



## A web brasileira na Covid-19: arquivamento da web e preservação digital

*The Brazilian web at Covid-19: web archiving and digital preservation*

Moisés Rockembach <sup>a,\*</sup> 

**RESUMO:** O artigo trata do arquivamento de páginas web referentes à COVID 19 no contexto brasileiro. Discorre sobre questões gerais da criação da web e seus objetivos iniciais, bem como do surgimento da necessidade de preservação da web. Foram analisadas iniciativas internacionais de preservação de conteúdos relativos à Covid 19, em destaque o IIPC (Consórcio Internacional de Preservação da Internet). O método de pesquisa utilizado foi o estudo de caso e de métodos mistos, qualitativo e quantitativo, revisão de literatura, pesquisas bibliográfica e documental. A pesquisa realizada coletou cerca de 16,4Gb de arquivos, até março de 2021 e pretendeu levantar a importância do arquivamento da web e da preservação de conteúdos sobre COVID 19 para acesso futuro.

**Palavras-chave:** Arquivos da Web; Preservação Digital; Arquivamento da Web; Covid-19; Estudo de Caso.

**ABSTRACT:** The article deals with the web archiving referring to COVID 19 in the Brazilian context. It discusses general questions about the creation of the web and its initial objectives, as well as the emergence of the need to preserve the web. International content preservation initiatives related to Covid 19 were analyzed, in particular the IIPC (International Consortium for the Preservation of the Internet). The research method used was the case study and mixed methods, qualitative and quantitative, literature review, bibliographic and documentary research. The research carried out collected approximately 16.4 GB of files, until March 2021 and intended to raise the importance of archiving the web and preserving content on COVID 19 for future access.

**Keywords:** Web Archive; Digital Preservation; Web Archiving; Covid-19; Case Study.

---

<sup>a</sup> Departamento de Ciências da Informação, Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, Brasil.

\* Correspondência para/Correspondence to: Moisés Rockembach. Endereço: FABICO/UFRGS - R. Ramiro Barcelos, 2705 - Santana, Porto Alegre - RS, 90035-007, Brasil. E-mail: moises.rockembach@ufrgs.br.

Recebido em/Received: 01/04/2021; Aprovado em/Approved: 16/06/2021.

Artigo publicado em acesso aberto sob licença [CC BY 4.0 Internacional](https://creativecommons.org/licenses/by/4.0/) 

## INTRODUÇÃO

Os desafios contemporâneos sobre a explosão da produção informacional, com dados conectados e em rede, suscitam questionamentos quanto à formação da memória digital e a garantia de seu acesso futuro.

As tecnologias de informação visam a imediata propagação de informações, mas nem sempre se comprometem com a sua recuperação em termos de preservação digital. As pesquisas envolvendo a preservação de conteúdos online são recentes e vem demandando uma atualização tão constante como das inovações tecnológicas. Na preservação e arquivamento de conteúdos web, por exemplo, após um ano de publicação na web, cerca de 80% dos hiperlinks são substituídos por novos ou deixam de apontar para o recurso informacional, invalidando o acesso (Ntoulas, Cho, Olston, 2004). Além disso, o grande volume de informações produzidas pela Internet abre espaço para a proliferação de desinformação, sem haver critérios de seleção ou avaliação eficientes. Alguns fatores decorrentes do boom informacional de 2007, como destaca Friedman (2016), possibilitaram a aceleração da conexão em rede, como tecnologias móveis 3G, bases de dados voltadas para a clusterização de dados (big data), desenvolvimento da inteligência artificial (IA) e outras tantas tecnologias que visam a disseminação ampla da informação. Tais tecnologias também possibilitaram práticas de desinformações, intencionais ou não, e do fenômeno da Infodemia, comparável e potencializado pela Pandemia, com o devido alerta da Organização Mundial da Saúde (WORLD HEALTH ORGANIZATION, 2020) quanto aos malefícios destas práticas.

No intuito de propor um estudo de caso de arquivamento da web no contexto brasileiro, optamos por escolher as informações sobre Covid-19 e a Pandemia no Brasil, tendo em consideração múltiplas fontes de informação, pois, como abordado em diversas iniciativas realizadas sobre arquivos da web da Covid-19 (International Internet Preservation Consortium, 2020), é possível perceber a relevância da preservação deste tipo de informação pública, que servirá como fonte de pesquisa e forma de documentar este período da história brasileira. Procurou-se enquadrar metodologicamente a pesquisa a partir de uma metodologia bibliográfica, documental e de estudo de caso, com aspectos quantitativos e qualitativos. Para obtermos uma melhor compreensão da pesquisa, abordamos os procedimentos adotados na investigação, o enquadramento das origens da web e de conceitos e abordagens sobre arquivamento da web, o cenário internacional, para então demonstrar os resultados do estudo de caso.

## A COVID-19 E A PRESERVAÇÃO DIGITAL

A Organização Mundial de Saúde (World Health Organization, 2020) utiliza o termo infodemia (infodemic) como uma forma de salientar que uma abundância de informações pode nos levar a difundir dados equivocados e que isto pode trazer consequências desastrosas à sociedade, prejudicando os indivíduos fisicamente e

mentalmente. A confiança nas informações é uma qualidade essencial, principalmente com as possibilidades de disseminação que a web, os aplicativos e as redes sociais nos permitem realizar.

O mundo todo passa por uma guerra de informações, a guerra de narrativas em relação a Covid-19, em muitos assuntos em que normalmente deveriam pesar os dados científicos, transforma o modo como a Sociedade enfrenta seus desafios. A preservação dos conteúdos da web, um processo que é realizado em diversas partes do mundo e no Brasil começa a se desenvolver de maneira mais consolidada, auxiliará a manter acessível recursos importantes para que os futuros pesquisadores e usuários interessados no tema possam analisar os fatos e chegar às suas próprias conclusões.

Diversos tipos de organizações em todo o mundo realizam a preservação de conteúdos da web, com políticas que vão desde a captura de domínios nacionais a sites governamentais e registros de interações entre o Estado e os seus cidadãos (Rockembach, Pavão, 2018).

Ao longo dos anos, algumas políticas de preservação digital no contexto brasileiro vem abordando diversos tipos de documentos nato digitais, a produção cada vez mais volumosa destes tipos de documentos impulsionam a tomada de medidas que visem seu acesso ao longo do tempo, como por exemplo a Carta para a Preservação do Patrimônio Arquivístico Digital (CONARQ, 2004), a Política da Rede Brasileira de Serviços de Preservação Digital (CARINIANA, 2014), e a Política de Preservação Digital (Arquivo Nacional do Brasil, 2016). Exemplos recentes vêm incluindo a preservação dos websites no seu escopo, como são os casos da Política de Preservação Digital da Biblioteca Nacional - PPDBN (Fundação Biblioteca Nacional, 2020) e da Política de Preservação de Acervos Digitais da Universidade Federal do Rio Grande do Sul (Universidade Federal do Rio Grande do Sul, 2021).

### **O processo de arquivamento da web**

Esta é uma situação que acontece a todos, não somente com os profissionais da informação, mas com todas as pessoas que precisam buscar algum recurso da web e que, por algum motivo ou outro, não conseguem recuperar o conteúdo. Este problema na recuperação da informação pode se dar por diversas questões, o mais comum é o chamado erro 404 (*error 404 not found*), ou página não encontrada.

Na produção de informações organizacionais, é comum que muitas delas sejam geradas e comunicadas unicamente na web, sejam organizações públicas, privadas ou do terceiro setor. Isto fica evidente nas relações que exijam a publicização e transparência, nas relações comerciais ou mesmo nas interações com os cidadãos (governo/terceiro setor) e clientes (setor privado), potencializado com plataformas da web 2.0. Todo este conteúdo é criado e permanece na web por tempo indeterminado, o que podem ser muitos anos ou apenas por um breve momento, vide o fenômeno de proliferação das ferramentas de stories nas redes sociais.

Surge também a necessidade de procurar uma referência para algum trabalho acadêmico e após algum tempo, também haver problemas em recuperar esta referência. Este tempo de acesso pode ser muito curto, entre o momento em que o recurso está acessível e quando ele deixa de estar disponível. O uso de links persistentes pode auxiliar na manutenção do vínculo entre o recurso da web e o identificador persistente que, diferente da URL (*Uniform Resource Locator*) a qual indica a localização do recurso no momento que ele é disponibilizado, representará o recurso de maneira inequívoca e com controle de sintaxe, identificando o produtor e o objeto de forma única.

Entretanto, com o arquivamento da web pode-se preservar toda a diversidade de conteúdos e formatos da maneira como foram disponibilizados originalmente, representando uma grande memória digital do nosso tempo, nas diversas iniciativas de preservação digital existentes pelo mundo.

No momento que a web foi criada, ela não foi pensada em termos de preservação digital, assim como grande parte das tecnologias. Isto tem relação com a obsolescência programada, mas é muito relacionado também com a dinamicidade própria dos meios digitais e com as atualizações constantes das tecnologias envolvidas, sejam de software ou de hardware.

Uma das questões em relação à tecnologia para os profissionais da informação é da teoria *black box* ou caixa preta, advinda da Teoria dos Sistemas, onde não compreendemos muito bem o funcionamento dos meios tecnológicos e trabalhamos somente nos inputs e outputs (entradas e saídas) dos sistemas de informação. Isto pode gerar diversos problemas, alguns muito presentes no nosso dia a dia, como as discussões sobre os algoritmos e como eles tem influenciado os sistemas de filtro e recomendações, como os filtros-bolha (PARISER, 2011) das redes sociais, sendo este fenômeno observado também em casos de falta de transparência algorítmica.

O conceito de *world wide web* remonta a proposição de Tim Berners-Lee, em 1989, e desde então, diversos elementos foram sendo incorporados para criar este sistema informacional e comunicacional em que é constituída a web, como o uso de um navegador (browser), o uso inicial de uma linguagem de marcação (HTML), onde posteriormente foram incorporadas diversas outras linguagens de programação e funcionalidades, a arquitetura cliente-servidor e a navegação por meio de hiperlinks. É neste uso de hiperlinks onde o recurso da web pode se perder, caso o recurso requisitado pelo usuário não esteja mais disponível no servidor, este retornará com a mensagem de erro.

A concepção da web por Tim Berners-Lee tem seu início em 1989, conforme referido, mas é somente em 1991 que este conceito se torna algo concreto, com a primeira página da web do mundo, ainda hoje disponível online, por meio de um projeto que recuperou o conteúdo original (<http://info.cern.ch/hypertext/WWW/TheProject.html>) e ainda proporcionou um emulador para a visualização do site da mesma forma como era visto no início dos anos 1990 (<https://worldwideweb.cern.ch/browser/>).

A primeira iniciativa de preservação da web do mundo surgiu em 1996, com o Internet Archive, isto é, cinco anos após o surgimento da primeira página da web e logo após o início da expansão do uso comercial da web. Portanto, isto representa a necessidade, que aparece logo após a criação da web, em preservar e manter estes conteúdos digitais acessíveis.

A preservação da web nasce desta necessidade em manter conteúdos disponibilizados online, com diversos formatos e tipos de informação, documentos textuais, sonoros, multimídias e essa complexidade pode ter atrasado a adoção de métodos de preservação digital por parte das instituições e de pesquisadores. O arquivamento da web exige uma atualização constante dos métodos e tecnologias (Masanes, 2006) que fazem parte do processo de preservação digital, consoante às próprias mudanças tecnológicas que surgem no desenvolvimento web.

Muitas questões têm despontado relativas à preservação de conteúdos web por usuários comuns, não especializados, pela própria necessidade em manter estes registros disponíveis e não ficar à mercê do acesso por parte do produtor ou servidor onde originalmente se encontra o recurso. Nas relações comerciais, por exemplo, quando uma empresa anuncia um produto ou serviço com determinadas características e valores, e depois modifica, se o cliente não conseguir recuperar a página ou postagem original, não é possível comprovar como a informação foi apresentada, pois não há um versionamento destes conteúdos, só há a web ao vivo. No momento em que é possível navegar por versões ao longo do tempo e perceber as modificações, o que é um dos objetivos dos arquivos da web, abrem-se diversas possibilidades de recuperação desta memória digital.

Alguns termos são importantes neste contexto do arquivamento da web. Relacionamos abaixo algumas definições utilizadas neste trabalho e que nos auxiliam no processo de preservação digital:

**Arquivabilidade do website:** condições que um website possui para a transferência segura de seu conteúdo para um arquivo da web, com fins de preservação digital, levando em consideração a acessibilidade, coesão, metadados e padrões em conformidade.

**Captura:** etapa do método de preservação dos websites, correspondente a recolha das páginas web baseada em uma lista de entradas ou URLs pré-definidas.

**Controle de qualidade:** todos os conteúdos da web capturados podem ter maior ou menor qualidade, conforme vários quesitos, como a completude dos dados, incluindo textos, imagens e reprodutores de multimídia (som e vídeo) ou na reprodução posterior do conteúdo. A fidelidade em reproduzir o website com o mesmo design original também é importante, porém a captura de todo o conteúdo das URLs é primordial. Por isto, torna-se necessário realizar uma etapa de controle da qualidade, verificando como o rastreador da web foi executado e a reprodutibilidade do conteúdo armazenado.

**Quebra de link:** um dos principais fenômenos que levam a necessidade de preservação dos conteúdos da web. Com um curto ciclo de vida dos links, perde-se rapidamente a conexão ao recurso informacional em um servidor.

**Rastreador da web:** o rastreador explora e navega na web de forma automatizada, buscando recursos pré-configurados e capturando-os para o armazenamento.

**Semente:** a URL que funciona como ponto de entrada para o rastreador da web.

**URL:** Acrônimo de *Uniform Resource Locator* ou Localizador Uniforme de Recursos, que representa a localização do recurso na web.

**WARC:** formato de arquivo em modelo *open source*, baseado na ISO 28500, para preservação de arquivos da web a longo prazo.

### **A preservação da web sobre a Covid-19**

O arquivamento da web é essencialmente um trabalho coletivo e colaborativo. A tarefa de tentar preservar toda a web mundial pode ser considerada utópica, mas com o trabalho conjunto de diversos organismos, instituições e indivíduos, é possível preservar aspectos importantes dos dias atuais, produzidos e publicados em um dos meios informacionais mais utilizados pela Sociedade, a web, e possibilitar acessos aos pesquisadores que não terão alternativas de uso, se não forem preservados hoje.

Algumas iniciativas internacionais na preservação de conteúdos da Covid-19 se destacam, uma delas é a do grupo de desenvolvimento de conteúdo do Consórcio Internacional de Preservação da Internet (International Internet Preservation Consortium, 2020), o projeto é representado pelos dados abaixo:

De forma geral, a cobertura dos arquivos da web acontece de forma desigual nos domínios (Rockembach, 2017), portanto faz-se necessário o trabalho colaborativo de diversas iniciativas ao redor do mundo para atingir uma maior preservação digital.

Uma iniciativa brasileira de arquivamento da web no âmbito da pesquisa científica vem desenvolvendo um trabalho no contexto da preservação digital, de forma que diversos conteúdos do domínio .br e da internet brasileira, de maneira geral, possam ser identificados e capturados para acesso retrospectivo, o Arquivo da Web Brasileira ([www.arquivo.org.br](http://www.arquivo.org.br)), com a colaboração do Núcleo de Pesquisa em Arquivamento da Web e Preservação Digital ([www.ufrgs.br/nuaweb](http://www.ufrgs.br/nuaweb)) e voluntários. Os escopos atingidos são os mais diversos, perpassando especialmente pelas áreas institucional, organizacional, esfera federal governamental e das eleições. O foco deste trabalho foi demonstrar o que vem sendo produzido na web em termos do tópico da Covid-19.

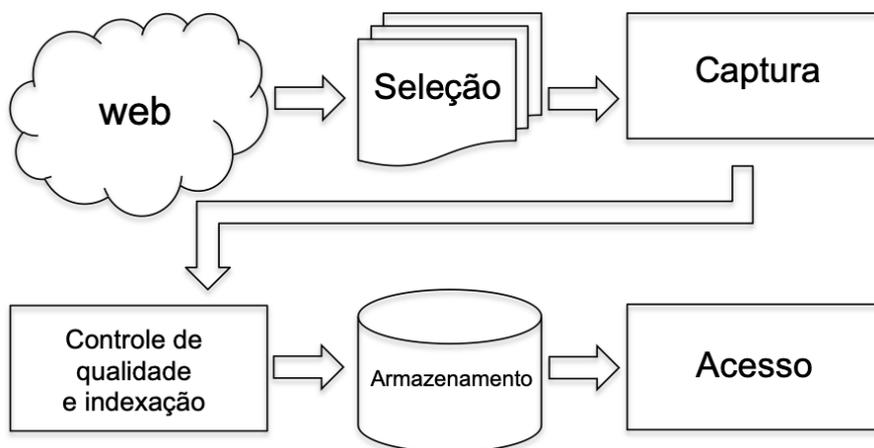
**Tabela 1.** Representação dos seeds capturados por idioma em Março de 2021

Espanhol	3666	Grego	18	Asl	1
Inglês	3162	Turco	15	Búlgaro	1
Português	1944	Sueco	14	Birmanês	1
Francês	775	Polonês	12	Cebuano	1
Alemão	707	Islandês	10	Chuukese	1
Holandês	352	Tailandês	9	Farsi	1
Sérvio	323	Malaio	8	Inuktitut	1
Dinamarquês	275	Vietnamita	8	Carélio	1
Norueguês	203	Urdu	6	Letão	1
Finlandês	175	Hebraico	5	Marshallês	1
Tcheco	167	Luxemburguês	4	Maori	1
Eslovaco	127	Papiamento	4	Nepalês	1
Estoniano	100	Ucraniano	4	Reto-românico	1
Chinês	84	Bengali	3	Romeno	1
Italiano	78	Catalão	3	Romanche	1
Húngaro	67	Dialeto Criolo/Francês	3	Sami	1
Japonês	54	Filipino	3	Cingalês	1
Árabe	44	Feroês	2	Somali	1
Croata	41	Hindi	2	Suaíli	1
Russo	35	Tagalo	2	Tâmil	1
Coreano	34	Albanês	1	Tigrinya	1
Indonésio	21	Amárico	1		
Persa	20	Armênio	1	<b>Total de seeds</b>	<b>12613</b>

Fonte: (INTERNATIONAL INTERNET PRESERVATION CONSORTIUM, 2020).

O processo de arquivamento da web executado segue o que é proposto na figura 1:

**Figura 1.** Modelo de Arquivamento da web.



O arquivamento da web compreende coletar, armazenar e disponibilizar a informação retrospectiva da web para futuros pesquisadores (Rockembach, 2018), mas também

preservar a informação da web como memória digital institucional, organizacional, coletiva ou individual (Rodrigues, Rockembach, 2021).

A pesquisa teve início em abril de 2020, na curadoria digital necessária para a seleção das URLs e sementes foi levado em consideração, em um primeiro momento, a repercussão na mídia da Covid-19 e os dados oficiais governamentais. Estes dois focos principais no início da pesquisa serviram de embasamento para a observação e monitoramento do que foi sendo publicado na web. Um dos fatores que nos levaram a rastrear estes conteúdos foram também as diferenças apontadas entre os dados oficiais e do Consórcio dos Veículos de Imprensa, que tinham números e critérios distintos nas suas apresentações. A comunicação (Luz, Weber, 2019) e a informação (Melo, Nunes, Rockembach, 2019) governamental precisam ser particularmente preservadas, garantindo o exercício da cidadania, transparência pública e acesso à informação.

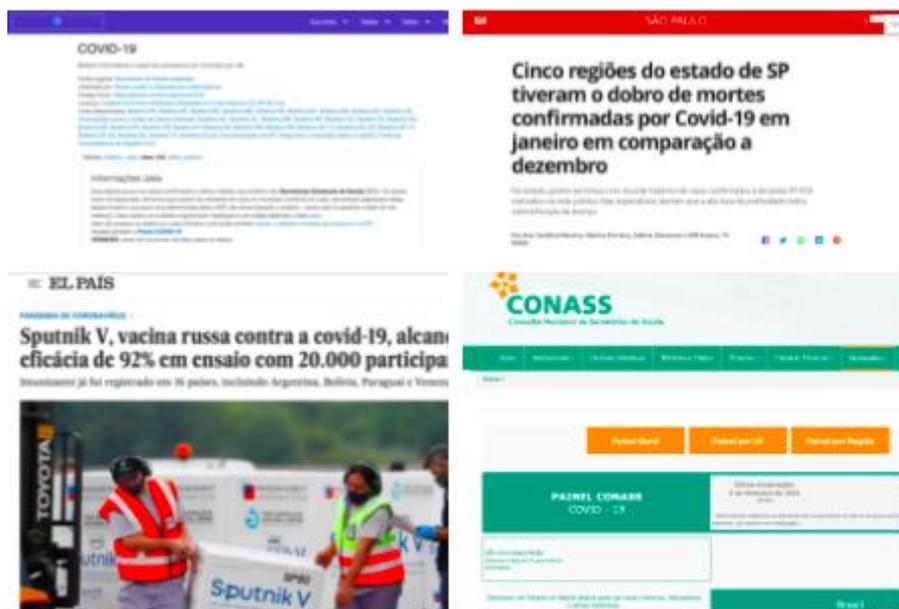
Entretanto, a curadoria digital da pesquisa não foi o único fator de seleção de conteúdos. Juntamente com este critério, foi adotado no decorrer da pesquisa, a partir do mês de agosto de 2020, uma seleção baseada em *crowdsourcing* (<https://www.ufrgs.br/nuaweb/covid19/>), chamando pesquisadores e usuários em geral para colaborar com a nomeação de sites que pudessem ser preservados. Esta ferramenta de *crowdsourcing* foi elaborada em formato de formulário, com os campos do endereço (URL) do site a ser preservado e a descrição resumida do site, procurando detalhar quando possível palavras-chave. Procurou-se tornar o processo de solicitação de recomendação de sites a serem preservados o mais simples possível para contar com as colaborações, mantendo as informações essenciais.

Na soma dos processos de curadoria digital e de *crowdsourcing*, obteve-se as seguintes nomeações: Folha de São Paulo, Estadão, Globo, El País, Uol, The Intercept Brasil e DW Brasil. Destes foram acrescentados também diversos sites oficiais do governo federal relativos a Covid-19.

O rastreamento e captura dos websites selecionados foram realizadas tendo em conta um cruzamento de métodos computacionais, levando em consideração o conhecimento em linguagens de programação e a disponibilidade de ferramentas Open Source. Desta forma foi possível realizar a captura de diversos conteúdos no que concerne a Covid-19, além de realizar o controle de qualidade das páginas capturadas.

Alguns dos exemplos dos sites preservados podem ser observados na figura 2:

Figura 2. Exemplos de websites capturados na pesquisa



Dentre estes exemplos acima, temos sementes relacionadas ao [brasil.io](#), [G1](#), [El País](#) e [Conass](#) (Conselho Nacional de Secretários de Saúde).

O armazenamento dos websites dessas URLs foi realizado no formato WARC (WebARChive), no padrão ISO 28500 (International Organization for Standardization, 2009), conforme a normatização adotada internacionalmente, e contou com 16,4 GB de armazenamento em março de 2021. A captura dos sites continua sendo realizada, portanto os números aqui apresentados representam o atual estado do acervo digital. O armazenamento final dos arquivos da web acontece nos servidores da instituição responsável pela pesquisa, com os devidos protocolos de backup e segurança da informação. Por sua vez, por uma definição baseada em boas práticas internacionais, a disponibilização dos arquivos da web para acesso público deve aguardar um período de embargo, geralmente a partir de um ano da captura e preservação dos websites, para que estes conteúdos não concorram com os websites ao vivo e para que os arquivos da web cumpram a sua função de memória digital e pesquisa retrospectiva (Arquivo da Web Portuguesa, 2017; British Library, 2014).

O modelo de arquivamento da web ilustrado identifica diversos pontos deste processo de preservação, partindo da seleção das sementes que servirão para as configurações de rastreadores e para a captura, o controle de qualidade e indexação dos conteúdos capturados, o armazenamento em padrão ISO 28500 e o acesso por meio da reprodução dos arquivos preservados. A observação de inconsistências na qualidade dos conteúdos capturados podem ainda sinalizar uma diminuição na arquivabilidade dos websites (MELO, ROCKEMBACH, 2020). A medição da arquivabilidade dos websites, a partir da acessibilidade, padrões em conformidade, coesão e metadados, advém da proposta de Banos e Manolopoulos (2015), com a possibilidade do uso da ferramenta Archive Ready ([www.archiveready.com](http://www.archiveready.com)).

O fenômeno da infodemia, o excesso informacional e a dificuldade em separar uma informação confiável, de outra com menos confiabilidade, vem afetando diversos setores da Sociedade. A propagação de *fake news*, que se referem as notícias falsas "no intuito de enganar os seus leitores e usuários, propagando conteúdos atraentes, mas enganosos, a fim de obter ganhos financeiros e políticos" (Neves, Borges, 2020) tem sido recorrente nos dias atuais. Por outro lado, informações na web são modificadas rapidamente e somente é possível estudar e comparar estas modificações, ou ainda acessar informações que são retiradas do ar pelo produtor da informação se houver a preservação destes conteúdos em outros locais, como os arquivos da web. O primeiro caso relacionado à Lei Geral de Proteção de Dados Brasileira (LGPD) é um exemplo bastante esclarecedor neste sentido, pois tratava-se de um site que comercializava milhares de dados de cidadãos brasileiros, com uma ação ajuizada em setembro de 2020. O Ministério Público do Distrito Federal acredita que 500 mil pessoas tiveram seus dados expostos apenas na cidade de São Paulo, com vítimas em todos os estados do Brasil, mas infelizmente, o site original não estava mais disponível online, tornando o julgamento arquivado (Rockembach, Silva, 2021). A construção de arquivos da web relacionados à Covid-19 possibilitará que, mesmo que haja a quebra dos links originais e os websites não estejam mais disponíveis, a informação possa ser recuperada e utilizada para estudos futuros.

## CONSIDERAÇÕES FINAIS

O contexto do estudo aqui relatado enquadrou-se no que chamamos de arquivamento da web, compreendido como um processo que visa o armazenamento e disponibilização retrospectiva e versionada das informações da web ao vivo, com a potencialidade de estabelecer uma memória digital de nosso tempo.

Para a realização da pesquisa, delimitamos o escopo da Covid-19, uma fonte de informação que tem possibilidades de ser muito acessada no futuro, por pesquisadores e interessados nos conteúdos produzidos neste período. A web, como ambiente informacional e comunicacional dinâmico, precisa ser monitorada, em termos de preservação digital, sob o risco de perdermos rapidamente os dados disseminados nos dias de hoje e passarmos para um momento de "idade das trevas digital", utilizando as palavras de Vint Cerf, um dos idealizadores da internet, ao tratar da necessidade de preservação digital.

Identificamos as iniciativas de arquivamento da web que vem trabalhando especificamente na captura, armazenamento e preservação de conteúdos relativos a Covid-19 em diversas línguas e países. Este movimento de preservação da web vem crescendo em qualidade, quantidade e ampliando a sua abrangência internacional. No Brasil, algumas iniciativas institucionais e de pesquisa vem fazendo a diferença na preservação digital, especificamente na preservação da web alguns esforços vêm chamando a atenção e formando comunidades de interessados, na memória dos nossos tempos e da urgência em pensarmos na manutenção desses registros hoje,

uma preocupação não somente na Ciência da Informação, mas também multi e interdisciplinar.

Perceber o problema aqui gerado na incapacidade de recuperar os recursos online é o que torna necessário o arquivamento da web e a preservação digital. Pensar na preservação do conteúdo do domínio .br e de todos os recursos da web brasileira é uma forma de constituir uma memória digital e possibilitar aos pesquisadores uma análise retrospectiva, neste estudo de caso, relacionado à pandemia. A infodemia por sua vez é um fenômeno contemporâneo, que a partir do excesso de informações e a propagação de fontes não verificáveis, acaba gerando diversos prejuízos à Sociedade. O arquivamento da web procura preservar a memória digital por meio de versões passadas dos conteúdos disponibilizados na internet e por meio destes acervos é possível identificar alterações ou acessar dados que se tornam indisponíveis com o tempo, podendo, portanto, servir de fonte de pesquisa para diversos campos do conhecimento, inclusive analisando o fenômeno da infodemia dos últimos anos.

A Covid-19 e o estado de Pandemia nos levaram, enquanto pesquisadores, a repensar nossos objetos e em como podíamos contribuir de forma efetiva, dentro dos nossos respectivos campos de conhecimento. É importante que a comunidade possa conhecer as pesquisas que vêm sendo executadas no âmbito da preservação dos conteúdos web. Espera-se o desenvolvimento de trabalhos futuros que levem ao engajamento de organizações, pesquisadores e usuários em geral no campo do arquivamento da web e da preservação digital, na complementação dos atuais acervos digitais e na produção de novos acervos.

## REFERÊNCIAS

ARQUIVO DA WEB PORTUGUESA. Termos e condições. 2017 Disponível em: <<http://sobre.arquivo.pt/pt/acerca/termos-e-condicoes>> Acesso em: 5 mar. 2021

ARQUIVO NACIONAL DO BRASIL. Política de Preservação Digital. Disponível em: [http://www.siga.arquivonacional.gov.br/images/an\\_digital/and\\_politica\\_preservacao\\_digital\\_v2.pdf](http://www.siga.arquivonacional.gov.br/images/an_digital/and_politica_preservacao_digital_v2.pdf). Acesso em 10 mar. 2021

BANOS, Vangelis, MANOLOPOULOS, Yannis. A quantitative approach to evaluate website archivability using the CLEAR+method. *International Journal on Digital Libraries*. 2015, v.17(2):119-41. Disponível em: <https://link.springer.com/article/10.1007/s00799-015-0144-4>. Acesso em: 10 mar. 2021

BRITISH LIBRARY. The British Library Collection Development Policy for websites. 2014.

CARINIANA. Política da Rede Brasileira de Serviços de Preservação Digital. Disponível em: <<https://cariniana.ibict.br/index.php/central-de-conteudos/publicacoes/item/27-politica-da-rede-brasileira-de-servicos-de-preservacao-digital>>. Acesso em 10 mar. 2021

CONARQ (Conselho Nacional de Arquivos). Carta para a Preservação do Patrimônio Arquivístico Digital. Disponível em:

<[http://conarq.gov.br/images/publicacoes\\_textos/Carta\\_preservacao.pdf](http://conarq.gov.br/images/publicacoes_textos/Carta_preservacao.pdf)>. Acesso em 10 mar. 2021

FRIEDMAN, Tomas L. What the Hell Happened in 2007. In: Thank You for Being Late: An Optimist's Guide to Thriving in the Age of Accelerations. New York: Penguin Random House, 2016. p. 19-35.

FUNDAÇÃO BIBLIOTECA NACIONAL. Política de Preservação Digital da Biblioteca Nacional (PPDBN). Disponível em: <<https://www.gov.br/conarq/pt-br/assuntos/noticias/politica-de-preservacao-digital-da-biblioteca-nacional-ppdbn>>. Acesso em 10 mar. 2021

INTERNATIONAL INTERNET PRESERVATION CONSORTIUM. Novel Coronavirus (Covid-19) Collection. Disponível em < <https://archive-it.org/collections/13529>>. Acesso em 10 mar. 2021

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION. ISO 28500:2009. Information and documentation - WARC file format, 2009

MASANÉS, Julien. Web Archiving: Issues and Methods. In: Web Archiving. Springer, Berlin, Heidelberg. 2006

LUZ, Ana Javes Andrade, WEBER, Maria Helena. A memória política do Brasil no site da presidência: acesso e desvios da Comunicação dos governos de Dilma Rousseff e Michel Temer. Liinc em Revista v.15, n.1 2019.

MELO, Jonas Ferrigolo, NUNES, Lúcia Andréia Nunes de Oliveira, ROCKEMBACH, Moisés. Preservação de websites governamentais a partir do arquivamento da web: abordagens e metodologias. Encontro Nacional de Pesquisa em Ciência da Informação. Anais [recurso eletrônico]. Florianópolis: ANCIB, 2019

NEVES, Barbara Coelho, BORGES, Jussara. Por que as fake news têm espaço nas mídias sociais?: uma discussão à luz do comportamento infocomunicacional e do marketing digital. Informação & Sociedade: estudos. João Pessoa. Vol. 30, n. 2, 2020.

NTOULAS, Alexandros, CHO, Junghoo, OLSTON, Christopher. What's new on the web?: the evolution of the web from a search engine perspective. Proceedings of the 13th international conference on World Wide Web. ACM, 2004.

PARISER, Eli. O filtro invisível: o que a internet está escondendo de você. Editora Schwarcz-Companhia das Letras, 2012.

ROCKEMBACH, Moisés. Inequalities in digital memory: ethical and geographical aspects of web archiving. The International Review of Information Ethics v.26, 2017

ROCKEMBACH, Moisés. Arquivamento da Web: estudos de caso internacionais e o caso brasileiro. RDBCI: Revista Digital de Biblioteconomia e Ciência da Informação, v.16(1), p. 7-24, 2018

ROCKEMBACH, Moisés, PAVÃO, Caterina Marta Groposo. Políticas e tecnologias de preservação digital no arquivamento da web. Revista Ibero-americana de Ciência da Informação. Brasília: UnB. Vol. 11, n. 1, p. 168-182 2018.

ROCKEMBACH, Moisés, SILVA, Armando Malheiro da. Web Data and the Relationship Between the General Data Protection Regulation in Europe and Brazil. Digital Transformation and Challenges to Data Security and Privacy. IGI Global, 2021

RODRIGUES, Vander Luis Duarte. ROCKEMBACH, Moisés. Arquivos da web como fonte historiográfica. RDBCI: Revista Digital de Biblioteconomia e Ciência da Informação, v.19, 2021

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL. Política de Preservação de Acervos Digitais da UFRGS. Disponível em <[http://www.ufrgs.br/consun/legislacao/documentos/politica-de-preservacao-digital/at\\_download/file](http://www.ufrgs.br/consun/legislacao/documentos/politica-de-preservacao-digital/at_download/file)>. Acesso em 09 mar. 2021

WORLD HEALTH ORGANIZATION. Managing the Covid-19 infodemic: Promoting healthy behaviours and mitigating the harm from misinformation and disinformation. Disponível em <<https://www.who.int/news/item/23-09-2020-managing-the-covid-19-infodemic-promoting-healthy-behaviours-and-mitigating-the-harm-from-misinformation-and-disinformation>>, Acesso em 10 mar. 2021