

ARQUIVAMENTO DA *WEB*: DEFINIÇÕES, ESTRATÉGIAS, FLUXOS E INICIATIVAS

WEB ARCHIVING: DEFINITIONS, STRATEGIES, FLOWS, AND INITIATIVES



Resumo

Introdução: A importância do arquivamento da *web* está em preservar e dar acesso, em longo prazo, às informações digitais publicadas na internet. Com a crescente quantidade de conteúdo digital, há um risco real de que informações valiosas sejam perdidas, corrompidas ou tornem-se inacessíveis, devido às mudanças tecnológicas, atualizações de plataformas, ou até mesmo exclusões acidentais ou intencionais. O arquivamento da *web* é fundamental para a preservação digital e para o acesso contínuo das informações institucionais, garantindo a disponibilidade, confiabilidade e continuidade do patrimônio digital da sociedade, e para o avanço da pesquisa, educação, transparência e preservação da história digital. **Objetivo:** Apresentar o tema, definindo arquivamento da *web*, mostrando como o processo é feito, as estratégias, os fluxos, e as iniciativas internacionais e nacionais, entre elas, o projeto ARQWEB, do Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT) e da Rede Cariniana. **Metodologia:** Pesquisa exploratória descritiva, utilizando o repositório de dados E-LIS e outras fontes, por meio da expressão *iniciativas e estratégias para o arquivamento da web, em português e inglês*. **Resultados:** Foram apresentadas sete iniciativas internacionais em variados contextos (bibliotecas e arquivos nacionais, organizações governamentais e sem fins lucrativos), e duas nacionais, onde pode-se destacar, o ARQWEB, que arquivará sítios de instituições parceiras e governamentais. **Conclusão:** Ainda são incipientes as iniciativas nacionais de arquivamento *web*, comparadas às internacionais, o que mostra possibilidades de estudo e a importância de estabelecer políticas, ter estratégias, e seguir fluxos para o arquivamento *web* nas instituições, garantindo qualidade na coleta e no resultado.

Palavras-chave: Arquivamento da *web*; Preservação digital; Estratégias de arquivamento; Fluxos de arquivamento da *web*; ARQWEB.



Sonia Araújo de Assis Boeres

Nome da Instituição

E-mail: sonia.boeres@gmail.com

Brasília – DF – Brasil



Autor [não preencher]

Nome da Instituição

E-mail: [xxxxxxxxxx](#)

Cidade – Estado - País



Abstract

Introduction: The importance of web archiving is to preserve and provide long-term access to digital information published on the internet. With the increasing amount of digital content, there is a real risk that valuable information will be lost, corrupted or become inaccessible, due to technological changes, platform upgrades, or even accidental or intentional deletions. Web archiving is essential for digital preservation and continuous access to institutional information, ensuring the availability, reliability, and continuity of society's digital heritage, and for advancing research, education, transparency, and preservation of digital history. **Objective:** To present the theme, defining web archiving, showing how the process is done, the strategies, flows, and international and national initiatives, among them, the ARQWEB project, from the Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT) e da Rede Cariniana. **Methodology:** Descriptive exploratory research, using the E-LIS data repository and other sources, through the expression *initiatives and strategies for archiving the web*, both in Portuguese and in English. **Results:** Seven international initiatives were presented in various contexts (national libraries and archives, governmental and non-profit organizations), and two national ones, where ARQWEB stands out, which will archive sites of partners and governmental institutions. **Conclusion:** National web archiving initiatives are still incipient, compared to international ones, which shows study possibilities and the importance of establishing policies, having strategies, and following flows for web archiving in institutions, guaranteeing quality in the collection and in the result.

Keywords: Web archiving; Digital preservation; Archiving strategies; Web archiving flows; ARQWEB.

LICENÇA DE USO

Os autores cedem à [Revista Brasileira de Preservação Digital](#) os direitos exclusivos de primeira publicação, com o trabalho simultaneamente licenciado sob a Licença Creative Commons Attribution (CC BY) 4.0 International. Esta licença permite que terceiros remixem, adaptem e criem a partir do trabalho publicado, atribuindo o devido crédito de autoria e publicação inicial neste periódico. Os autores têm autorização para assumir contratos adicionais separadamente, para distribuição não exclusiva da versão do trabalho publicada neste periódico (ex.: publicar em repositório institucional, em site pessoal, publicar uma tradução, ou como capítulo de livro), com reconhecimento de autoria e publicação inicial neste periódico.

PUBLISHERS

Universidade Estadual de Campinas – Sistema de Bibliotecas / Instituto Brasileiro de Informação em Ciência e Tecnologia – Rede Brasileira de Serviços de Preservação Digital – Cariniana. As ideias expressadas neste artigo são de responsabilidade de seus autores, não representando, necessariamente, a opinião dos editores ou da universidade.

EDITORES

Gildenir Carolino Santos, Miguel Angel Márdero Arellano.

CREDIT

Submetido em: 02/05/2023 – Aceito em: 15/05/2023 – Publicado em: 10/06/2023

1 INTRODUÇÃO

Preservar a informação digital da *web* tem se tornado uma preocupação crescente na era digital em que vivemos. Com o rápido avanço da tecnologia e a constante evolução da internet, muitas informações importantes estão em constante risco de serem perdidas ou tornarem-se inacessíveis. É nesse contexto que surgem o arquivamento da *web*, sistemas computacionais criado com o objetivo de coletar, armazenar e disponibilizar informações retrospectivas da *World Wide Web*, permitindo o acesso a longo prazo para pesquisadores, e garantindo a preservação da história digital (GOMES 2010).

A importância deste tipo de arquivamento reside na necessidade de preservação e acesso a longo prazo de informações digitais publicadas na internet. Com a crescente quantidade de conteúdo digital disponibilizado *online* há um risco real de que informações valiosas sejam perdidas, corrompidas ou tornem-se inacessíveis devido a mudanças tecnológicas, atualizações de plataformas ou até mesmo exclusões acidentais ou intencionais. Diante desses aspectos, o arquivamento da *web* é fundamental para a preservação e acesso às informações digitais, garantindo a disponibilidade, confiabilidade e continuidade do patrimônio digital da sociedade, bem como para o avanço da pesquisa, educação, transparência e preservação da história digital.

O presente artigo tem como objetivo apresentar o tema do arquivamento da *web* considerando sua importância, desafios e iniciativas globais e locais nessa área, destacando a necessidade de preservar a informação digital em um mundo cada vez mais dependente da internet.

2 ARQUIVAMENTO DA WEB

De acordo com Rockembach (2018) para entender o arquivamento da *web* tem-se que verificar o que é a *World Wide Web* (*web*). Ela nasceu em 1989, resultado da proposta de Bernes-Lee para suprir a necessidade de compartilhamento de informações por cientistas, e a primeira página *web* foi publicada em 20 de dezembro de 1990. A proposta inicial da *web* trazia o conceito de organizar, dar acesso à informação e evitar a perda de detalhes importantes dos projetos desenvolvidos no Conseil Européen pour la Recherche Nucléaire (CERN). (ROCKEMBACH, 2018).

Ligado a este conceito de rede de computadores, o *World Wide Web Consortium* (W3C), criado em 1994, discute padrões e diretrizes para garantir o crescimento da *web* em longo prazo, fundamentado em uma *web* aberta e colaborativa. (ROCKEMBACH, 2018). A preocupação da W3C vai ao encontro das ideias de Costa, Gomes e Silva (2016), que revelam que 80% das páginas *web* não estão disponíveis na sua forma original após um (1) ano; 13% das referências da *web* em artigos acadêmicos desaparecem após 27 meses, e 11% das postagens em redes sociais são perdidas após um (1) ano. Sendo assim, o arquivamento da *web* emerge como uma necessidade para reduzir esse impacto, e se faz necessária uma compreensão de como ele funciona. Para Silva (2019), o arquivamento da *web* é o processo de coleta, armazenamento e disponibilização de conteúdo digitais para fins de preservação histórica e pesquisa. Esse processo é realizado por meio de ferramentas automatizadas, conhecidas como "*crawlers*", que rastreiam a *web* e coletam conteúdo para armazenamento.

Conforme Gomes (2010), as três etapas de arquivamento da *web* envolvem recolher a informação, indexar, e disponibilizar serviços de pesquisa e acesso, sendo que a primeira etapa se subdivide em coletar o arquivo, armazená-lo, extrair os endereços para outros arquivos a partir dos *hiperlinks*, e inserir os novos endereços descobertos para a recolha.

O arquivamento da *web* pode ser visto como um "*workflow*" através do qual os recursos são selecionados, coletados, preservados, e disponibilizado para o usuário final. Para Rockembach (2018) a seleção e coleta são realizadas de forma contínua, e isso poderá levar em consideração uma série de fatores, como o conteúdo a ser coletado, se os *links* externos ao *site* selecionado também serão coletados, e a frequência da coleta. Para o autor, considerar os fatores citados tem como objetivo garantir a qualidade da coleta para assegurar um melhor arquivamento e recuperação dessa informação.

O fluxo de coleta do conteúdo da *web* pode ser feito através do envio das informações para o arquivamento, ou através de coleta ativa pela instituição responsável pelo arquivamento. Para Gomes (2010) a automatização para a coleta ativa é fundamental, visto o crescimento acelerado do conteúdo disponibilizado na internet. Nessa automatização, ao considerar a implementação de sistemas de arquivamento da *web*, o uso dos metadados é um ponto muito importante na

informação digital, eles são informações estruturadas que descrevem, localizam, gerenciam, recuperam e acessam recursos de informação digital. (MELO, 2020, p. 69 e 70). Eles ainda apoiam diferentes funções, como o acesso, registro de contexto e proveniência de objetos, documentação de ações e políticas. Quanto aos formatos, os arquivos da *web* podem ter uma ampla variedade, sejam páginas de conteúdo *online* textual, imagens, vídeos, arquivos em formato *PDF* e outros.

Para arquivamento da *web* desse conteúdo usam-se diferentes formatos de armazenamento e técnicas de compactação. (KHAN, RAHMAN, 2019). O formato ARC começou a ser usado desde meados dos anos 1990, em 2008 a Internet Organization for Standardization (ISO) especificou o formato WARC como padrão oficial para arquivos *web*, e em 2009 a organização Internet Archive adotou o formato ARC para salvar os arquivos de *websites*. (GOMES; MIRANDA; COSTA, 2011). Em 2011 foi identificado que cerca de 54% das iniciativas de arquivamento da *web* utilizavam os formatos ARC e WARC para arquivar conteúdo. (GOMES; MIRANDA; COSTA, 2011, apud MELO, 2020, p. 70). O formato WARC é usado internacionalmente seguindo a ISO 28500 e incentivado por autores como Rosa, Craveiro e Domingues (2017), Alam *et al.* (2019), Schafer e Winters (2020) e Rockembach (2021). Sobre o uso do arquivamento da *web* deve-se ir além da custódia e da preservação. (SILVA, et al., 1999). Devem-se compreender os usos potenciais desta tecnologia e do conteúdo armazenado, para os mais diversos fins, principalmente na recuperação de memórias pessoais e organizacionais, complementa Rockembach (2018).

Quanto às questões legais envolvendo o arquivamento da *web*, ainda falta o consenso sobre os cuidados que se deve ter. Destacam-se alguns destes problemas legais, como os direitos do autor e demais responsabilidades pelos conteúdos disponibilizados. A European Union Web Archive, por exemplo, captura apenas conteúdo que está disponível gratuitamente no domínio público, e orienta os gerentes de sítios da *web* a otimizar a qualidade das versões arquivadas, criar *sites* “preserváveis”, remover o conteúdo que não deve ser preservado (e estar acessível) em longo prazo, por motivos de direitos de propriedade intelectual, confidencialidade, privacidade, proteção de dados etc. (GUIDELINES, 2021). Para eles há circunstâncias em que pode ser necessário ocultar publicamente as páginas da *web*, a instituição

sugere que assim o seja quando a página incluir informações pessoais ou sensíveis, conforme definidas pelo Regulamento 2018/1725:

- material protegido por direitos autorais para os quais eles não são mantidos;
- material ou mensagens difamatórias ou obscenas;
- conteúdos de páginas que podem causar dificuldades administrativas ao proprietário do *site*;
- página publicada de boa-fé, mas que as circunstâncias mudaram, e sua remoção foi considerada apropriada;
- se a página foi publicada com erro e precisa ser corrigida.

Nas instituições algumas demandas de preservação e arquivamento de conteúdo chegam da comunidade, mas é necessário lembrar a importância de se ter um profissional da informação responsável por fazer a curadoria do processo, e selecionar segundo critérios pré-estabelecidos, como o de estar de acordo com a missão da instituição. Ele também deve ser capacitado e estar atualizado e atento ao tempo necessário para o descarte, afinal, mesmo a nuvem tem espaço limitado. Estas e outras discussões devem estar registradas na política de preservação digital da instituição. (BOERES, 2017).

3 INICIATIVAS INTERNACIONAIS E NACIONAIS DE ARQUIVAMENTO DA WEB

Variadas e diferentes são as iniciativas entre os países, a seguir começa-se pelas iniciativas internacionais, e logo em seguida as nacionais.

3.1 *Iniciativas Internacionais*

Quanto às iniciativas internacionais de preservação da *web* pode-se destacar o International Internet Preservation Consortium (IIPC), criado pela Biblioteca Nacional da França em 2003. Segundo Melo (2020, p. 36) o IIPC tem por objetivo a coleta de conteúdo da internet para as futuras gerações, promovendo o intercâmbio global e as relações internacionais. Ele promove o desenvolvimento e uso de ferramentas, técnicas e padrões comuns que incentivem e apoiem bibliotecas nacionais, assim como organizações de pesquisa para tratarem do arquivamento e preservação da

internet. Em 2020 o consórcio contava com aproximadamente 53 membros de mais de 35 países, incluindo bibliotecas e arquivos nacionais.

Entre outras iniciativas pioneiras no arquivamento da *web* destacam-se a Internet Archive, o projeto da Biblioteca Nacional Australiana PANDORA e a plataforma britânica UK Web Archive, que se consolidou como arquivamento da *web* nacional em 2004. A Internet Archive, um dos membros fundadores do IIPC, foi um dos primeiros na coleta dos arquivos da *web*, e vem arquivando sítios desde 1996, se consolidando como a principal plataforma de arquivamento da internet no mundo (MASANÈS, 2006, apud MELO, 2020, p. 37). Em 2002, a *Internet Archive* lançou o Heritrix, um rastreador da *web* de código aberto, ferramenta que captura o conteúdo da *World Wide Web* e que se tornou a tecnologia de rastreamento de arquivos da *web* mais popular e difundida atualmente. (MELO, 2020, p. 37). Os resultados dos rastreamentos são armazenados em um arquivo WebARChive (WARC), um formato desenvolvido pelo IIPC, que em 2009 foi adotado como extensão padrão para arquivos *web* (MAEMURA et.al, 2018, apud MELO, 2020, p.37), definido na ISO 28500. Outra ferramenta em destaque, desenvolvida pela Internet Archive, foi a Wayback Machine, que permite acessar como um determinado sítio era em um dado momento no tempo.

De acordo com Rockembach (2018), iniciativas como a da Biblioteca Regional da Catalunha e a da Biblioteca Nacional da França arquivam domínios específicos. A primeira arquiva os domínios ".cat" e a segunda os domínios ".fr" e de territórios franceses como ".re" (Reunião), ".nc" (Nova Caledônia) e ".gf" (Guiana Francesa). No caso do Reino Unido, a iniciativa é colaborativa e liderada pela Biblioteca Britânica, conhecida como UK Web Archive (UKWA). O objetivo é coletar e preservar os sítios britânicos com domínio ".uk", principalmente os governamentais de interesse público. (MELO, 2020, p. 38). Já em Portugal, a iniciativa Arquivo.pt, que teve início em 2008, permite pesquisar e aceder a páginas da *web* portuguesa e lusófona, arquivadas desde 1996.

3.2 Iniciativas Nacionais

Para Luz (2022), ao analisar as iniciativas internacionais, percebe-se que há uma preocupação em promover a salvaguarda dos conteúdos da *web* com objetivo de garantir o acesso à informação de interesse público. Contudo, quando se olha para o

panorama nacional percebe-se uma carência de iniciativas oficiais. Esse panorama já havia sido percebido a partir de levantamento e análise de estudos realizado por Rockembach sobre arquivamento da *web* na base de dados Scopus, compreendendo um intervalo de 2002 a 2016, onde constatou-se, até então, não haver programas permanentes de arquivamento da *web* em andamento no Brasil (ROCKEMBACH, 2018, apud LUZ, 2022).

A partir desse levantamento algumas iniciativas nacionais têm surgido, podendo-se destacar o Núcleo de Pesquisa em Arquivamento da *Web* e Preservação Digital (NUAWEB). Criado em 2017 pela Universidade Federal do Rio Grande do Sul, ele tem como objetivo investigar características do arquivamento da *web* por meio de iniciativas nacionais e internacionais, lidando tanto com as políticas, quanto com as tecnologias envolvidas no processo. (NUAWEB, 2023). Segundo Melo (2020, p. 44), "o NUAWEB está desenvolvendo dois projetos de pesquisa simultaneamente: o projeto AWEB - Arquivamento da *Web* das Eleições Brasileira de 2018; e o projeto Arquivamento da *Web* brasileira: políticas de preservação e modelos tecnológicos".

Além dos projetos de pesquisa, o NUAWEB captura e arquivava sítios da *web* de assuntos específicos, ou os de interesse do Núcleo. Pode-se destacar o monitoramento e captura das páginas da *web* referentes à pandemia do Coronavírus (COVID-19). Dentre os sítios da *web* relacionados à doença e preservados encontram-se a Folha de São Paulo, o Estadão, o Globo, o El País, o Uol, o The Intercept Brasil e o DW Brasil, além dos sítios oficiais do governo federal. (NUAWEB, *website*, 2023). O arquivamento se dava através de indicação de páginas da *web* relacionadas à pandemia, por meio de formulário.

O sítio do NUAWEB tem um material introdutório sobre o entendimento de arquivamento da *web*, projetos gerenciados por eles, e as iniciativas de arquivamento. Contudo, sobre a arquitetura tecnológica adotada para o arquivamento das páginas da *web* seguido pelo projeto NUAWEB, não foi encontrado nenhum material público que possa orientar no entendimento do fluxo escolhido.

Outra iniciativa que merece destaque é o projeto piloto ARQWEB (Serviço de Preservação de Páginas Web do IBICT). Em 2022, na comemoração dos 10 anos da Rede Brasileira de Serviços de Preservação Digital (Cariniana), o IBICT iniciou o projeto piloto ARQWEB, que tem como objetivo arquivar os sítios da *web* das

instituições parceiras da Rede Cariniana, e aos sítios governamentais, e não se limitando apenas a essas categorias. Até o presente momento, foram arquivados três sítios da *web* de instituições parceiras, onde o resultado do arquivamento se manteve idêntico aos *sites* ao vivo das instituições.

3.3 A iniciativa do portal ARQWEB, do IBICT

O portal ARQWEB¹ iniciou o arquivamento com o antigo sítio do IBICT², para validar o processo de arquivamento antes de se estender para os demais sítios. Segundo o resultado do estudo, na prática, a aplicação do arquivamento da *web* pode ser feita para outros sítios institucionais, usando os mesmos princípios.

O portal ARQWEB considera os seguintes aspectos estruturais no processo de arquivamento dos sítios:

- 1) **aparência (leiaute da web ou apresentação)**: esta categoria indica o esboço geral, a apresentação e a aparência da página da *web* (MELO, 2020, p. 62);
- 2) **comportamento (navegação de código)**: caracterizado por navegação a partir de *links*, podendo direcionar para o mesmo sítio da *web*, ou para outros diferentes, ou ainda para documentos externos ou recursos dinâmicos e animados, como o *feed* de uma rede social, comentários, marcações ou favoritos. (KHAN; RAHMAN, 2019, p. 75).

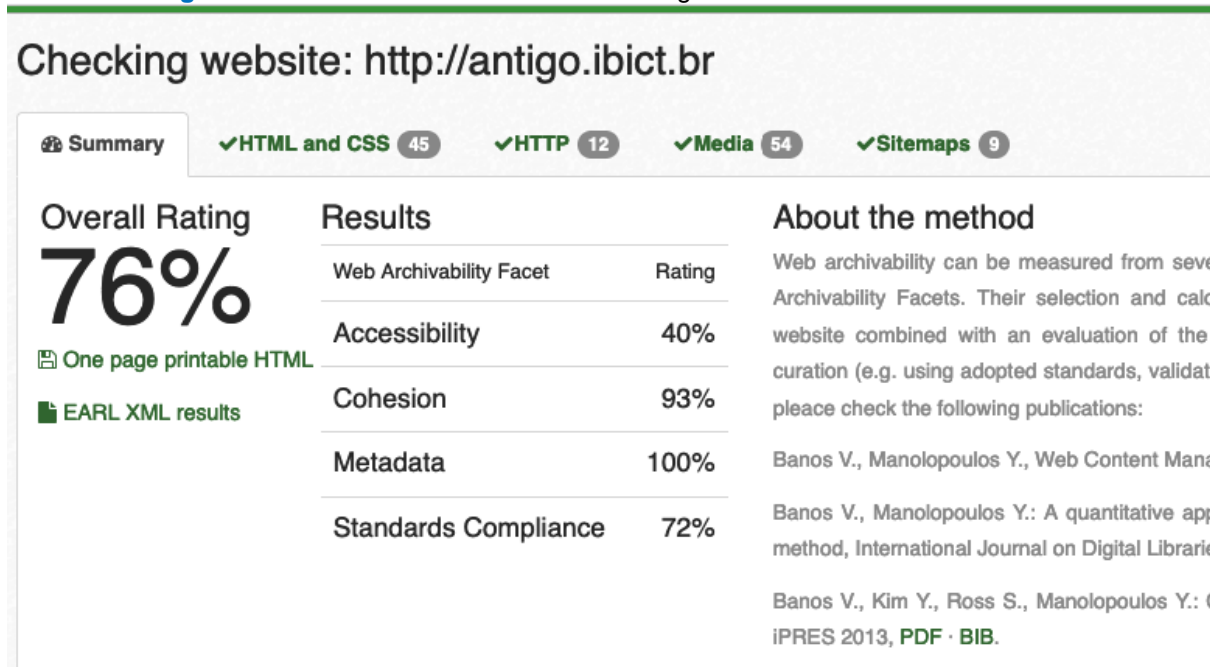
O processo de arquivamento utilizou a ferramenta de avaliação de arquivabilidade de sítios da *web* do ArchiveReady como forma de indicar se os aspectos estruturais serão mantidos, e se haverá diminuição na qualidade do arquivamento da *web*. A ferramenta faz uma análise na estrutura do sítio avaliado e no seu conteúdo, levando em consideração a acessibilidade, os padrões em conformidade, a coesão e os metadados. Ao avaliar o sítio da *web*, o ArchiveReady gera uma pontuação que demonstra a preservabilidade, e produz um relatório que permite visualizar as conformidades e as não conformidades, e as sugestões, para aumentar o potencial de ser preservável. (ARCHIVEREADY, 2023).

¹ <https://arqweb.ibict.br>

² <https://antigo.ibict.br/>

A consulta da preservabilidade de um *site* usando o ArchiveReady é pública, e pode ser aplicada por qualquer pessoa a qualquer sítio. Foi feito o teste considerando o antigo *site* do IBICT e o resultado determinou 76% de preservabilidade.

Figura 1. Resultado da análise do sítio³ antigo do IBICT feito em 22/05/2023.



Fonte: Elaborado pelos autores (2023).

Após a validação do fluxo de arquivamento adotado pelo ARQWEB, o projeto piloto iniciou a coleta de mais dois (2) sítios da web, o da UNICAMP, e o da Biblioteca Nacional. A seleção foi feita através da manifestação de interesse por parte das instituições que participam da Rede Cariniana. Uma vez selecionados, a coleta foi feita através do software open-source Heritrix, escrito em linguagem Java, e desenvolvido pela Internet Archive. O resultado do rastreamento foi armazenado em arquivos WARC (WebARChive), formato este que desde 2009 foi adotado como padrão para arquivos da web, definido na ISO 28500.

A profundidade da coleta permitiu que todos os sítios da web fossem arquivados, garantindo que as páginas e conteúdos que pertencem aos mesmos domínios fossem preservados. Páginas e objetos que estavam em outros domínios deverão ser tratados conforme a sua importância, para manter o layout da página e sua relevância.

³ <https://antigo.ibict.br>

Como resultado do arquivamento do sítio da web do IBICT foi possível manter o leiaute e o conteúdo idênticos ao sítio ao vivo. Contudo, conteúdos como vídeos, integração com agenda do Google e integrações com redes sociais não foram coletados e arquivado. Para a coleta desses itens, pertencente a um domínio externo, é preciso que faça parte do escopo de arquivamento. Assim é possível medir a profundidade da coleta e evitar que o rastreador seja direcionado para um novo ponto de entrada e faça o arquivamento inteiro do domínio externo, e não apenas do objeto do *layout*.

O sítio ARQWEB traz como interface principal a relação de *sites* arquivados e com um filtro que permite selecionar o ano da busca, coincidindo com o período do endereço pesquisado. Esta experiência é possível através do uso da ferramenta pywb, desenvolvida e mantida pela Webrecorder (2023), que possibilita a reprodução dos arquivos WARC's capturados.

Figura 2. Resultado da captura do antigo *site* em 15/02/2022.



Fonte: Elaborado pelos autores (2023).

4 CONSIDERAÇÕES FINAIS

Buscou-se reunir e analisar aspectos ligados ao arquivamento da *web*, que é a coleta de sítios para garantir que as informações sejam preservadas para pesquisas ao longo do tempo. Deve-se compreender os usos potenciais da tecnologia e do conteúdo armazenados sob a ótica da informação e da comunicação, pois eles estão envolvidos no arquivamento para a seleção, manutenção, recuperação, custódia e preservação digital, para a memória mundial.

No que tange ao que faz ou fará parte dos acervos digitais é importante tomar cuidado com os conteúdos “voláteis” ou que mudam de forma rápida, como os sítios da *web* externos à instituição, os conteúdos dinâmicos e os que advêm da mídia social, os sítios protegidos por senha e os *streamings* e mídias para *download*. Também há que se observar, para estabelecer estas estratégias e fluxos para arquivamento, se o conteúdo pretendido foi coletado integralmente, se o conteúdo intelectual pode ser reproduzido na ferramenta de acesso, e se a cópia coletada pode ser reproduzida com a mesma aparência e com o mesmo comportamento do sítio ao vivo.

O fluxo de ações para um arquivamento da *web* “ideal” deve contemplar a manutenção do material protegido por direitos autorais, e tomar cuidado com os conteúdos de páginas digitais que podem causar dificuldades administrativas ao proprietário do sítio, como os que passaram por circunstâncias que se alteraram e precisaram ser retirados do ar, ou ainda os com páginas publicadas com erro e que carecem de correção.

Uma necessidade para o arquivamento da *web* (importante para o mundo digital) é investir no desenvolvimento constante das competências individuais e da equipe, e mostrar que a cooperação, com pesquisadores e especialistas, é vital e faz diferença positiva nos projetos.

Com relação aos dados coletados como referências para a pesquisa, embora originalmente se pretendesse priorizar a pesquisa em artigos, além destes, observou-se que foram constituídos por livros, teses e dissertações, apresentações em eventos, documentos oficiais (leis), entre outros. Verificou-se que grande parte da literatura brasileira se encontra disponível em sítios de acesso livre, sendo obtido com acesso não pago, como por exemplo alguns dos disponibilizados no Portal de Periódicos da

CAPEs. Diferentemente, na literatura estrangeira, boa parte do conteúdo está indisponível gratuitamente, e para seu acesso é exigida ou a vinculação com a universidade que o disponibiliza, ou o pagamento para o acesso, e/ou envio da cópia, dificultando o livre acesso ao conteúdo.

Considera-se ter atingido o objetivo geral deste levantamento, que foi a reunião de literatura sobre arquivamento de páginas da *web* considerando suas definições, como o processo é feito, as estratégias, os fluxos de arquivamento, e as iniciativas internacionais e nacionais. Como resultado, do conteúdo tratado e dos exemplos encontrados foram abordadas sete iniciativas internacionais com variedades de contexto (bibliotecas nacionais, organizações governamentais e sem fins lucrativos e arquivos nacionais), e duas nacionais, onde pode-se destacar a mais recente iniciativa, o ARQWEB, do IBICT, que se lança como uma experiência de arquivamento de sítios de instituições parceiras da Rede Cariniana, e de sítios governamentais. A investigação apresentou um cenário incipiente de iniciativas nacionais de arquivamento da *web*, comparado às internacionais, o que traz ainda muitas possibilidades de estudo e mostra o quão importante é estabelecer políticas, ter estratégias e definir fluxos para o arquivamento *web* nas instituições, garantindo qualidade na coleta e no seu resultado.

REFERÊNCIAS

ALAM, S., *et al.* Supporting web archiving via web packaging. **arXiv**:1906.07104 [cs.NI], 2019. Disponível em: <http://arxiv.org/abs/1906.07104>. Acesso em: 7 abr. 2023

BANOS, Vangelis. **ARCHIVEREADY**, 2017. Disponível em: <https://archiveready.com/>. Acesso em: 20 mar. 2023.

BOERES, Sonia Araújo de Assis. Necessidade de capacitação de gestores para preservação digital na biblioteconomia, museologia e arquivologia. **Revista Ibero-Americana de Ciência da Informação**, Brasília, DF, v. 9 No 2, n. 2, p. 517-527, 2016. Disponível em: <https://periodicos.unb.br/index.php/RIC/article/view/2426/2164>. Acesso em: 10 abr. 2023.

BOERES, Sonia Araújo de Assis. **Competências necessárias para equipes de profissionais de preservação digital**. 2017. 293 f., il. Tese (Doutorado em Ciência da Informação) — Universidade de Brasília, Brasília, 2017. Disponível em: <https://repositorio.unb.br/handle/10482/24354?mode=full>. Acesso em: 22 fev. 2023.

COSTA, Miguel; GOMES, Daniel; SILVA, Mário J. The evolution of web archiving. **International Journal on Digital Libraries**, Heidelberg, p. 1-15, 2016. DOI: <https://doi.org/10.1007/s00799-016-0171-9>. Acesso em: 8 abr. 2023.

GUIDELINES to make archivable websites. **Publications Office of The European Union**. [S.l.], 2021. Disponível em: <http://lnnk.in/hFgV>. Acesso em: 8 abr. 2023.

GOMES, D. Preservar a Web: um desafio ao alcance de todos. In: CONGRESSO NACIONAL DE BIBLIOTECÁRIOS, ARQUIVISTAS E DOCUMENTALISTAS, 10., 2010, Guimarães, Portugal, Políticas de Informação na Sociedade em Rede. **Actas**. Guimarães: BAD, 2010. Disponível em: <https://lnnk.in/hqhF>. Acesso em: 8 abr. 2023.

IBICT. **ARQWEB**, 2022. Disponível em: <http://argweb.ibict.br/>. Acesso em: 20 abr. 2023.

KHAN, Muzammil; RAHMAN, Arif Ur. A systematic approach towards web preservation. **Informational Technology and libraries**, Chicago, v.38, n.1, p. 71 – 90. 2019. Disponível em: <https://ejournals.bc.edu/index.php/ital/article/view/10181>. Acesso em 3 abr. 2023.

LUZ, Ana Javes. Preservação de sites oficiais: exemplos internacionais e o caso brasileiro. **Revista Brasileira de Preservação Digital**, Campinas, SP, v.3, 2022. DOI: [10.20396/rebpred.v3i00.16587](https://doi.org/10.20396/rebpred.v3i00.16587). Acesso em: 18 maio 2023.

MELO, J. F. **Arquivamento dos websites do governo federal brasileiro: preservação do domínio gov.br**. 2020. 133 f. Dissertação (Mestrado – Programa de Pós-Graduação em Comunicação e Informação) – Universidade Federal do Rio Grande do Sul, Porto Alegre, 2020. Disponível em: <http://hdl.handle.net/10183/210671>. Acesso em: 04 abr. 2023.

MELO, J. F.; ROCKEMBACH, M. Arquivabilidade de websites para preservação digital: estudo a partir da área da saúde. **RECIIS: Revista Eletrônica de Comunicação, Informação e Inovação em Saúde**, Rio de Janeiro; v. 14, n. 3, p. 529-545, jul./set. 2020. Disponível em: <https://www.arca.fiocruz.br/handle/icict/43726>. Acesso em: 4 abr. 2023.

UFRGS. **NUAWEB – Núcleo de Pesquisa em Arquivamento da Web e Preservação Digital**, 2019. Disponível em: <https://www.ufrgs.br/nuaweb/>. Acesso em: 20 abr. 2023.

ROCKEMBACH, M. Arquivamento da Web: estudos de caso internacionais e o caso brasileiro. **RDBCI: Revista Digital de Biblioteconomia e Ciência da Informação**, Campinas, SP, v. 16, n. 1, p. 07-24, 2018. Disponível em: <http://eprints.rclis.org/32269/1/8648747-30995-7-PB.pdf>. Acesso em: 4 abr. 2023.

ROCKEMBACH, M. A web brasileira na Covid-19: arquivamento da web e preservação digital. **Liinc Em Revista**, Rio de Janeiro, v. 17, n. 1, p. 5713, 2021. Disponível em: <http://revista.ibict.br/liinc/article/view/5713>. Acesso em: 4 abr. 2023

ROSA, C. A.; CRAVEIRO, O.; DOMINGUES, P. Open-source software for digital preservation repositories: a survey. **International Journal of Computer Science & Engineering Survey**, Tamil Naduv, v.8, n. 3, 2017. Disponível em: <https://tinyurl.com/5fbch72n>. Acesso em: 4 abril. 2023

SCHAFER, V.; WINTERS, J. The values of web archives. **International Journal of Digital Humanities**, Cham, SW, v. 2, p. 129–144, 2021. DOI: <https://doi.org/10.1007/s42803-021-00037-0>. Acesso em 4 abr. 2023.

SILVA, K. M. da. **Um modelo de ciclo de vida de dados na web**. 2019. 133 f. Dissertação (Mestrado – Programa de Pós-graduação em Ciência da Computação) – Universidade Federal de Pernambuco, Recife, 2019. Disponível em: <https://repositorio.ufpe.br/handle/123456789/34147> . Acesso em: 10 abr. 2023.

SILVA, Armando Malheiro *et al.* **Arquivística**: teoria e prática de uma ciência da informação. Porto: Edições Afrontamento, 1999.

WEBRECORDER. **Web archiving for all!** 2023. Disponível em: <https://webrecorder.net> . Acesso em 20 maio 2023.