquisa, projetos e resoluções dedicados à temática (Boeres, 2023; Conselho

Nacional de Arquivos, 2022, 2023a, 2023b; Formenton; Gracioso, 2020; Luz, 2022; Rockembach, 2018), dentre os quais podemos citar:

- a) grupo *Estudos e Práticas de Preservação Digital* – ou Rede DRÍADE –, criado em 2014 pela Rede Brasileira de Serviços de Preservação Digital (Cariniana) do Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT), que integra grupos de estudo sobre arquivamento da web, arquivamento de e-mail, preservação de dados de pesquisa, etc., tendo por enfoque a preservação digital (Rede Brasileira de Serviços de Preservação Digital, c2022);
- b) grupo *Núcleo de Pesquisa em Arquivamento da web e Preservação Digital* (NUAWEB), criado em 2017, na Universidade Federal do Rio Grande do Sul (UFRGS), para investigar, por meio de iniciativas nacionais e internacionais, abrangendo as políticas e as tecnologias envolvidas, assim como, aspectos temáticos do tema (UFRGS, 2017);
- c) Câmara Técnica Consultiva (CTC) *Preservação de websites e mídias sociais*, instituída em 2021, no âmbito do CONARq, com o propósito de definir diretrizes para a elaboração de estudos, proposições e soluções para a preservação desses documentos dinâmicos e interativos (CONARq, 2022);
- d) projeto piloto ARQWEB – Serviço de Preservação de Páginas web –, criado em 2022 (em comemoração aos dez anos da Rede Cariniana, do IBICT, já citada), para arquivar os sites das instituições parceiras da Cariniana, sites governamentais, etc. (ARQWEB, [2022]);
- e) Resoluções n. 52 e n. 53, ambas instituídas em 2023, no contexto do CONARq, com os intuiitos de: estabelecer a política de preservação de sites e mídias sociais no âmbito do Sistema Nacional de Arquivos (SINAR) (o qual implementa a política nacional de arquivos públicos e privados); e, expor aos integrantes do SINAR as condições mínimas necessárias para se atender ao objetivo de preservar estes materiais por longo prazo (CONARq, 2020, 2023a, 2023b).

Considera-se que no Brasil, apesar dos exemplos citados, a área segue ainda carente de iniciativas organizadas em favor do desenvolvimento de

arquivos da web, e de estudos teóricos e práticos que investiguem as várias facetas do assunto. Neste cenário, a presente pesquisa procura proporcionar uma visão amplamente reflexiva das razões e justificativas para arquivar sites a partir de casos de uso para os arquivos da web e o arquivamento da web.

Acredita-se que, a partir de um mapeamento dos principais argumentos e defesas apontadas na literatura especializada, assim como pelas próprias iniciativas internacionais de arquivos da web, seja possível fornecer um apoio decisório eficiente, com informações tão embasadas quanto adequadas sobre políticas, estratégias e/ou formas de investimentos em preservação digital e arquivamento da web.

Para este fim, optou-se por uma abordagem exploratória referente ao tema de arquivamento da web, no escopo da preservação digital (Cordeiro *et al.*, 2007; Gil, 2010; Severino, 2016), no âmbito da Ciência da Informação, adotando-se como metodologias: a construção de *corpus*/coleta de dados, realizadas por meio de uma revisão de literatura, pesquisa bibliográfica e documental; e, a análise dos dados coletados, utilizando-se da análise do conteúdo (Bardin, 2016; Cavalcante; Calixto; Pinheiro, 2014) expresso em produções científicas recentes, nacionais e internacionais.

Os documentos de interesse foram artigos de periódicos, anais de eventos, dissertações e livros da Ciência da Informação, e áreas afins, buscados em fontes indexadoras, tais como: o Google Scholar (Google Scholar, c2024); a Scientific Electronic Library Online (SCIELO) (SCIELO, c2024), as bases de dados Scopus (Elsevier), ScienceDirect (Elsevier), Web of Science (Clarivate Analytics) e Emerald Insight (Emerald Publishing), disponíveis via Portal de Periódicos da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) (CAPES, c2020).

Sites oficiais, políticas e diretrizes acerca de iniciativas de arquivos da web no mundo também foram utilizados, cabendo observar que, no tocante a essas fontes supracitadas, foram reconhecidas e sistematizadas algumas das principais razões para se arquivar sites tendo por apoio uma série de casos de uso, ou seja, baseando-se nas experiências de aplicação identificadas.

Para o tratamento de assunto de pesquisa, o estudo fundamentou-se,

principalmente, em teorias da Ciência da Informação e da Biblioteconomia, fazendo uso dos procedimentos de revisão de literatura, e de análise de conteúdo bibliográfico/documental, identificando e categorizando os motivos mais comuns (apontados por experiências empíricas e/ou discutidos na teoria) para o desenvolvimento do arquivamento da web e dos arquivos da web.

Ademais, tais desenvolvimentos se coadunam com a exemplificação de casos de uso de algumas iniciativas representativas de arquivos da web pelo mundo (identificadas, sobretudo, em bibliotecas nacionais e universitárias, e arquivos nacionais), relacionando-os aos motivos identificados e descritos, destacando-se interconexões, e, obtendo-se *insights* práticos.

Deste modo, o trabalho em questão se dispôs a expor os resultados e as análises dos conteúdos coletados, sendo que o produto deste mapeamento previu colaborar tanto na “sensibilização” da comunidade acadêmico-científica, do governo, das instituições públicas ou privadas e do público brasileiro em geral, sobre a importância do arquivamento da web e dos arquivos da web, visando estimular um maior engajamento do tema no Brasil.

2 As razões e justificativas para o arquivamento da web, segundo os casos de uso relacionados aos arquivos da web

Como indicado por Pennock (2013), apesar de se configurar como uma das principais razões para arquivar sites (sobretudo na comunidade de patrimônio cultural), preservar o conteúdo da web para evitar meramente que o mesmo seja perdido, é um argumento fraco quando comparado a obrigação legal que algumas instituições possuem de capturar e arquivar conteúdo web.

Por exemplo, a Biblioteca Nacional da França (do francês *Bibliothèque Nationale de France* (BnF)) se fundamenta juridicamente no depósito legal da web francesa (BnF, c2022), por meio da Lei n. 2006-961, de 2006 (*République Française*, 2006) – lei relativa aos direitos de autor e direitos conexos na sociedade da informação (do francês *loi relative au droit d'auteur et aux droits voisins dans la société de l'information*), conhecida como Lei DADVSI –, garantindo que sites do domínio francês sejam coletados e preservados, consultados em salas de leitura e reproduzidos segundo o Código de Propriedade

Intelectual francês (*Code de la Propriété Intellectuelle*) (République Française, 2024).

Os Regulamentos de Bibliotecas de Depósito Legal (obras não impressas) (do inglês *Legal Deposit Libraries – Non-Print Works – Regulations*), de 2013 (The National Archives, 2013), também habilitam as Bibliotecas de Depósito Legal do Reino Unido (ALDL, 2013)⁶ que compõem o arquivo da web do Reino Unido (do inglês *UK web Archive*) (The National Archives, [2024]) a coletar qualquer site baseado neste país, preservá-los para as gerações futuras, e disponibilizá-los aos usuários em suas instalações. Além dos requisitos legislativos para se coletar conteúdo da web sob depósito legal – os tornando parte do patrimônio de nações –, outras iniciativas de arquivamento da web – como o *Coca-Cola web Archive*, em colaboração com o serviço da *Hanzo Archives* (Hanzo, c2024) – arquivam sites tanto para apoio jurídico-processual – permitindo que parte/totalidade do site possa ser solicitada em tribunal – como para conformidade normativa para aplicações de gestão de registros.

Para mais, o arquivamento de sites pode ocorrer ainda devido a um interesse social em documentar a evolução e o conteúdo da Internet como um todo, possibilitando sua disponibilização para usuários, como no caso do *Internet Archive* (Pennock, c2013).

No arquivo da web da Nova Zelândia (do inglês *New Zealand web Archive*) da Biblioteca Nacional da Nova Zelândia, por exemplo, o arquivo da web é usado para obter um registro visual de como sites da Nova Zelândia e do Pacífico mudaram ao longo do tempo. A biblioteca nacional da Nova Zelândia coleta por depósito legal, arquiva e preserva para pesquisa as publicações neozelandesas (livros, sites, blogs, etc.), embasando-se na *National Library of New Zealand (Te Puna Mātauranga o Aotearoa) Act*, de 2003, e na responsabilidade social de preservar a história social e cultural do país (National Library of New Zealand, [2022?a, 2022?c]). A obrigação de depósito legal desta lei permite que a Biblioteca Nacional da Nova Zelândia colete e preserve para a sua comunidade, e para as gerações futuras, as publicações neozelandesas (isto é, qualquer trabalho físico e digital publicado na Nova Zelândia, incluindo obras publicadas por neozelandeses em websites hospedados em plataformas de auto

publicação no exterior) como, por exemplo, livros, sites, músicas gravadas, periódicos, mapas, partituras, dentre outros (National Library of New Zealand, [2022?b]).

As *Stanford Libraries* ou, melhor, as bibliotecas da Universidade Stanford, nos Estados Unidos (do inglês *Stanford University Libraries*), listam, em seu projeto de arquivamento da web (Stanford University, 2016) alguns motivos para seus esforços, justificando-se por via de uma vasta gama de casos de uso local, entre eles:

- a) preservação do legado institucional na web - os artefatos impressos que a muito tempo contam a história da universidade deram lugar ao endereço da web “www.stanford.edu” como a representação mais consolidada da evolução da universidade. Isto pode ser exemplificado com algumas coleções no *Archive-It* (isto é, o serviço de arquivamento da web por assinatura, lançado em 2006 pelo *Internet Archive* (Internet Archive, 2014a), para que instituições criassem, armazenassem e dessem acesso a coleções de conteúdo web), tal como a *Stanford University website Collection* (Internet Archive, 2015a), a qual inclui sites dos seus departamentos, laboratórios, institutos, eventos, etc ;
- b) contribuição para aprendizagem - o arquivamento da web possibilita a captura das páginas ou instantâneos (*snapshots*) dos sites de empresas que são objetos de estudo da *Graduate School of Business* (Stanford University, [2024]), da universidade, e as coleções no *Archive-It* (Internet Archive, 2014b) com potencial de valor acadêmico são recursos exclusivos, tal como a *Digital Games* (Internet Archive, 2008), que oferece um contexto complementar para a coleção de Stephen M. Cabrinety (OAC, [ca. 2009]), contendo software, jogos, literatura sobre a indústria de jogos de microcomputação, etc;
- c) materiais complementares para coleções especiais físicas - o Departamento de Coleções Especiais e Arquivos Universitários seleciona e preserva materiais de valor histórico duradouro, buscando apoiar as necessidades de pesquisa de alunos e docentes da universidade. O arquivamento da web permite adicionar materiais complementares e/ou

ausentes nestas coleções, como os sites arquivados das coleções *Patrick Suppes* (Internet Archive, 2015b), e *Philip G. Zimbardo* (Internet Archive, 2015c) no *Archive-It*, que documentam as trajetórias destes professores;

- d) gerência de dados governamentais - o governo eletrônico (*electronic government*, ou *e-gov*) oferece novas opções para a divulgação das informações do governo, e novos desafios à sua preservação. United States (2002, sec. 101, 3601, tradução nossa) explica que governo eletrônico é definido como a utilização, pelo Governo dos Estados Unidos “[...] de aplicações da Internet baseadas na web e outras tecnologias da informação, combinadas com processos que implementam estas tecnologias [...]”, visando “[...] aumentar o acesso e a entrega de informações e serviços do Governo ao público, outras agências e outras entidades governamentais; ou”, objetivando “[...] trazer melhorias nas operações do Governo que podem incluir eficácia, eficiência, qualidade de serviço [...]. Neste cenário: o arquivamento da web expande o escopo da informação documental que a universidade pode coletar e organizar para as suas comunidades; e, as coleções têm respostas aos pedidos da *Freedom of Information Act* (FOIA) (FOIA, [2024]), com presenças web de governos da região da Baía de São Francisco, na Califórnia, por meio das coleções de sites como *Freedom of Information* (FOIA) (FOIA, 2007) e *Bay Area Governments* (Internet Archive, 2007), no *Archive-It*;
- e) salvaguarda de resultados acadêmicos - os projetos de alunos e docentes resultam cada vez mais na criação de sites como subprodutos auxiliares, ou mesmo, centrais. Exemplos incluem as coleções de conteúdos da web *Carolyn Bertozzi* (Internet Archive, 2015d), *Carl Djerassi* (Internet Archive, 2015e), *Stanford University Student Organizations website Collection* (Internet Archive, 2015f) e *Center for Relationship Abuse Awareness* (Internet Archive, 2015g), no *Archive-It*, que têm sites de professores, e dos seus grupos de pesquisas, de órgãos estudantis da universidade, e de centros de treinamento;

f) conformidade e gerenciamento de registros - o recredenciamento, as questões jurídicas e as demais ações de conformidade, podem requerer o acesso a versões de informações partilhadas nos sites da universidade. O arquivamento da web oferece um mecanismo forense na gestão de registros, cumprimento e redução de riscos de litígio para preservar as políticas e a documentação baseadas na web, à medida que estas mudem no decorrer do tempo. Por exemplo, a coleção *Stanford University COVID-19 Response* (Internet Archive, 2020a), no *Archive-It*, inclui sites arquivados que documentam a resposta da instituição à pandemia de coronavírus.

A Biblioteca do Congresso, também, de fato biblioteca nacional dos Estados Unidos, como um dos membros fundadores do IIPC (IIPC, 2000) demonstra esse interesse no arquivamento web, pois, embora hoje não seja legalmente obrigada a arquivar sites, arquiva conteúdo online nascido digital, ou em risco de perda, por meio do seu programa de arquivamento da web, desde 2000, em um esforço para prover acesso e preservar esses objetos efêmeros, assim como a instituição tem feito com materiais impressos (Library of Congress, [2022?b]).

No seu arquivo da web, a Biblioteca do Congresso preserva e cede acesso para pesquisa a sites arquivados, notificando seus proprietários que gostaria de incluir o seu conteúdo no arquivo, antes de seu arquivamento (exceto no caso de sites do governo americano, ou aqueles que usam *Creative Commons* (Library of Congress, 2022a, [2022?c]). Cabendo, complementarmente, definir que, para a *Society of American Archivists* (*Creative commons*, c2022a, tradução nossa) *Creative Commons* é “[...] um tipo de licença, baseada em direitos autorais, que fornece uma forma padronizada para os criadores concederem a outras pessoas o direito de compartilhar e usar seu trabalho”.

Assim, a Biblioteca do Congresso dos Estados Unidos (Library of Congress, 2022a) se esforça para construir coleções que registrem a criatividade americana e refletem a diversidade e complexidade do país, com prioridade na aquisição de material em diversas perspectivas e vozes sub-representadas, buscando assegurar a variabilidade de autoria, identidades culturais e outros

fatores histórico-culturais. A título de exemplos, as coleções web *Women's and Gender Studies web Archive* (Library of Congress, 2018a), *LGBTQ+ Politics and Political Candidates web Archive* (Library of Congress, 2018b) e *LGBTQ+ Studies web Archive* (Library of Congress, 2018c), incluem conteúdo online sobre movimentos culturais, sociais e políticos pela igualdade de gênero, órgãos políticos e jurídicos LGBTQ+ nos Estados Unidos, e a história, o saber e a cultura LGBTQ+ americana e mundial, complementando os acervos físicos da biblioteca.

Junto ao desenvolvimento das suas coleções, a biblioteca ainda vem cooperando ativamente com outras organizações para documentar fatos que se manifestam na web ao redor do mundo, como a coleção *Ukraine Conflict* (Internet Archive, 2014) – construída com a equipe do *Archive-It*, especialistas da Universidade de *Stanford*, etc. –, que documenta o conflito na Ucrânia.

2.1 Arquivos da web como testemunhas digitais de fatos e guardiões da memória pessoal, social, organizacional e científica

Os casos de uso associados aos arquivos da web podem similarmente fornecer inúmeros motivos para arquivar páginas web que documentam tanto eventos recentes, dados governamentais, reações públicas, notícias históricas, instituições culturais, informações de fonte para pesquisas acadêmico-científicas quanto uma ampla diversidade de tópicos, com conteúdo produzido em várias nações, em diferentes idiomas e plataformas (sites, blogs, mídias sociais etc.).

A lista destas ocorrências e tópicos compõe-se de:

- a) eventos espontâneos - constituindo uma classe de conteúdo da web em risco, os eventos espontâneos (isto é, catástrofes, acidentes, revoluções, tópicos sociais populares, etc.) podem ocupar brevemente os holofotes do público e depois sumir de vista (Stanford Libraries, [c2022?]) – como muitas notícias mudam para *feeds* informais de mídia social de rápida atualização, volumes de dados sobre eventos atuais podem ser perdidos. Os arquivos preservam tal conteúdo como partes do registro histórico, conforme *International Internet Preservation Consortium* (IIPC, c2024). Por exemplo: a *Ukraine Conflict* (Internet Archive, 2014d), no *Archive-*

It, do *Internet Archive Global Events*, já citada no final da seção anterior, inclui blogs, mídias sociais, etc., acerca do conflito na Ucrânia, desde 2014;

- b) preservação de citações e referências na web - os arquivos da web podem servir para a citação de versões de um conteúdo web, tais versões podem ser usadas como referências em obras acadêmicas, aumentando a longevidade da citação e o seu valor para futuros leitores (Pennock, c2013). Ou seja, os arquivos da web podem fornecer links para versões específicas e estáveis do site, utilizando-se de identificadores persistentes formais atribuídos a cada recurso, ou por uma estrutura de URL consistente e estável para acessar recursos (como é o caso do arquivo da web da Biblioteca do Congresso Americano) que recebem um ID exclusivo da citação, redirecionando a busca para o local do site arquivado, assegurando, ainda, de acordo com *International Internet Preservation Consortium* (IIPC, c2024), que os sites citados sejam localizados mesmo que a estrutura de URL padrão do arquivo seja alterada. A *Digital Preservation Coalition* (c2015, p. 36-37, tradução nossa) define o termo Identificador persistente (*persistent identifier*) como “[...] uma referência duradoura a um recurso digital.”, compondo-se de um identificador único para “[...] garantir a proveniência de um recurso digital (que é o que propõe ser) [...]”, e de um serviço duradouro “[...] que localiza o recurso ao longo do tempo mesmo quando sua localização muda.”, assegurando que “[...] o identificador aponte para a localização atual correta.”, visando “[...], assim, solucionar o problema da persistência de acesso ao recurso citado, em particular na literatura acadêmica.”. Exemplos de uso desse identificador podem ser identificados em: *Digital Object Identifier* (DOI), *Persistent Uniform Resource Locator* (PURL), *Uniform Resource Name* (URN), *Handle System* e outros esquemas (*schemes*) de identificadores persistentes. Para o Conselho Nacional de Arquivos (2020, p. 34), Identificador persistente se refere ao “Identificador de longa duração de um recurso na Internet”.

- que se mantém válido mesmo que a tecnologia de acesso ou a localização física do recurso identificado se modifique no tempo.”;
- c) comunicação científica - diferentes coleções de sites arquivados contêm conteúdo web que auxiliam na comunicação e na divulgação de assuntos da ciência. Nesse sentido, existe, no arquivo da web da Biblioteca do Congresso americano: a coleção *Science Blogs web Archive* (Library of Congress, 2013) que, considerando os blogs de ciência como periódicos ou diários online que enriquecem o acervo analógico de revistas científicas da biblioteca, provê recursos para acadêmicos e outros interessados em pesquisas sobre redação, ensino e comunicação científica nos Estados Unidos;
 - d) guerra Russo-Ucraniana - com a invasão russa na Ucrânia em 2022, várias iniciativas têm identificado, coletado, gravado e arquivado sites ucranianos antes destes se perderem durante a guerra, buscando preservar a memória cultural e digital do país. Como, por exemplo, o *Saving Ukraine Cultural Heritage Online* (SUCHO) (SUCHO, 2022a), composto por uma equipe internacional de mais de mil voluntários (incluindo bibliotecários, arquivistas, etc.) que trabalham para criar arquivos web de sites de instituições culturais ucranianas em risco de perda – tal como o site do governo ucraniano sobre o arquivo oficial de *Kharkiv* (SUCHO, 2022b) –, utilizando-se: do envio de URLs ao software *Wayback Machine* (Internet Archive, 2014e), do *Internet Archive*, que reproduz páginas da web de sites arquivados; ou do *Conifer* (Conifer, [2020?]), que gera gravações de experiência de navegação nos sites; além de outras tecnologias usadas para rastrear, arquivar e auxiliar na reconstrução de sites (Adams; Fernandez, 2022; Serrano, 2022);
 - e) ataques terroristas na França - após os atentados de Paris, em 2015, e de Nice, em 2016, muitas instituições lançaram coleções especificamente centradas no massacre que atingiu o jornal satírico francês Charlie Hebdo, como a coleção *Charlie Hebdo* (Internet Archive, 2015h), da equipe do *Archive-It*, que contém mídias sociais, notícias e sites institucionais relacionados a esse ataque, em Paris. A BnF e o *Institut*

National de l'audiovisuel (INA) (INA, [c2024]) criaram “coleções de emergência”, buscando capturar uma amostra das reações online oficiais e populares sobre os ataques, por meio dos rastros deixados na Internet, e no Twitter, incluindo homenagens, apoio, opiniões críticas e hostis, etc., que para Schafer *et al.* (2019), oferece, complementando-se com demais materiais (artigos de imprensa, fotografias, entrevistas, etc.), tanto um “quadro” de uso potencial para fonte de pesquisa quanto uma resposta social aos ataques;

- f) comunidade LGBTQIA+ - algumas instituições arquivam sites referentes a população LGBTQIA+, a fim de preservar as memórias culturais, sociais e políticas destes grupos no mundo. Por exemplo, por meio do uso do *Archive-It*, a Universidade da Califórnia, em Berkeley, nos Estados Unidos, criou as coleções *Southeast Asia LGBT web Archive* (Internet Archive, 2015i) e *Archiving the LGBT web: Eastern Europe and Eurasia* (Internet Archive, 2015j); e o UKWA tem a coleção *LGBTQ+ Lives Online*, o qual se oferece como recurso para pesquisas sobre o assunto, enriquecendo as coleções impressas das bibliotecas parceiras desta iniciativa. Cocciole (2016) e Pendse (2014) também exploraram os desafios e a utilidades específicas da criação de arquivos web de comunidades LGBT para documentar e preservar (para fins de pesquisa acadêmica) os movimentos de luta por seus direitos civis, contra os efeitos da epidemia de AIDS, pelo aumento de aceitação social, etc.;
- g) jornalismo digital - as coleções da web de notícias históricas criadas pelas instituições de arquivamento servem como fonte para estudos baseados em jornalismo online. Por exemplo, existem a *ABC News - Australian Internet sites*, no *Preserving and Accessing Networked Documentary Resources of Australia* (PANDORA Archive), com notícias da *Australian Broadcasting Corporation (ABC) News* na Austrália (National Library of Australia, 2018); e a *Hurricane Katrina blogs web collection* (Internet Archive, 2016), no *Archive-It*, da Universidade do Mississippi, nos Estados Unidos, o qual documenta uma

amostra representativa de blogs e jornalismo, produzidos por vítimas do furacão Katrina;

- h) historiografia da web - os arquivos da web, criados por instituições de arquivamento, configuram-se como fontes historiográficas em que as informações culturais, etc., preservadas nestes repositórios, têm possibilidades de uso por historiadores, e outros pesquisadores de estudos históricos da web, e da Internet (Rodrigues; Rockembach, 2021). Exemplificando-se, por meio das coleções: *Personal stories of Australians in war*; e, *Historic gold mining sites* no PANDORA Archive (National Library of Australia, [c2024a], [c2024b]), ambas, podem ser usadas por historiadores da web para fornecer dados e documentar estudos;
- i) comunidade indígena - diversas bibliotecas e universidades desenvolvem, por meio do *Archive-It*, coleções na web, voltadas as comunidades indígenas. Como, por exemplo: a *Hawaii – Hawaiians*, da Universidade do Havaí (Internet Archive, 2010), que inclui sites com informações sobre havaianos indígenas, questões nativas e de soberania; e, a *Policing, Racism, and Indigenous People in Thunder Bay*, da Universidade de Lakehead, no Canadá (Internet Archive, 2017), que contém notícias e respostas de *Thunder Bay* tanto a ocorrências de racismo anti-indígena quanto de questões de policiamento. No *New Zealand web Archive* (National Library of New Zealand, [2022?c]), os sites arquivados por, e para, os povos da etnia *Māori*, têm esses povos fielmente representados em sua coleção, o que, segundo Ka‘ai-Mahuta (2019), contribui na preservação e disponibilização de informações digitais, partilhadas pelos povos indígenas, garantindo, às gerações futuras, a continuidade cultural e transmissão de saberes;
- j) comunidade negra - determinadas iniciativas se dedicam a criar coleções centradas nas comunidades e cultura negra, pelo mundo. Por exemplo: o arquivo da web do Reino Unido (do inglês UK web Archive - UKWA) tem a coleção *Black and Asian Britain*, que inclui sites referentes a cultura e a história da presença negra e asiática no Reino Unido; e, o

Middlebury College, por meio do *Archive-It*, criou a coleção *Community Responses to Anti-Black Racism and Police Violence* (Internet Archive, 2020b), que inclui sites com respostas e reações de indivíduos e organizações ao assassinato de George Floyd, em 2020⁷ (Cardoso; Cartchuk, c2024.), assim como pela luta por justiça social para os negros nos Estados Unidos. Rollason-Cass e Reed (2015) examinam a criação de uma coleção da web em torno do movimento *#blacklivesmatter*⁸ (Black Lives Matter Global Network Foundation, 2013), isto é, o “*#blacklivesmatter web Archive*” (Internet Archive, 2014f), no *Archive-It*, que provê um recurso valioso para pesquisadores, ativistas e historiadores sobre o movimento contra os maus-tratos de afro-americanos pelas mãos das autoridades;

- k) mudanças climáticas e desastres naturais - muitas coleções web documentam os efeitos e as respostas aos eventos climáticos ocorridos no mundo. Por exemplo: a *Indian Ocean Tsunami December 2004*, no UKWA, com sites de órgãos que prestam ajuda e socorro, ou para registro de experiências pessoais em situações emergenciais, etc., como no desastre causado pelo *tsunami* de 2004, na Ásia; e a *Japan Earthquake* (Internet Archive, 2011), no *Archive-It* da *Virginia Tech: Crisis, Tragedy, and Recovery Network* (Virginia Tech, c2024), com blogs, sites de notícias, etc., que retratam os eventos acerca do terremoto, e o consequente *tsunami*, ocorridos no Japão, em 2011, assim como a reconstrução pós-desastre. Outros exemplos são mencionados por: Rockembach e Serrano (2021), que demostram a relevância da preservação do conteúdo web sobre mudanças climáticas, mostrando o que foi, e o que terá de ser preservado no futuro; e, Radinsky e Horvitz (2013), que mineraram a web visando a previsão de eventos futuros por meio de seus registros, estabelecendo uma relação entre secas e tempestades em Angola, e o estímulo dos surtos de cólera.

Assim, segundo Stanford Libraries ([c2022?], local. tradução nossa) “assegurar a capacidade contínua de acesso ao conteúdo da web [...]” é considerado um “[...] imperativo para objetivos tão diversos como pesquisa,

ensino, construção de coleção de biblioteca, legado institucional, conformidade legal e gestão da informação governamental.”. Alnoamany, Weigle e Nelson (2016) complementam essa consideração observando que esse conteúdo é um repositório importante da nossa história recente, e da nossa herança cultural.

Por essa via, os arquivos da web, como coleções web temáticas acessadas por endereços eletrônicos, ou baseadas em registros de regiões e eventos, são uma fonte útil de informações únicas, e historicamente valiosas para pesquisa, sobretudo pelos seus tópicos e à sua qualidade de coleta, propiciando explicar as histórias do passado e a conjecturar eventos futuros, tendo por instrumentos, em conformidade com Cadavid (2017) e Costa, Gomes e Silva (2017), a extração, modelagem e análise da evolução de dados.

O *International Internet Preservation Consortium* (c2022) e Reynolds (2013) elucidam outros casos de uso para os arquivos da web (e o arquivamento da web), dentre os quais se destacam:

- a) análise de links (*link analysis*) - como a coleta de grandes conjuntos de sites abrange também a captura de links e conexões entre eles, essas redes de sites e dados vinculados podem ser extraídos para observar relações entre pessoas, ideias, organizações, etc., ao longo do tempo. Assim como em sites na web ao vivo, tal análise poderá ser usada com dados de arquivos da web, observando-se mudanças em períodos de tempo presente ou em períodos do passado;
- b) atividade de extensão e educação (*outreach and education*) - já que a web se tornou parte dos serviços de instituições educacionais e de patrimônio cultural, os arquivos da web têm sido utilizados em exposições em formato presencial (físico) ou virtual (online) – de museus, por exemplo. Existindo, além disso, esforços para incluir alunos na criação de coleções de arquivos da web, com o propósito de envolvê-los com a história e realçar a relevância de se coletar sites;
- c) prestação de contas (*accountability*) - como o rastreamento de sites ao longo de um período de tempo, permite-se analisar alterações no conteúdo da web – esse acesso é útil para garantir a prestação de contas de um conteúdo que não existe mais. Por exemplo: as empresas podem

arquivar seu conteúdo web como estratégia defensiva contra ações judiciais; ou, ainda, arquivos públicos da web podem mostrar mudanças evolutivas nas políticas ou práticas de governos, organizações, etc.;

- d) vinculação persistente (*persistent linking*) - enquanto conteúdos web podem mudar ou desaparecer sem aviso prévio, os arquivos da web oferecem aos usuários links para acesso a versões específicas e estáveis do conteúdo de interesse, via, por exemplo, identificadores persistentes (já mencionados na lista anterior). Isto permite que os usuários consultem tal conteúdo e o acessem conhecendo, exatamente, qual versão do site está sendo utilizada, no caso de uma citação ou referência, por exemplo;
- e) acesso a conteúdo excluído ou modificado (*access to deleted or modified content*) - os arquivos da web disponibilizam sites que já foram excluídos ou alterados, de modo que os usuários podem visualizar facilmente conteúdos inacessíveis na web ao vivo, como são os casos das ferramentas *Wayback Machine* e *Memento Time Travel* (Memento [...], c2021) que permite a captura, acesso e visualização de versões anteriores de sites, e páginas da web, existentes em algum momento do passado;
- f) análise de tendências tecnológicas (*analysis of technology trends*) – *JavaScript*⁹ (*JavaScript*, [2022?], *HTML*, *Resource Description Framework* (*RDF*) (*Resource* [...], 2014) e mais outros formatos de arquivo, linguagens de programação e de marcação capturados em coleções de arquivos web, servem de linha temporal do desenvolvimento de tecnologias web. A análise em páginas coletadas pode mostrar transformações no uso de formatos da web ao longo do tempo, indicando tendências em marcação e formatação digital.

Niu (2012), baseada, em parte, nos casos de uso do IIIPC, definiu igualmente quais usos e funcionalidades que se esperam que os arquivos da web suportem para suprirem as necessidades dos seus usuários, que podem auxiliar a informar o design de funcionalidade de futuros arquivos da web a serem construídos como, ainda, a avaliar ou autoavaliar os arquivos da web já existentes.

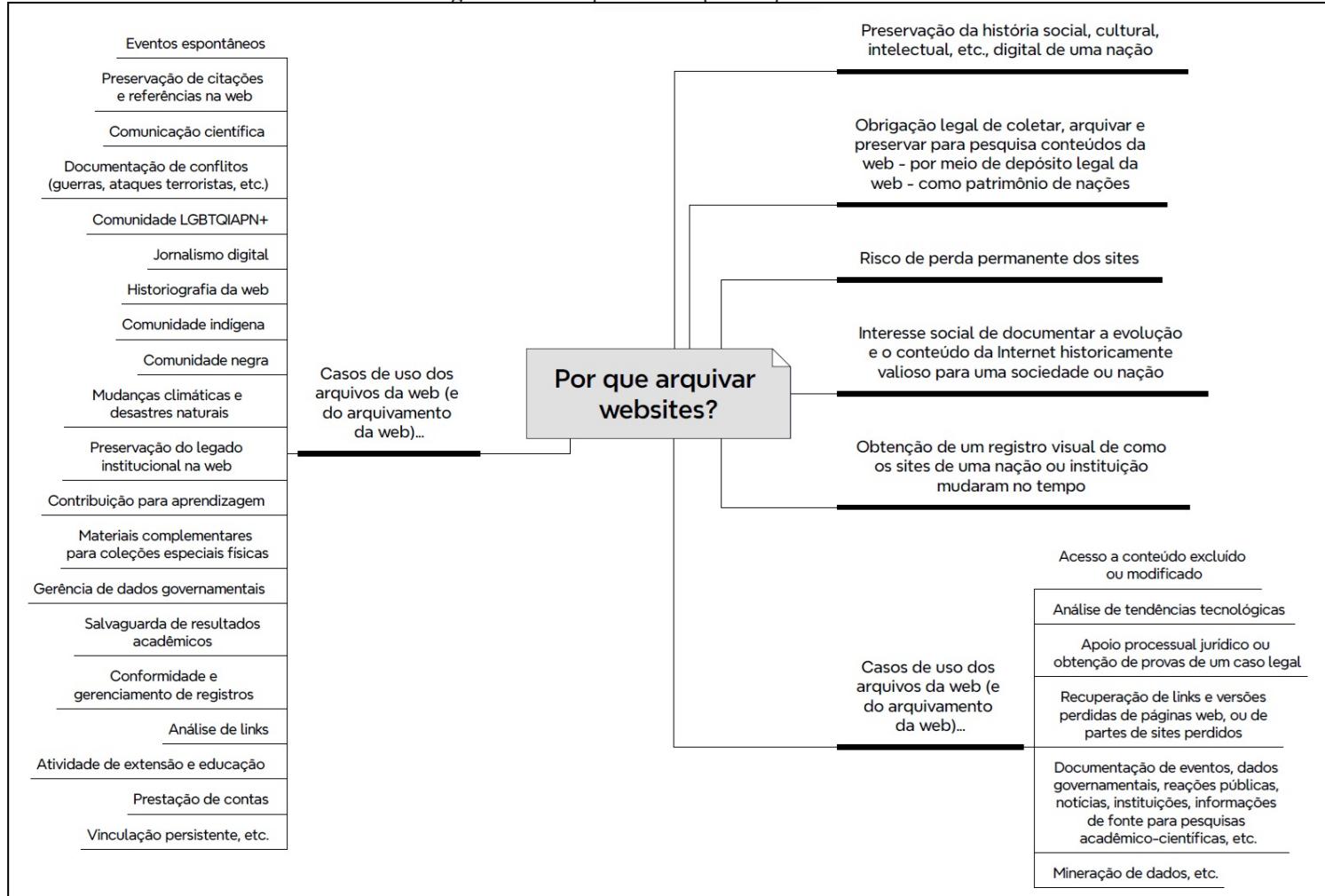
Dentre os usos e funcionalidades apresentadas pela autora, destacam-se:

- a) mineração de dados (*data mining*) - segundo a *Society of American Archivists* (Data mining , c2022b, tradução nossa), *data mining* possui relação com o “[...] processo de identificação de padrões previamente desconhecidos pela análise de relações em grandes quantidades de dados reunidos a partir de diferentes aplicações.”, e, em conformidade com a *International Organization for Standardization* (ISO, 2014, p. 2, tradução nossa) trata do “processo computacional que extrai padrões através da análise de dados quantitativos de diferentes perspectivas e dimensões, categorizando-os, e resumindo potenciais relacionamentos e impactos”. Assim, por meio dessa funcionalidade, o arquivo da web pode: apresentar gráficos que ilustrem como certos sites arquivados se associam a determinados eventos num período de tempo; fornecer informações de link para uma página da web arquivada (links de entrada, de saída e internos); permitir que usuários extraiam um subconjunto do arquivo da web, baseando-se em critérios, como idioma, formato de arquivo e metadados, podendo processar e analisar os dados no próprio arquivo da web, ou até exportar o subconjunto extraído, processando-o em outro lugar; e, preservar arquivos de *log* (*log files*) do site que contém informações quanto a sistemas operacionais, servidores web¹⁰, versões, etc. ;
- b) recuperação de pelo menos partes de sites perdidos - os usuários e os proprietários de sites podem usar arquivos da web para reconstruir uma certa versão de um site perdido, mantendo-se a estrutura do site original. Portanto, como apresentado a partir da revisão de literatura e análise de conteúdo das fontes de pesquisas bibliográficas e documentais sobre o tema em pauta, citadas concomitante durante as descrições, nas subseções anteriores, há várias e diferentes razões para se arquivar sites ou, ainda, para a realização do arquivamento da web e a criação de arquivos da web.

3 Discussões e resultados

A Figura 1, procura sintetizar as razões para o arquivamento da web, funcionando como representação gráfico-textual do resultado final desta pesquisa, que permite tecer algumas considerações antes da conclusão do trabalho.

Figura 1 - Principais razões para arquivar sites



Fonte: Elaborado pelos autores.

Procurou-se demonstrar, no presente trabalho, os atuais motivos para se arquivar sites. Esses motivos foram divididos em um primeiro conjunto de razões mais gerais, pelas quais instituições (sobretudo, bibliotecas e arquivos nacionais) devam desenvolver arquivos da web e, em um segundo conjunto, outros fatores que, de fato, faz com que essas instituições se engajem nessa atribuição.

No primeiro grupo, tem-se a percepção da urgência de se preservar os conteúdos web, pois, eles estão sendo perdidos veloz e diariamente, e que sua proteção salvaguarda o patrimônio histórico-cultural digital em risco, produzido pelas nações, pautando-se, inclusive, na missão social de arquivar esses conteúdos, enquanto extensão das tarefas clássicas atribuídas a instituições ligadas à preservação da memória, complementando-se coleções físicas, já existentes e tradicionalmente protegidas.

Cabendo, ainda, considerar os casos de uso dos arquivos da web (por exemplo), como potenciais fontes de dados permanentes para pesquisa e ensino, ou, adicionalmente, para fornecerem informações para análise de eventos e períodos de tempo, do presente e do passado, possibilitando, inclusive a previsão, projeção, de condições futuras (aspecto decisório baseado em dados concretos desses conteúdos).

Já, no segundo grupo, observa-se a questão da legislação, que pode: (1) conferir o direito e a proteção às instituições em capturar, abrigar e arquivar, sob a perspectiva de um depositário legal, juridicamente constituído, materiais da web; e, (2) impor, ou exigir que conteúdos publicados no passado (registros públicos de organizações, governos, etc.) sejam arquivados para conformidade normativa de acesso, inclusive como apoio material, de modo a servir de provas de defesa em processos judiciais.

Quanto a este último grupo, Santos (2020) traz uma interessante análise geral da legislação federal brasileira em vigor e das proposições legislativas que tem relação com a preservação de sítios web do país. Para o autor embora não haja uma legislação própria referente à proteção, captura e preservação de sites institucionais, e tendo por base na nossa legislação atual, considera ser possível defender que tal serviço seja atribuído às instituições de memória, via leis arquivísticas.

Contudo, o autor sinaliza que é preciso uma análise do tratamento a ser aplicado aos sítios institucionais, considerando-se a falta de consenso entre os profissionais de preservação digital, haja vista que as experiências de arquivos da web podem contemplar o viés bibliográfico, oriundo do conceito de depósito legal (por exemplo, o arquivo da web espanhola, da Biblioteca Nacional de

España (BNE, c2024)) ou o viés arquivístico de preservação do patrimônio digital de um país (por exemplo, o Arquivo.pt (2008) de Portugal).

4 Conclusões

De fato, cada caso de uso ligado aos arquivos da web pode revelar incentivos singulares para o arquivamento, assim como o mapeamento detalhado dessas motivações pode tornar claro o valor e a relevância do arquivamento da web. Entender os *insights* providos pelos arquivos da web justifica, em certa medida, os aspectos decisórios de futuros investimentos financeiros, políticos, científicos, etc., na preservação digital e no arquivamento da web, sobretudo, em países que ainda não detêm um arquivo da web nacional, como é caso do Brasil.

Além das possibilidades tratadas no presente trabalho, os futuros estudos nacionais que pretendem, da mesma forma, discutir as razões e as justificativas para o arquivamento da web, poderiam focar em outros casos de uso dos arquivos da web, incluindo, buscar compreender como estas estruturas seriam capazes de: (1) preservar as artes, a literatura, a música e outros bens que compõem o patrimônio histórico-cultural digital brasileiro; (2) proporcionar *insights* acerca das mudanças, transformações e evoluções da nossa sociedade na era da informação digital; (3) ajudar a fortalecer a defesa contra a censura ditatorial, assim como contra a manipulação da informação online, a desinformação produzida e/ou publicada na web pública e privada; (4) servir de fonte para pesquisas e inovações em múltiplas áreas em desenvolvimento, como inteligência artificial, por exemplo; (5) auxiliar a rastrear e a responsabilizar as ações antiéticas online (notícias falsas, comportamento indevido na web, etc.) de instituições, empresas, governos e pessoas; dentre outros.

Assim, considera-se que este trabalho atingiu o objetivo de apresentar as razões para se arquivar sites por meio da descrição de casos de uso para arquivos da web, assim como do arquivamento da web, almejando-se que a sistematização de iniciativas apresentadas nessas páginas, possa contribuir para o avanço de estudos desse assunto, sobretudo no território brasileiro.

Financiamento

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

Referências

ADAMS, Kimberly; FERNANDEZ, Sasha. Digital archivists race to preserve Ukrainian heritage. **Marketplace Tech**, [Saint Paul], 11 Mar. 2022. Disponível em: <https://bit.ly/467Mlw0>. Acesso em: 11 nov. 2023.

AGENCY FOR THE LEGAL DEPOSIT LIBRARIES (ALDL). Edimburgo, 2013. Disponível em: <https://www.legaldeposit.org.uk/>. Acesso em: 30 out. 2023.

ALNOAMANY, Yasmin; WEIGLE, Michele C.; NELSON, Michael L. Detecting off-topic pages within timemaps in web archives. **International Journal on Digital Libraries**, Berlim, v. 17, n. 3, p. 203-221, 2016. Disponível em: <https://doi.org/10.1007/s00799-016-0183-5>. Acesso em: 17 nov. 2023.

ARQUIVO.PT. [Arquivo da web portuguesa]. Lisboa, 2008. Disponível em: <https://arquivo.pt/>. Acesso em: 26 out. 2023.

ARQWEB. [Projeto piloto serviço de preservação de páginas web]. Brasília: IBICT, [2022]. Disponível em: <https://arqweb.ibict.br/pt-br/search/>. Acesso em: 26 out. 2023.

BARDIN, Laurence. **Análise de conteúdo**. São Paulo: Edições 70, 2016.

BAUCOM, Erin. Planning and implementing a sustainable digital preservation program. **Library Technology Reports**, Chicago, v. 55, n. 6, p. 22-27, 2019. Disponível em: <https://doi.org/10.5860/ltr.55n6>. Acesso em: 13 nov. 2023.

BIBLIOTECA DE CATALUNYA. Patrimoni digital de Catalunya (PADICAT). **l'Arxiu web de Catalunya**. Catalunya, c2011. Disponível em: <https://www.padicat.cat/ca>. Acesso em: 26 out. 2023.

BIBLIOTECA NACIONAL DE ESPAÑA (BNE). Archivo de la Web Española. Madrid, c2024. Disponível em: Archivo de la Web Española | Biblioteca Nacional de España (bne.es) Acesso em: 14 jan. 2024.

BIBLIOTHÈQUE NATIONALE DE FRANCE (BnF). Accueil. Collaborer. Déposer. **Qu'est-ce que le dépôt légal?** Paris, c2022. Disponível em: <https://www.bnf.fr/fr/quest-ce-que-le-depot-legal>. Acesso em: 18 nov. 2023.

BLACK LIVES MATTER GLOBAL NETWORK FOUNDATION.

#BlackLivesMatter. United States; United Kingdom; Canada, 2013. Disponível em: <https://blacklivesmatter.com/herstory/>. Acesso em: 16 nov. 2023.

BOERES, Sonia Araújo de Assis. Arquivamento da web: definições, estratégias, fluxos e iniciativas. **Revista Brasileira de Preservação Digital**, Campinas, São Paulo, v. 4, p. 1-15, 2023. Disponível em:
<https://doi.org/10.20396/rebpred.v4i00.17934> 51. Acesso em: 18 nov. 2023.

BROWN, Adrian. **Archiving websites: a practical guide for information management professionals**. London: Facet Publishing, c2006.

CADAVID, Jhonny Antonio Pabón. Evolution of legal deposit in New Zealand: from print to digital heritage. **International Federation of Library Associations and Institutions**, Haia, v. 43, n. 4, p. 379-390, 2017. Disponível em: <https://doi.org/10.1177/0340035217713763>. Acesso em: 1 nov. 2023.

CARDOSO, Clarice; CARTCHUCK, Ana. Justiça para George Floyd. **UOL**, São Paulo, c2024. Disponível em: George Floyd: Como negro morto pela polícia inspira luta antirracista (uol.com.br). Acesso em: 16 nov. 2023.

CAVALCANTE, Ricardo Bezerra; CALIXTO, Pedro; PINHEIRO, Marta Macedo Kerr. Análise de conteúdo: considerações gerais, relações com a pergunta de pesquisa, possibilidades e limitações do método. **Informação & Sociedade**, João Pessoa, v. 24, n. 1, p. 13-18, 2014. Disponível em:
<https://periodicos.ufpb.br/ojs/index.php/ies/article/view/10000/10871>. Acesso em: 18 nov. 2023.

COCCIOLO, Anthony. Community archives in the digital era: a case from the lgbt community. **Preservation, Digital Technology & Culture**, Berlim v. 45, n. 4, p. 157-165, 2016. Disponível em: <https://doi.org/10.1515/pdtc-2016-0018>. Acesso em: 18 nov. 2023.

COLUMBIA UNIVERSITY LIBRARIES. **Web archives at Columbia**. New York, c2021. Disponível em: <https://library.columbia.edu/collections/web-archives.html>. Acesso em: 26 out. 2023.

CONIFER. **Rhizome.org**. [New York], [2020?]. Disponível em:
<https://conifer.rhizome.org/>. Acesso em: 14 nov. 2023.

CONSELHO NACIONAL DE ARQUIVOS (CONARq) (Brasil). Assuntos. Câmaras técnicas consultivas. **Para definir diretrizes para a elaboração de estudos, proposições e soluções para a preservação de websites e mídias sociais**. Rio de Janeiro, 2022. Disponível em: <https://bit.ly/47noucy>. Acesso em: 11 nov. 2023.

CONSELHO NACIONAL DE ARQUIVOS (CONARq) (Brasil). Câmara Técnica de Documentos Eletrônicos (CTDE). **Glossário: documentos**

arquivísticos digitais. Versão 8. Rio de Janeiro, 2020. Disponível em: <https://bit.ly/3Sv1fcu>. Acesso em: 11 nov. 2023.

CONSELHO NACIONAL DE ARQUIVOS (CONARq) (Brasil). Diretrizes para a presunção de autenticidade de documentos arquivísticos digitais. Rio de Janeiro, 2012. Disponível em: <https://bit.ly/3vDJnCQ>. Acesso em: 12 jan. 2024.

CONSELHO NACIONAL DE ARQUIVOS (CONARq) (Brasil). Resolução n. 52, de 25 de agosto de 2023. Estabelece a política de preservação de websites e mídias sociais no âmbito do Sistema Nacional de Arquivos (SINAR). Rio de Janeiro, 2023b. Disponível em: <https://bit.ly/3O439xK>. Acesso em: 9 jan. 2024.

CONSELHO NACIONAL DE ARQUIVOS (CONARq) (Brasil). Câmara Técnica Consultiva Preservação de websites e Mídias Sociais. Resolução n. 53, de 25 de agosto de 2023. Requisitos mínimos de preservação para websites e mídias sociais. Rio de Janeiro, 2023a. Disponível em: <https://bit.ly/3HFuEdv>. Acesso em: 9 jan. 2024.

CONSELHO NACIONAL DE ARQUIVOS (CONARq) (Brasil). Sistema Nacional de Arquivos (SINAR). Rio de Janeiro, 2020. Disponível em: <https://www.gov.br/CONARq/pt-br/conexoes/sinar>. Acesso em: 9 jan. 2024.

COORDENAÇÃO DE APERFEIÇOAMENTO DE PESSOAL DE NÍVEL SUPERIOR (CAPES). Portal de Periódicos. Brasília: MEC, c2020. Disponível em: <https://www-periodicos-capes-gov-br.ez1.periodicos.capes.gov.br/index.php>? Acesso em: 27 out. 2023.

CORDEIRO, Alexander Magno *et al.* Revisão sistemática: uma revisão narrativa. **Comunicação científica**, Rio de Janeiro, v. 34, n. 6, p. 428-431, 2007. Disponível em: <https://doi.org/10.1590/S0100-69912007000600012>. Acesso em: 13 nov. 2023.

COSTA, Miguel; GOMES, Daniel; SILVA, Mário J. The evolution of web archiving. **International Journal on Digital Libraries**, Berlim, v. 18, n. 3, p. 191-205, 2017. Disponível em: <https://doi.org/10.1007/s00799-016-0171-9>. Acesso em: 17 nov. 2023.

CREATIVE COMMONS. In: SOCIETY of American Archivists. **Dictionary of archives terminology**. [Chicago], c2022a. Disponível em: <https://dictionary.archivists.org/entry/creative-commons.html>. Acesso em: 9 nov. 2023.

DATA MINING. In: SOCIETY of American Archivists. **Dictionary of archives terminology**. [Chicago], c2022b. Disponível em: <https://dictionary.archivists.org/entry/data-mining.html>. Acesso em: 22 nov. 2023.

DIGITAL PRESERVATION COALITION (DPC). **Digital preservation handbook.** Technical solutions and tools. 2th ed. [Glasgow], c2015. Disponível em: <https://bit.ly/48UkroT>. Acesso em: 20 nov. 2023.

DIGITAL PRESERVATION COALITION (DPC). [Glasgow], c2024. Disponível em: <https://www.dpconline.org/>. Acesso em: 26 out. 2023.

DURANTI, Luciana. The long-term preservation of the digital heritage: a case study of universities institutional repositories. **Italian Journal of Library and Information Science**, Macerata, v. 1, n. 1, p. 157-168, 2010. Disponível em: <https://doi.org/10.4403/jlis.it-12>. Acesso em: 22 nov. 2023.

FLORES, Daniel. Preservação de páginas web e redes sociais em cadeia de custódia: identificação, seleção e arquivamento. In: ENCONTRO NACIONAL DE MEMÓRIA DO PODER JUDICIÁRIO, 1., 2021, Brasília. **Anais** [...]. Brasília: Conselho Nacional de Justiça, 2021. 1 vídeo (29 min). Disponível em: <https://bit.ly/3tXLgK7>. Acesso em: 13 jan. 2024.

FREEDOM OF INFORMATION ACT STATUTE (FOIA). Washington, [2024]. Disponível em: <https://www.foia.gov/foia-statute.html>. Acesso em: 16 nov. 2023.

FREEDOM OF INFORMATION ACT STATUTE (FOIA). Washington, 2007. Collected by Stanford University Social Sciences Resource Group. Disponível em: <https://www.archive-it.org/collections/924>. Acesso em: 16 nov. 2023.

FORMENTON, Danilo. **Identificação de critérios de seleção de conteúdos para o arquivamento da Web.** 2023. Tese (Doutorado em Ciência, Tecnologia e Sociedade) - Centro de Educação e Ciências Humanas, Universidade Federal de São Carlos, São Carlos, 2023. Disponível em: <https://repositorio.ufscar.br/handle/ufscar/18426>. Acesso em: Acesso em: 3 fev. 2024.

FORMENTON, Danilo; GRACIOSO, Luciana de Souza. Padrões de metadados no arquivamento da web: recursos tecnológicos para a garantia da preservação digital de websites arquivados. **RDBCi: Revista Digital de Biblioteconomia e Ciência da Informação**, Campinas, v. 20, p. 1-29, 2022. Disponível em: <https://doi.org/10.20396/rdbc.v20i00.8666263>. Acesso em: 7 fev. 2024.

FORMENTON, Danilo; GRACIOSO, Luciana de Souza. Preservação digital: desafios, requisitos, estratégias e produção científica. **RDBCi: Revista Digital de Biblioteconomia e Ciência da Informação**, Campinas, v. 18, p. 1-27, 2020. Disponível em: <https://doi.org/10.20396/rdbc.v18i0.8659259>. Acesso em: 7 fev. 2024.

GIL, Antonio Carlos. **Como elaborar projetos de pesquisa.** 5. ed. São Paulo: Atlas, 2010.

GOOGLE SCHOLAR. Mountain View: Google, c2024. Disponível em: <https://scholar.google.com.br/>. Acesso em: 27 out. 2023.

HANZO. **We are the data navigators:** delivering a deeper perspective. London, c2024. Disponível em: <https://www.hanzo.co/>. Acesso em: 14 nov. 2023

HIRTLE, Peter B. Archival authenticity in a digital age. In: COUNCIL ON LIBRARY AND INFORMATION RESOURCES. Authenticity in a digital environment. May 2000. Washington: Council on Library and Information Resources, 2000. p. 8-23. Disponível em: <https://www.clir.org/wp-content/uploads/sites/6/pub92.pdf>. Acesso em: 16 jan. 2024.

INSTITUT NATIONAL DE L'AUDIOVISUEL (INA) (France). Paris, [c2024]. Disponível em: <https://www.ina.fr/institut-national-audiovisuel/collections-audiovisuelles/le-web-media>. Acesso em: 30 maio 2023.

INTERNATIONAL INTERNET PRESERVATION CONSORTIUM (IIPC). **Library of Congress web Archive.** United States, 2000. Disponível em: <https://netpreserve.org/about-us/members/library-congress/>. Acesso em: 30 maio 2023.

INTERNATIONAL INTERNET PRESERVATION CONSORTIUM (IIPC). **web archiving.** United States, c2024. Disponível em: <https://netpreserve.org/web-archiving/>. Acesso em: 21 nov. 2023.

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION (ISO). **ISO 16439:2014:** Information and documentation: methods and procedures for assessing the impact of libraries, Geneva: ISO, 2014. Disponível em: <https://www.iso.org/standard/56756.html>. Acesso em: 20 nov. 2023.

INTERNET ARCHIVE. **Archive-it.** [San Francisco], 2007. Collection by Stanford University, Social Sciences Resource Group - Bay area governments. Disponível em: <https://www.archive-it.org/collections/903>. Acesso em: 16 nov. 2023.

INTERNET ARCHIVE. **Archive-it.** [San Francisco], 2008. Collection by Silicon Valley Archives - Digital Games. Disponível em: <https://www.archive-it.org/collections/1023>. Acesso em: 16 nov. 2023.

INTERNET ARCHIVE. **Archive-it.** [San Francisco], 2010. Collection by University of Hawaii - Hawaii-Hawaiians. Disponível em: <https://archive-it.org/collections/1279>. Acesso em: 16 nov. 2023

INTERNET ARCHIVE. **Archive-it.** [San Francisco], 2011. Collection by Virginia Tech: Crisis, Tragedy, and Recovery Network - Japan Earthquake. Disponível em: <https://archive-it.org/collections/2438>. Acesso em: 16 nov. 2023.

INTERNET ARCHIVE. **Archive-it.** [San Francisco], 2014a. Disponível em: <https://archive-it.org/>. Acesso em: 16 nov. 2023.

INTERNET ARCHIVE. **Archive-it.** Projects. Projects and programs. Spontaneous event collections. [San Francisco], 2014b. Disponível em: <https://archive-it.org/blog/projects/spontaneous-events/>. Acesso em: 22 nov. 2023.

INTERNET ARCHIVE. **Archive-it.** [San Francisco], 2014d. Collection by Internet Archives Global Events - Ukraine Conflict. Disponível em: <https://archive-it.org/collections/4399>. Acesso em: 30 maio 2023.

INTERNET ARCHIVE. **Archive-it.** [San Francisco], 2014f. Collection by Internet Archives Global Events - #blacklivesmatter Web Archive. Disponível em: <https://archive-it.org/collections/4783>. Acesso em: 16 nov. 2023

INTERNET ARCHIVE. **Archive-it.** [San Francisco], 2015a. Stanford University website Collection. Disponível em: <https://archive-it.org/collections/5591>. Acesso em: 16 nov. 2023.

INTERNET ARCHIVE. **Archive-it.** [San Francisco], 2015b. Collection by Stanford University Archives - Patrick Suppes. Disponível em: <https://archive-it.org/collections/5605>. Acesso em: 16 nov. 2023.

INTERNET ARCHIVE. **Archive-it.** [San Francisco], 2015c. Collection by Stanford University Archives - Philip G. Zimbardo. Disponível em: <https://archive-it.org/collections/5604>. Acesso em: 16 nov. 2023.

INTERNET ARCHIVE. **Archive-it.** [San Francisco], 2015d. Collection by Stanford University Archives - Carolyn Bertozzi. Disponível em: <https://archive-it.org/collections/6434>. Acesso em: 16 nov. 2023.

INTERNET ARCHIVE. **Archive-it.** [San Francisco], 2015e. Collection by Stanford University Archives - Carl Djerassi. Disponível em: <https://archive-it.org/collections/5590>. Acesso em: 30 maio 2023.

INTERNET ARCHIVE. **Archive-it.** [San Francisco], 2015f. Collection by Stanford University Archives - Stanford University Student Organizations website Collection. Disponível em: <https://archive-it.org/collections/5593>. Acesso em: 30 maio 2023.

INTERNET ARCHIVE. **Archive-it.** [San Francisco], 2015g. Collection by Stanford University Archives - Center for Relationship Abuse Awareness. Disponível em: <https://archive-it.org/collections/6063>. Acesso em: 30 maio 2023.

INTERNET ARCHIVE. **Archive-it.** [San Francisco], 2015h. Collection by Internet Archives Global Events - Charlie Hebdo. Disponível em: <https://archive-it.org/collections/5190>. Acesso em: 30 maio 2023.

INTERNET ARCHIVE. **Archive-it.** [San Francisco], 2015i. Collection by UCLA - Southeast Asia LGBT web Archive. Disponível em: <https://archive-it.org/collections/6459>. Acesso em: 16 nov. 2023.

INTERNET ARCHIVE. **Archive-it.** [San Francisco], 2015j. Collection by University of California Berkeley - Archiving the LGBT web: Eastern Europe and Eurasia. Disponível em: <https://archive-it.org/collections/6165>. Acesso em: 16 nov. 2023.

INTERNET ARCHIVE. **Archive-it.** [San Francisco], 2016. Collection by University of Mississippi Meek School of Journalism - Hurricane Katrina blogs web collection. Disponível em: <https://archive-it.org/collections/7625>. Acesso em: 16 nov. 2023.

INTERNET ARCHIVE. **Archive-it.** [San Francisco], 2017. Collection by Lakehead University - Policing, Racism, and Indigenous People in Thunder Bay. Disponível em: <https://archive-it.org/collections/9394>. Acesso em: 16 nov. 2023.

INTERNET ARCHIVE. **Archive-it.** [San Francisco], 2020a. Collection by Stanford University Archives - Stanford University COVID-19 Response. Disponível em: <https://archive-it.org/collections/13658>. Acesso em: 30 maio 2023.

INTERNET ARCHIVE. **Archive-it.** [San Francisco], 2020b. Collection by Middlebury College - Community Responses to Anti-Black Racism and Police Violence. Disponível em: <https://archive-it.org/collections/14467>. Acesso em: 16 nov. 2023.

INTERNET ARCHIVE. **Internet Archive is a non-profit library of millions of free books, movies, software, music, websites, and more.** [San Francisco], 2014c. Disponível em: <https://archive.org/>. Acesso em: 26 out. 2023.

INTERNET ARCHIVE. **Wayback Machine.** [San Francisco], 2014e. Disponível em: <https://web.archive.org/>. Acesso em: 14 nov. 2023.

INTERNET ENGINEERING TASK FORCE (IETF). Wilmington, [2022]. Disponível em: <https://www.ietf.org/>. Acesso em: 26 out. 2023.

JAVASCRIPT. In: THE NATIONAL Archives UK. Richmond, [2022?]. Disponível em: <https://www.nationalarchives.gov.uk/webarchive/about/glossary/#j>. Acesso em: 21 nov. 2023.

KA‘AI-MAHUTA, Rachael. Preserving indigenous voices: web archiving in Aotearoa/New Zealand. **Interaction Design and Architecture(s) Journal - IxD&A**, Italy, n. 41, p. 24-30, 2019. Disponível em: <https://doi.org/10.55612/s-5002-041-002>. Acesso em: 13 nov. 2023.

LIBRARY OF CONGRESS. **Digital Collections**. Washington, 2018a. Collection Women's and Gender Studies Web Archive. Disponível em: <https://www.loc.gov/collections/womens-and-gender-studies-web-archive/about-this-collection/>. Acesso em: 14 nov. 2023.

LIBRARY OF CONGRESS. **Digital Collections**. Washington, 2018b. Collection LGBTQ+ Politics and Political Candidates web Archive. Disponível em: <https://www.loc.gov/collections/lgbtq-politics-and-political-candidates-web-archive/about-this-collection/#:~:text=The%20LGBTQ%2B%20Politics%20and%20Political,know%20local%20and%20state%20politics>. Acesso em: 30 maio 2023.

LIBRARY OF CONGRESS. **Digital Collections**. Washington, 2018c. Collection LGBTQ+ Studies web Archive. Disponível em: <https://www.loc.gov/collections/lgbtq-studies-web-archive/about-this-collection/#:~:text=The%20LGBTQ%2B%20Studies%20web%20Archive,%2C%20historical%20records%2C%20and%20more>. Acesso em: 30 maio 2023.

LIBRARY OF CONGRESS. **Digital Collections**. Washington, 2013. Collection Science Blogs web Archive. Disponível em: <https://www.loc.gov/collections/science-blogs-web-archive/about-this-collection/>. Acesso em: 30 maio 2023.

LIBRARY OF CONGRESS. **Library of Congress collections policy statements supplementary guidelines**: web archiving. Washington, July 2022a. Disponível em: <https://www.loc.gov/acq/devpol/webarchive.pdf>. Acesso em: 1 nov. 2023.

LIBRARY OF CONGRESS. Programs. **Web archiving**. About this program. Washington, [2022?b]. Disponível em: <https://www.loc.gov/programs/web-archiving/about-this-program/>. Acesso em: 1 nov. 2023.

LIBRARY OF CONGRESS. Programs. **Web archiving**. For site owners. Washington, [2022?c]. Disponível em: <https://www.loc.gov/programs/web-archiving/for-site-owners/>. Acesso em: 1 nov. 2023.

LUZ, Ana Javes. Preservação de sites oficiais: exemplos internacionais e o caso brasileiro. **Revista Brasileira de Preservação Digital**, Campinas, São Paulo, v. 3, p. 1-14, 2022. Disponível em: <https://doi.org/10.20396/rebpred.v3i00.1658780>. Acesso em: 20 nov. 2023.

MEMENTO Time Travel. [The TimeTravel service only displays links of Mementos to publicly accessible web archives. We do not own or store the

contente]. **Research Library of the Los Alamos National Laboratory**, United States, c2021. Disponível em: <https://timetravel.mementoweb.org/>. Acesso em: 14 nov. 2023.

NATIONAL LIBRARY OF AUSTRALIA. Pandora Australia's web Archive. Canberra, 2018. Collection ABC News. Disponível em: <http://pandora.nla.gov.au/col/16241>. Acesso em: 16 nov. 2023.

NATIONAL LIBRARY OF AUSTRALIA. Pandora Australia's web Archive. Canberra, [c2024a]. Collection Personal stories of Australians in war. Disponível em: <http://pandora.nla.gov.au/col/12925>. Acesso em: 16 nov. 2023.

NATIONAL LIBRARY OF AUSTRALIA. Pandora Australia's web Archive. Canberra, [c2024b]. Collection Historic gold mining sites. Disponível em: <http://pandora.nla.gov.au/col/13023>. Acesso em: 16 nov. 2023.

NATIONAL LIBRARY OF NEW ZEALAND. New Zealand web archive. Wellington, [2022?a]. Collections A-Z of our collections Disponível em: <https://natlib.govt.nz/collections/a-z/new-zealand-web-archive>. Acesso em: 7 nov. 2023.

NATIONAL LIBRARY OF NEW ZEALAND. Legal deposit. What's legal deposit? Our services for publishers and authors. Wellington, [2022?b]. Disponível em: <https://natlib.govt.nz/publishers-and-authors/legal-deposit/whats-legal-deposit>. Acesso em: 21 nov. 2023.

NATIONAL LIBRARY OF NEW ZEALAND. Web harvesting. Our services for publishers and authors. General information for publishers. New Zealand web Archive. Whole of domain web harvest. Wellington, [2022?c]. Disponível em: <https://natlib.govt.nz/publishers-and-authors/web-harvesting>. Acesso em: 7 nov. 2023.

NIU, Jinfang. Functionalities of web archives. **D-Lib Magazine**, United States, v. 18, n. 3/4, 2012. Disponível em: <https://doi.org/10.1045/march2012-niu2>. Acesso em: 11 nov. 2023.

ONLINE ARCHIVE OF CALIFORNIA (OAC). California, [ca. 2009]. Guide to the Stephen M. Cabrinety Collection in the History of Microcomputing, ca. 1975-1995. Disponível em: <http://www.oac.cdlib.org/findaid/ark:/13030/kt529018f2/>. Acesso em: 16 nov. 2023.

PENDSE, Liladhar. R. Archiving the Russian and east European lesbian, gay, bisexual, and transgender web, 2013: a pilot project. **Slavic & East European Information Resources**, United States, v. 15, n. 3, p. 182-196, 2014. Disponível em: <https://doi.org/10.1080/15228886.2014.930973>. Acesso em: 21 nov. 2023.

PENNOCK, Maureen. **Web Archiving**. DPC Technology Watch Report 13-01. Great Britain: Digital Preservation Coalition, 2013. Disponível em: <http://dx.doi.org/10.7207/twr13-01>. Acesso em: 13 nov. 2023.

PINHEIRO, Lena Vania Ribeiro; FERREZ, Helena Dodd. **Tesauro brasileiro de ciência da informação**. Rio de Janeiro; Brasília: Instituto Brasileiro de Informação em Ciência e Tecnologia, 2014. Disponível em: <https://bit.ly/4657GWF>. Acesso em: 11 nov. 2023.

RADINSKY, Kira; HORVITZ, Eric. Mining the web to predict future events. In: ASSOCIATION FOR COMPUTING MACHINERY (ACM); INTERNATIONAL CONFERENCE ON WEB SEARCH AND DATA MINING, 6., February 2013, Rome. **Proceedings** [...]. New York: ACM, 2013. p. 255-264.

REDE BRASILEIRA DE SERVIÇOS DE PRESERVAÇÃO DIGITAL. Rede Cariniana. [Rede de Pesquisa DRÍADE]. Brasília: IBCIT/Rede Cariniana, c2022. Disponível em: https://cariniana.ibict.br/?page_id=341. Acesso em: 26 out. 2023.

RÉPUBLIQUE FRANÇAISE. LOI n. 2006-961 du 1er août 2006 relative au droit d'auteur et aux droits voisins dans la société de l'information. **Journal Officiel**: Paris, n. 178 du 3 août 2006. Disponível em: <https://www.legifrance.gouv.fr/jorf/id/JORFTEXT000000266350>. Acesso em: 30 out. 2023

RÉPUBLIQUE FRANÇAISE. Code de la propriété intellectuelle. Paris, **Légifrance**, version en vigueur au 16 février 2024 Disponível em: <https://www.legifrance.gouv.fr/codes/id/LEGITEXT000006069414/>. Acesso em: 30 out. 2023.

RESOURCE Description Framework. In: SEMANTIC Web Standard. Wakefield, **W3C**, 2014. Disponível em: <https://www.w3.org/RDF/>. Acesso em: 14 nov. 2023.

REYNOLDS, Emily. **Web archiving use cases**. Washington: Library of Congress, UMSI, ASB13, Mar., 2013. Disponível em: https://netpreserve.org/resources/IIPC_archive-UseCases_Final.pdf. Acesso em: 20 nov. 2023.

ROCKEMBACH, Moisés. Arquivamento da web: estudos de caso internacionais e o caso brasileiro. **RDBCi: Revista Digital Biblioteconomia Ciência da Informação**, Campinas, v. 16, n. 1, p. 7-24, 2018. Disponível em: <https://doi.org/10.20396/rdbcii.v16i1.8648747>. Acesso em: 21 nov. 2023.

ROCKEMBACH, Moises; PAVÃO, Caterina Marta Groposo. Políticas e tecnologias de preservação digital no arquivamento da web. **RICI: Revista Ibero-Americana de Ciência da Informação**, Brasília, v. 11, n. 1, p. 168-182,

2018. Disponível em: <https://doi.org/10.26512/rici.v11.n1.2018.8473>. Acesso em: 22 nov. 2023.

ROCKEMBACH, Moisés; SERRANO, Anabela. Climate change and web archives: an ibero-american study based on the portuguese and brazilian contexts. **Records Management Journal**, United Kingdom, v. 31, n. 3, p. 222-239, 2021. Disponível em: <https://doi.org/10.1108/RMJ-11-2020-0039>. Acesso em: 13 nov. 2023.

RODRIGUES, Vander Luis Duarte; ROCKEMBACH, Moisés. Arquivos da web como fonte historiográfica. **RDBCi: Revista Digital Biblioteconomia Ciência da Informação**, Campinas, v. 19, p. 2-6, 2021. Disponível em: <https://doi.org/10.20396/rdbcii.v19i00.8663680>. Acesso em: 21 nov. 2023.

ROLLASON-CASS, Sylvie; REED, Scott. Living movements, living archives: selecting and archiving web content during times of social unrest. **New Review of Information Networking**, London, v. 20, p. 241-247, 2015. Disponível em: <https://doi.org/10.1080/13614576.2015.1114839>. Acesso em: 21 nov. 2023.

SANTOS, Vanderlei Batista dos. Arquivamento web: legislação correlata. **Revista Brasileira de Preservação Digital**, Campinas, v. 1, p. 1-11, 2020. Disponível em: <https://doi.org/10.20396/rebpred.v1i00.14800>. Acesso em: 9 jan. 2024.

SAVING UKRAINIAN CULTURAL HERITAGE ONLINE (SUCHO). [S. l.], 2022a. Disponível em: <https://archive-it.org/collections/4783>. Acesso em: 16 nov. 2023.

SAVING UKRAINIAN CULTURAL HERITAGE ONLINE (SUCHO). **Archives**. [S. l.], 2022b. Disponível em: <https://www.sucho.org/archives>. Acesso em: 30 maio 2023

SCHAFFER, Valérie *et al.* Paris and Nice terrorist attacks: exploring twitter and web archives. **Media, War & Conflict**, New Jersey, v. 12, n. 2, p 153-170, 2019. Disponível em: <https://doi.org/10.1177/1750635219839382>. Acesso em: 7 nov. 2023.

SCIENTIFIC ELECTRONIC LIBRARY ONLINE (SCIELO). São Paulo, c2024. Disponível em: <https://scielo.org/>. Acesso em: 27 out. 2023.

SERRANO, Jody. How to stop Ukrainian websites from vanishing during war. **Gizmodo**, New York, 4 Feb. 2022. Disponível em: <https://bit.ly/47p6O0g>. Acesso em: 12 nov. 2023.

SEVERINO, Antônio Joaquim. **Metodologia do trabalho científico**. 24. ed. rev. e atual. São Paulo: Cortez, 2016.

STANFORD UNIVERSITY. Libraries. **Web archiving**. Stanford: Stanford University, Dec. 2016. Disponível em:
<https://swap.stanford.edu/was/20160912010919/https://library.stanford.edu/projects/web-archiving>. Acesso em: 22 nov. 2023.

STANFORD UNIVERSITY. **Graduate School of Business**. Stanford, [2024]. Disponível em: <https://www.gsb.stanford.edu/faculty-research/case-studies>. Acesso em: 16 nov. 2023.

TERRADA, Gabriela Ayres Ferreira. **Preservação digital da web: uma reflexão sobre políticas e práticas**. 2022. Dissertação (Mestrado em Ciência da Informação) - Instituto de Arte e Comunicação Social, Universidade Federal Fluminense, Rio de Janeiro, 2022.

THE NATIONAL ARCHIVES (United Kingdom). The Legal Deposit Libraries (Non-Print Works) Regulations 2013 Richmond, **legislation.gov.uk**, 2013. Disponível em:
<https://www.legislation.gov.uk/ukdsi/2013/9780111533703/contents>. Acesso em: 30 out. 2023

THE NATIONAL ARCHIVES (United Kingdom). **UK Government Web Archive**. About the UK Government web Archive. Richmond, [2024]. Disponível em: <https://www.nationalarchives.gov.uk/websarchive/about/>. Acesso em: 30 out. 2023.

UNITED STATES. 107th United States Congress. H.R.2458. E-government act of 2002. **Public law 107–347**. [Washington], Dec. 17 2002. Disponível em:
<https://www.congress.gov/107/plaws/publ347/PLAW-107publ347.pdf>. Acesso em: 20 nov. 2023.

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL (UFRGS). **Núcleo de Pesquisa em Arquivamento da Web e Preservação Digital (NUAWEB)**. Porto Alegre: UFRGS, 2017.

VIRGINIA TECH: crisis, tragedy, and recovery network. Virginia, c2024. Disponível em: <https://www.vt.edu/>. Acesso em: 14 nov. 2023.

WORLD WIDE WEB CONSORTIUM (W3C). **Making the web work**. Wakefield, c2024. Disponível em: <https://www.w3.org/>. Acesso em: 26 out. 2023.

Web archiving: systematization of reasons and justifications for archiving websites

Abstract: The objective of this study is to present reasons for archiving websites through use cases of web archives and web archiving. It is

characterized as exploratory research that adopts the bibliographic and documentary method based on a literature review on the subject. Content analysis was applied to the review of current national and international scientific productions in Information Science, as well as official websites, policies, and guidelines of web archives around the world. It was concluded that there are broader motivations for institutions to develop web archives, including, among such motivations, the urgency of preserving web content, due to the risk of losing this content, thus safeguarding the digital heritage produced by nations and the countless cases of use of web archives. In addition, however, the issue of legislation is effectively recognized, which can oblige and encourage these initiatives, granting the right and protection of these contents to institutions, delegating the functions of capturing, sheltering, and archiving web materials, whether in the role of a legal depository, aiming to: preserve social memory; or, for regulatory compliance and support in legal proceedings. Understanding the insights provided by web archives justifies, to a certain extent, future financial, political, scientific, etc. investments in digital preservation and web archiving, especially in countries that do not have a national web archive, as in this case from Brazil.

Keywords: web archiving; digital preservation; websites; justifications; Information Science

Recebido: 12/12/2023

Aceito: 06/02/2024

Declaração de autoria:

Concepção e elaboração do estudo: Formenton, Danilo.

Coleta de dados: Formenton, Danilo.

Análise e interpretação de dados: Formenton, Danilo.

Redação: Formenton, Danilo.

Revisão crítica do manuscrito: Formenton, Danilo. Gracioso, Luciana de Souza.

Como citar

FORMENTON, Danilo; GRACIOSO, Luciana de Souza. Arquivamento da web: sistematização de razões e justificativas para arquivar sites. **Em Questão**, Porto Alegre, v. 30, e-137396, 2024. DOI: <https://doi.org/10.1590/1808-5245.30.137396>



- ¹ O artigo origina-se de Tese de Doutorado intitulada “Identificação de critérios de seleção de conteúdos para o arquivamento da web”, defendida em 2023, na Universidade Federal de São Carlos (UFSCar), apresentando mudanças em relação ao texto original (Formenton, 2023).
- ² Documento arquivístico é uma expressão que tem o sentido de “documento produzido (elaborado ou recebido), no curso de uma atividade prática, como instrumento ou resultado de tal atividade, e retido para ação ou referência.” (Conselho Nacional de Arquivos, 2020, p. 24).
- ³ CONSELHO NACIONAL DE ARQUIVOS (Brasil). Diretrizes para a presunção de autenticidade de documentos arquivísticos digitais. Rio de Janeiro, RJ: Conarq, 2012. *Apud Flores* (2021).
- ⁴ Cadeia de custódia ininterrupta se refere à “[...] linha contínua de custodiadores de documentos arquivísticos (desde o seu produtor até o seu legítimo sucessor) pela qual se assegura que [...]” tanto “[...] esses documentos são os mesmos desde o início [...]” como “[...] não sofreram nenhum processo de alteração e, portanto, são autênticos.” (Conselho Nacional de Arquivos, 2012, p. 1).
- ⁵ CONSELHO NACIONAL DE ARQUIVOS (Brasil). Diretrizes para a presunção de autenticidade de documentos arquivísticos digitais. Rio de Janeiro, RJ: Conarq, 2012. *Apud Flores* (2021).
- ⁶ Existem seis Bibliotecas de Depósito Legal do Reino Unido (do inglês UK *Legal Deposit Libraries*) que fazem parte do sistema de depósito legal do país, a saber: *British Library; National Library of Scotland; Bodleian Libraries, University of Oxford; Cambridge University Library; National Library of Wales; e Library of Trinity College Dublin, University of Dublin* (ALDL, 2013).
- ⁷ Em 25 de maio de 2020, o afro-americano George Perry Floyd Jr. foi assassinado em Minneapolis, estado norte-americano de Minnesota, depois que o policial branco, Derek Chauvin, o estrangulou se ajoelhando no seu pescoço durante uma abordagem. Após a morte de Floyd, uma onda de protestos e debates antirracistas - e contra a violência policial - foi gerada nos Estados Unidos, difundindo-se por todo o mundo. O julgamento de Chauvin teve início em 2021, culminando em sua condenação (Cardoso; Cartchuck, c2024).
- ⁸ *Black Lives Matter* (em português: Vidas Negras Importam) é um movimento social, iniciado nos Estados Unidos e difundido ao redor do mundo, pelo qual se defende a luta contra a brutalidade policial, a discriminação racial dirigida às comunidades negras, etc. Este movimento se iniciou em 2013, quando houve o uso da hashtag *#BlackLivesMatter* nas mídias sociais, após a absolvição de George Zimmerman pela morte do afro-americano Trayvon Martin, em Sanford, Flórida, ganhando mais força a partir de 2014, após as mortes dos afro-americanos: Michael Brown, pelo policial branco, Darren Wilson, em Ferguson, Missouri; e, Eric Garner, por policiais de Nova Iorque, gerando uma grande onda de protestos e manifestações. (*Black Lives Matter Global Network Foundation*, 2013).
- ⁹ *JavaScript* se refere à “[...] uma linguagem de programação de computador comumente usada para criar efeitos interativos dentro de navegadores web.” (*JavaScript*, [2022?], tradução nossa) ou, de acordo com Brown (c2006, p. XII, tradução nossa), remete à “[...] uma linguagem de *script* orientada a objetos, comumente usada para adicionar funcionalidade às páginas web.”
- ¹⁰ Servidor web (*web server*) é “um programa de computador que recebe solicitações HTTP de clientes (em geral navegadores web), e ‘fornece’ o conteúdo web solicitado para eles.”, porém “o termo também pode ser aplicado ao computador no qual o software do servidor web está sendo executado.” (Brown, c2006, p. 14, tradução nossa).