

# PRESERVAÇÃO DIGITAL: CONSIDERAÇÕES SOBRE ARQUIVAMENTO DE CONTEÚDO DA WEB

**Antonio Paulo Carretta,**

Doutorando em Ciência da Informação pelo PPGCI ECA-USP

**RESUMO:** Artigo aborda questões sobre preservação e justificativas para arquivamento de conteúdo Web. Apresenta breve panorama de evolução da Web; formato de arquivamento (WARC), padrões e recomendações de preservação digital. Por fim, observa dimensões de preservação, seleção e políticas de Web arquivamento.

**PALAVRAS-CHAVE:** Preservação Digital; Arquivamento da Web; Preservação da Web.

## DIGITAL PRESERVATION: CONSIDERATIONS ON WEB ARCHIVING

**ABSTRACT:** Article presents digital preservation issues and justifications for Web archiving. It contains a brief overview of the evolution of the Web; archiving format (WARC), digital preservation standards and recommendations. Finally, it observes dimensions of preservation, selection, and Web archiving policies.

**KEYWORDS:** Digital Preservation; Web Archiving; Web Preservation.

## INTRODUÇÃO

Presente em nosso cotidiano digital, a Web é uma importante camada tecnológica com registros que constroem a memória de nossa sociedade. Para entender melhor esse quadro, vale a pena fazer a seguinte associação: assim como publicações impressas e outros documentos físicos compõem o patrimônio da humanidade, permitindo resgatar a memória de diversas épocas por meio da representação do pensamento e da cultura em sociedade, de maneira similar o conteúdo digital disponível na Web registra nosso passado mais recente e se constitui fonte de referência para estudos futuros, bem como herança cultural de uma nação.

Neste sentido, em 2003, a UNESCO reconheceu que estamos diante de um novo legado cultural: o patrimônio digital, que consiste em recursos únicos, derivados do conhecimento ou expressão de seres humanos, gerados exclusivamente em formato digital, e que incluem informações de diversas naturezas: cultural, educacional, científica e outras.

Porém, a condição de existência e manutenção do conteúdo Web depende de ações de preservação que minimizem as perdas contínuas de informação digital, decorrentes de erros de endereçamento de páginas Web, mudanças de plataformas, descontinuidade

de projetos ou desaparecimento de blogs e sites. Enfrentar com estratégias e políticas de preservação essa fragilidade do ambiente é o grande desafio dos países que pretendem manter registros de memória para futuras gerações.

Para refletir sobre esse tema, este artigo traça um breve panorama de evolução da Web; considera aspectos atrelados às tecnologias e estruturas da Internet, o formato adequado para arquivamento, recomendação de preservação e, por fim, levanta pontos de discussão sobre políticas de arquivamento, curadoria e reuso de conteúdo da Web.

## CONTEXTO HISTÓRICO E PANORAMA DA WEB

Passados mais de 30 anos da criação da Web, proposta por Tim Berners-Lee, observamos um ambiente de comunicação e disseminação de informação digital que se tornou parte do cotidiano das pessoas. A evolução ocorreu com velocidade, volume e variedade de conteúdo. De apenas uma página experimental, em 1990, atualmente existem mais de 1.5 bilhões de sites na World Wide Web. Desse total, segundo dados do órgão de controle de domínios nacionais, o Registro.br, a Web brasileira já representa mais de 4 milhões de domínios com terminação “.br”.

Entretanto, do total de sites no mundo, menos de 200 milhões estão ativos<sup>1</sup> e a explicação está no fato que, ao longo do tempo, os conteúdos publicados na Web podem perder a dinâmica de atualização ou atividade. Mike Ashenfelder, membro da Biblioteca do Congresso dos EUA, destaca que a vida útil de uma página na Web é de cerca de 100 dias, pouco mais de 3 meses (ROCHA, 2015). Essa natureza efêmera indica que depois dos 3 meses, muitos endereços deixam de ser atualizados ou perdem a frequência de postagens e acabam caindo no esquecimento.

Segundo estimativa do Internet Live Stats, que reforça ainda mais a fragilidade de manutenção do ambiente Web, 75% dos sites já criados estão vivos, mas não são ativos. Outros indicadores apontam que 80% das páginas web não estão disponíveis na sua forma original após 1 ano; 13% de artigos acadêmicos online desaparecem após 27 meses e 11% dos recursos de mídias sociais são perdidos depois de 1 ano (COSTA, GOMES, SILVA, 2016).

A percepção dessa condição é mais evidente ao buscarmos por páginas da Web e nos depararmos com o “Erro 404”, mensagem automática enviada pelo servidor de hospedagem do site avisando que determinada página não foi encontrada. Na opinião de especialistas, respectivamente em estrutura da internet e curadoria digital, estamos diante de um grande problema que exige analisarmos o impacto da ausência de ações de preservação da Web. Segundo Vint Cerf, vivemos o momento da chamada “era das trevas digital”, marcada pela incerteza de recuperação de conteúdos produzidos e que levanta uma questão estratégica e com projeção para o futuro: o que poderia acontecer

<sup>1</sup> Dados estatísticos indicados nos sites Internet Live Stats (<https://www.internetlivestats.com>) e Registro.br (<https://registro.br/dominio/estatisticas/>), consultados em julho de 2020.

com o conteúdo de publicações acadêmicas e diversas referências online, quando daqui a dez, 20 ou 50 anos os domínios mudarem, forem abandonados ou simplesmente deixarem de ser pagos? Nessa linha de raciocínio, Aquiles Brayner alerta que este século vai se configurando como a ‘Era do Erro 404’ e “o futuro tende a ser marcado pela escassez de fontes históricas. Não se pode pensar a história do século 21 sem levar em conta o que se é publicado na internet, incluindo as redes sociais” (ABDALLA, 2018).

Em oposição a esse contexto alarmante, ao longo do tempo de existência da Web, ações e serviços foram surgindo para preservação de parte da memória digital de nossa sociedade. Por exemplo, iniciado em 1996 e atuando como arquivo para armazenamento de páginas da Web, o Internet Archive é referência nos estudos sobre preservação digital e colaborou no desenvolvimento do atual formato WARC (Web ARChive), utilizado pelos participantes do International Internet Preservation Consortium (IIPC) como padrão para armazenamento de documentos gerados na Web (INTERNET ARCHIVE, 2020; INTERNATIONAL ORGANIZATION FOR STANDARDIZATION, 2009).

Por sua vez, o IIPC surgiu em 2003 para agregar esforços no desenvolvimento de recomendações e melhores práticas para preservação digital. Formado por instituições públicas, privadas e sem fins lucrativos, em sua maioria bibliotecas e arquivos nacionais, localizados em alguns países da África, Ásia, América do Norte, América do Sul (apenas o Chile), Europa e Oceania. Atualmente conta 37 países e 57 membros, dos quais 16 países declaram em sua legislação de depósito legal propósitos para preservação do conteúdo Web (IIPC, c2020). Aliás, o diferente contexto jurídico de cada país participante esbarra em questões legais e, por isso, apresentam diferentes diretrizes de preservação e autorizações de acessos aos documentos Web armazenados.

Entre os membros participantes do IIPC, podemos mencionar o caso do Arquivo.pt, serviço de arquivamento e preservação da Web portuguesa, que iniciou oficialmente em 2008. Este serviço contém páginas no idioma português desde 1996, mantém estrutura amigável de pesquisa, não apenas por url, mas também por termos, e oferece espaço para reflexão do tema, bem como estimula pesquisas inovadoras com base no reuso do conteúdo histórico arquivado (ARQUIVO.PT, 2020a).

Paralelamente, outras iniciativas e projetos especiais tem se preocupado com a preservação do conteúdo originado no ambiente Web. Com dimensões e propósitos diferentes, diversas instituições atuam em ações de preservação de seu próprio conteúdo, registros de redes sociais e informações temáticas, de contexto político ou cultural, como é o caso do consórcio The New York Art Resources Consortium (NYARC, 2020), que integra três importantes museus de arte na cidade de Nova York: Museu do Brooklyn, Coleção Frick e Museu de Arte Moderna. O projeto da NYARC arquivava coleções selecionadas de sites de exposições, em áreas que correspondem ao escopo e pontos fortes dessas instituições, bem como sites de projetos do consórcio e sites institucionais dos três museus.

Embora o Brasil tenha representatividade na Web e intuições públicas preocupadas

com o tema, como Arquivo e Biblioteca Nacional, alinhadas com programas de preservação digital, o país não possui uma legislação de apoio, nem serviços de abrangência nacional dedicados à preservação de informações do domínio.BR. Pontualmente, existem casos e projetos especiais relacionados com a Web brasileira (ROCKEMBACH, 2018), por exemplo, o Latin American Web Archiving Project, hospedado em endereço eletrônico da Universidade do Texas em Austin<sup>14</sup>, cuja coleta para arquivamento iniciou em 2005 e tem foco em documentos governamentais e de expressão política.

Outra iniciativa pontual foi do Grupo de Trabalho de Desenvolvimento de Conteúdo vinculado ao Consórcio Internacional de Preservação da Internet, que coletou sites, artigos, notícias, blogs e mídias sociais (Twitter, Facebook) sobre as Olimpíadas Rio 2016. O grupo realizou formulário público e incentivou pessoas para colaborarem com a seleção de temas relacionados às Olimpíadas 2016. A hashtag '#RIO2016WA' no Twitter também foi uma forma de marcar e acompanhar informações relativas ao arquivamento das Olimpíadas 2016. Resumidamente, este recorte histórico e o panorama apresentado trazem um contexto preocupante, principalmente para países que retardam o enfretamento desse processo de perda de memória digital. Embora o Brasil seja exemplo de discussões sobre governança da internet, educação e direitos nas redes digitais, faltam ainda estratégias e ações, concretas e colaborativas, que impulsionem um programa nacional para arquivamento de conteúdo da Web brasileira. Além disso, carece do fortalecimento de base teórico-empírica, apoiada por pesquisas nacionais e amplas reflexões nos campos da Ciência da Informação e Humanidades Digitais.

## **ARGUMENTOS E RECOMENDAÇÕES SOBRE ARQUIVAMENTO E PRESERVAÇÃO DA WEB**

O processo de arquivamento e preservação da Web é amplo e complexo. Para refletir sobre sua função e a natureza desse recurso, podemos destacar alguns conceitos ou argumentos de apoio: memória digital, o formato digital e as recomendações de preservação de conteúdos em ambientes digitais.

Segundo Vera Dodebei (2006), na Web a característica de acumulação do conhecimento ocorre no domínio coletivo, provocando uma permanente construção e reconstrução da informação, gerando não apenas bancos de dados ou bases de dados, mas centros de conhecimento compartilhado. Neste contexto, a Web passa ser vista como espaço da memória social, permitindo pesquisas sobre comportamento e meios de produção do conhecimento, seja ele de natureza histórica, artística ou técnica. Este domínio coletivo “faz da Web um grande centro virtual da memória do mundo.”

Analisando essa construção de uma memória digital, Dodebei (2006) elabora uma questão essencial: “Como propiciar a preservação das memórias que circulam na web?” Nesta indagação a autora deixa claro que “a existência do patrimônio digital e sua

permanência na memória virtual do mundo estarão intimamente ligadas às condições de preservação que forem proporcionadas por aqueles que se responsabilizarem pela inclusão e manutenção de um objeto na rede mundial.” Muito embora exista consciência da quase impossibilidade de preservação total da informação em meio digital, grandes são os esforços para manter um processo de transmissão para outras gerações, impedindo seu desaparecimento, absorção ou dispersão entre domínios hegemônicos, de ordem idiomática ou geopolítica. Para evitar isso, é necessário não apenas favorecer a migração de informação para novos suportes, mas dotar os objetos digitais de significado, preservando seu conteúdo intelectual e garantindo sua autenticidade e acessibilidade.

Preservar a integridade de um documento online é outro ponto essencial de análise. A Web opera com camadas. Percebemos que a busca por uma melhor solução para lidar com o problema dos dados estruturados, considerando sua migração para um formato de arquivamento adequado, está baseada na moderna construção de websites que, basicamente, funcionam com uma arquitetura em três camadas: camada de metadados oferecem estrutura e capacidade de recuperação do conteúdo; camada de apresentação dos dados, definida pela codificação em HTML do documento acessado; e camada lógica, definida pelo software que permite comunicação, leitura e exibição dos elementos das camadas de dados e apresentação (MASANÉS, 2006).

Essa estrutura em camadas é complexa e opera de forma dinâmica. Uma página da Web pode conter vários elementos (texto, vídeo, áudio, imagens), que refletem um esforço editorial e estético, por padrão ou livre desenvolvimento, e pode conter diferentes combinações de visualização e interatividade. Todas estas características precisam ser mantidas ao migrarmos esse conteúdo para um formato de preservação. No caso da Web, para fins de armazenamento, o formato atualmente utilizado é WARC (Web ARChive), cujo requisito geral permite carregar de forma segura todos os objetos digitais constituintes da estrutura do site, facilitando o processo de coleta, acesso e organização do conteúdo armazenado.

O formato WARC oferece um padrão estruturador para gestão e armazenamento de milhões de recursos digitais coletados da web, por meio da construção de aplicativos para coleta, gerenciamento, acesso, mineração e troca de conteúdo. Embora seja considerado um formato exclusivo para arquivos da web, pode ainda ser adotado para armazenar objetos natos digitais ou digitalizados. Adicionalmente, a maneira como os arquivos WARC são criados e os recursos armazenados depende do uso de software e aplicativos específicos.

Segundo a norma ISO 28500:2009 (INTERNATIONAL ORGANIZATION FOR STANDARDIZATION, 2009) o formato de arquivo WARC tem como propósito:

- armazenar o conteúdo da carga útil (corpo de dados) e as informações de controle dos principais protocolos da camada de aplicativos da Internet, como o HTTP (Hypertext Transfer Protocol), o DNS (Sistema de Nomes de Domínio) e o FTP (File Transfer Protocol);

- armazenar metadados complementares vinculados a outros dados armazenados (por exemplo, classificador de assunto, idioma e codificação);
- apoiar a compactação de dados e manter a integridade do registro de dados;
- armazenar todas as informações de controle do protocolo de coleta (por exemplo, cabeçalhos de solicitação), não apenas informações de resposta;
- armazenar os resultados de transformações de dados vinculados a outros dados armazenados;
- armazenar indicações de duplicidade vinculadas a outros dados armazenados (para reduzir o armazenamento na presença de recursos idênticos ou substancialmente semelhantes);
- ser estendido sem interromper funcionalidades existentes;
- suportar o manuseio de registros excessivamente longos por truncamento ou segmentação, quando desejado.

Por conta dessas características e a natureza que adquire esse formato, outro conceito pode ser empregado para definir os arquivos que surgem da transição estrutural: o renascido digital. Segundo Brügger (2017), materiais digitais podem ter diferentes processos de formação: digitalizado, nato digital e renascido digital.

O material digital renascido é um material digitalizado ou nato digital que foi coletado e preservado, mas em algum grau foi alterado neste processo, por exemplo, como acontece após a transição de um formato para um arquivo da Web. Considerando essa tríade digital, a Web pode desempenhar três papéis diferentes: 1º como plataforma para distribuição de materiais analógicos digitalizados; 2º como fonte que já nasce de processo e formato digital, e 3º como fonte digital renascida em um arquivo da Web, estruturado com propósito de armazenamento e preservação.

Nas últimas décadas, propostas para organização e armazenamento da memória digital tentam encontrar maneiras adequadas para coletar, em grande escala, a imensa quantidade de conteúdo gerado na Web. Para auxiliar nessa tarefa, são utilizadas ferramentas como “crawlers”, rastreadores que navegam por websites de maneira automatizada e conforme um conjunto de políticas de coleta. O fluxo de operação desses rastreadores segue a seguinte lógica: a partir de uma lista de URLs, este programa salva cada página contida em determinada URL, encontra os hiperlinks da página (links para outros endereços, imagens, vídeos, scripts ou alguma instrução do código fonte) e os adiciona em uma lista controlada de URLs que visita recursivamente para novas ações de coleta.

Esse processo para armazenar e gerenciar tal quantidade de páginas da Web representa um grande desafio. Por conta disso, diversas instituições de memória definem algumas recomendações e práticas que podem auxiliar a operação dos rastreadores e, de certa forma, tornar um website mais ou menos preservável. Entre as recomendações

sugeridas pelo Arquivo.pt (2020b) podemos apontar:

1. Utilize diretrizes de acessibilidade, além disso, evite lacunas na identificação do conteúdo, por exemplo: observe se incluiu corretamente a data de publicação dos documentos;
2. Defina códigos e use provedores que ajudem os rastreadores de arquivamento, para isso, use corretamente o protocolo de exclusão de robots > robots.txt é um arquivo, salvo na pasta raiz do site, que serve para dar ordens específicas aos robôs de busca e rastreadores indicando páginas do site que podem (ou não) ser acessadas para arquivamento e apresentam conteúdo textual relevante;
3. Considere usar uma licença Creative Commons, indicações que expressam claramente quando determinado conteúdo é livre para uso ou possui restrições de coleta;
4. Mantenha URLs estáveis e redirecione quando necessário. Qualquer mudança na URL pode tornar um site inacessível e fora do alcance dos rastreadores; use um endereço para cada conteúdo;
5. Utilize formatos adequados e recomendados para preservação, por exemplo: HTML ou XML (texto), PNG (imagem), AVI (vídeo);
6. Publique metadados para enriquecer os conteúdos, por exemplo, metadados descritivos do padrão Dublin Core, estas informações enriquecem ou complementam conteúdos capturados. Este conjunto de aspectos nos ajuda entender a complexidade de preservação da Web e quais protocolos apoiam o processo de arquivamento. No entanto, para pensarmos em ações mais amplas, a definição de políticas de coleta, estratégias de curadoria e acesso ao conteúdo arquivado são fundamentais para proteção do patrimônio digital em evolução.

## **POLÍTICAS DE ARQUIVAMENTO WEB, CURADORIA E REUSO DE INFORMAÇÃO**

Abrindo a discussão sobre políticas e estratégias de preservação da herança digital, Masanès (2006) considera 3 pontos críticos muito comuns em argumentações iniciais contra o arquivamento de conteúdo da Web: a qualidade do conteúdo publicado; a condição tecnológica de preservação do ambiente digital e a imensurável tarefa de armazenar e permitir acesso ao conteúdo preservado. Como contraponto a complexidade dessas questões, destacamos categorias possíveis para seleção de conteúdo, dimensões estruturais do processo de preservação digital e os desafios para promover políticas de preservação da Web brasileira.

De forma objetiva, podemos definir o arquivamento da web como um processo que compreende coletar, armazenar e disponibilizar a informação retrospectiva da World Wide Web para futuros pesquisadores. Para isso, a seleção é uma questão fundamental da web arquivamento. Dito de outra forma, o arquivamento na Web vai além das questões tecnológicas e da mera coleção de material da Internet. Questões relacionadas ao uso precisam ser vistas pela perspectiva organizacional, funcional e tecnológica. Pensar no uso

de uma iniciativa de arquivamento na Web abrange da seleção e coleta, ao gerenciamento e preservação da informação (ROCKEMBACH, 2018; MASANÈS, 2006).

Nesse contexto, o processo de seleção para arquivamento pode ser classificado em quatro categorias (BRAYNER, 2016; ROCKEMBACH, PAVÃO, 2018):

- a) seleção extensiva (que abrange domínios inteiros como .fr ou .br);
- b) seleção intensiva (que coleta em profundidade, arquivando os diversos níveis e hierarquias de um site selecionado por estratégias de curadoria)
- c) seleção temática: definida por critérios para escolha de um tema ou evento relevante, por exemplo, a coleta de conteúdo sobre o impacto social da COVID- 19;
- d) seleção autorizada: sites que possuem diretrizes de coleta autorizadas por seus editores e autores.

Devido à natureza descentralizada e a rápida obsolescência tecnológica, muitos desafios são impostos para preservação da Web. O ponto de partida para enfrentá-los é o planejamento e a definição de políticas de arquivamento (se centrado em um tema, site ou domínio), a metodologia e as plataformas para coletar, armazenar e disponibilizar o conteúdo Web.

Estas palavras-chave (coletar, armazenar e disponibilizar) correspondem às etapas do processo de arquivamento Web que, em certa medida, conforme a Figura 1, nos permitem visualizar três dimensões de preservação entrelaçadas:

- Dimensão política > baseada em questões de governança da internet, legislação, normas e padrões para coleta de conteúdo, procedimentos de preservação digital autorizada e condições de acesso;
- Dimensão tecnológica > baseada em tecnologias de preservação, aplicação de metadados, capacidade de interoperabilidade e nas ferramentas disponíveis para rastreamento, transição de formato e recuperação do conteúdo armazenado;
- Dimensão curatorial > baseada na escolha de estratégias para produção de conteúdo da Web, medidas de preservação seletiva ou integral, formas de exploração e reuso do conteúdo armazenado.



Figura 1: Dimensões de preservação do conteúdo Web.

Fonte: Elaborado pelo autor.

Estas dimensões configuram um quadro complexo para elaboração de estratégias e políticas de preservação e, quando projetados no contexto nacional, esbarramos numa lista de desafios (DANTAS, 2015):

- Apoio governamental e envolvimento de empresas públicas e privadas
- Definição de técnicas e tecnologias acessíveis e seguras
- Brechas jurídicas (por exemplo, regulamentação da Lei de Depósito Legal 10994/04 e 12192/10, que define: “publicações produzidas por qualquer meio ou processo, para distribuição gratuita ou venda”)
- Formação de competências profissionais e ampla inserção em programas acadêmicos
- Sustentabilidade econômica e tecnológica de serviços de arquivamento nacional
- Avaliação e monitoramento sistemático de coletas nacionais
- Acesso e difusão do conteúdo Web armazenado

Cabe ainda inserir nesse contexto de desafios a preocupação apresentada no relatório “Best Practice Forum on Local Content”, produzido em 2020 pelo Fórum de Governança da Internet. Nesse documento são destacadas questões sobre proteção, preservação e promoção das línguas locais e indígenas e do patrimônio cultural na era digital. Considerando a existência de conteúdos digitais vulneráveis, que registram idiomas ou práticas culturais de diversas comunidades, como de indígenas e artesões populares brasileiros, podemos integrar ao escopo de ações a segurança e preservação de heranças

culturais fragilizadas (IGF, 2020).

Buscar respostas para esses desafios indicam o primeiro passo para o desenvolvimento de um serviço de preservação e armazenamento da Web brasileira e, conseqüentemente, formação do patrimônio digital nacional.

## CONSIDERAÇÕES FINAIS

É inevitável mencionar que as interações sociais em ambiente digital intensificaram a produção e disseminação de informação. O contexto da sociedade digital indica, agora de forma mais evidente, que lidamos com um intenso processo de transformação do comportamento digital adquirido na última década. Além disso, a exemplo de outros momentos históricos, existem esforços mundiais para arquivar as diversas narrativas digitais da memória social: ações governamentais, conteúdo científico, textos jornalísticos, recursos educacionais, relações profissionais e trabalhistas, manifestações sociais em rede e seu impacto no cotidiano das pessoas.

Tomando esse momento como referência, ressaltamos a importância do arquivamento de conteúdo da Web não apenas para construção de registros digitais de memória, mas também para evitar um processo de esquecimento digital. Por essa perspectiva, parece urgente desenvolver projetos estruturais de apoio ao Web arquivamento, assim como ampliar abordagens de pesquisas sobre esse tema e analisar ações concretas sobre o patrimônio digital nacional, a amplitude de governança da internet, direitos e garantias de preservação, arquivamento, acesso à informação e ao conhecimento produzido na Web brasileira.

## REFERÊNCIAS

ABDALLA, Anita. Era do Erro 404: vamos perder a história registrada na internet? **UOL Tecnologia**, 15/08/2018. Disponível em: <https://www.uol.com.br/tilt/noticias/redacao/2018/08/15/a-era-do-erro-404-ja-esta-entre-nos.htm>. Acesso em: 20 jun. 2020.

ARQUIVO.PT. Arquivo.pt: informações acerca do serviço de pesquisa sobre o passado. 18/03/2020a. Disponível em: <https://sobre.arquivo.pt/pt/> Acesso em: 12 jun 2020.

\_\_\_\_\_. Criação e gestão de sites preserváveis: Recomendações do Arquivo.pt. [Slide]. 19/06/2020b. Disponível em: [https://sobre.arquivo.pt/wp-content/uploads/Sessao11\\_versao1.pdf](https://sobre.arquivo.pt/wp-content/uploads/Sessao11_versao1.pdf) Acesso em: 26 jun.2020.

BRAYNER, Aquiles Alendar. UK Web Archive Programme. A Brief History of Opportunities and Challenges. **Revista Digital de Biblioteconomia e Ciência da Informação (RDBCI)**. Campinas, v. 14, nº 2, mai/ago. 2016, p. 318–333. Disponível em: <<https://periodicos.sbu.unicamp.br/ojs/index.php/rdbci/article/view/8645982>>. Acesso em: 10 jun. 2020.

- BRÜGGER, Niels. When the present web is later the past: web historiography, digital history, and internet studies. **Historical Social Research**, 37 (2012) 4. p. 102–117.
- COSTA, Miguel; GOMES, Daniel; SILVA, Mário J. **The evolution of web archiving** **International Journal on Digital Libraries**, p. 1-15, 2016.
- DANTAS, Camila Guimarães. **Criptografias da memória**: um estudo teórico-prático sobre o arquivamento da web no Brasil. 2015.
- DODEBEI, Vera. Patrimônio e Memória Digital. **Morpheus** - Revista Eletrônica em Ciências Humanas, v.4, n.8, 2006.
- INTERNATIONAL INTERNET PRESERVATION CONSORTIUM (IIPC), c2020. Disponível em: <https://netpreserve.org/about-us/> Acesso em: 15 jun. 2020.
- INTERNATIONAL ORGANIZATION FOR STANDARDIZATION. **ISO 28500:2009. Information and documentation - WARC file format**. Geneva: ISO, 2009. Disponível em: <https://www.iso.org/obp/ui/#iso:std:iso:28500:ed-1:v1:en> Acesso em: 18 jun.2020.
- INTERNET ARCHIVE. About the Internet Archive. Disponível em: <https://archive.org/about/> Acesso em: 9 jun. 2020.
- IGF – INTERNET GOVERNANCE FORUM. **Best Practice Forum on Local Content**: Local and indigenous content in the digital space: Protection, preservation and sustainability of creative work and traditional knowledge. United Nations, December 2020. Disponível em: [https://www.intgovforum.org/multilingual/index.php?q=filedepot\\_download/5005/2372](https://www.intgovforum.org/multilingual/index.php?q=filedepot_download/5005/2372) Acesso em: 20 jan.2021.
- MASANÈS, Julien. **Web Archiving**. Berlin, Heidelberg, Springer, 2006.
- NEW YORK ART RESOURCES CONSORTIUM (NYARC). About [Página]. Disponível em<<http://nyarc.org/content/about>> Acesso em 22 jun. 2020.
- ROCHA, Caroline. Quantos sites existem na internet? **OLHAR DIGITAL**, 30/09/2015. Disponível em: <https://olhardigital.com.br/noticia/quantos-sites-existem-na-internet/51803> Acessado em: 15 jul. 2020.
- ROCKEMBACH, Moisés. Arquivamento da web: estudos de caso internacionais e o caso brasileiro. **Revista Digital de Biblioteconomia e Ciência da Informação (RDBCI)**. Campinas, v. 16, nº 1, já. /abr. 2018, p. 7-24. Disponível em:<<https://periodicos.sbu.unicamp.br/ojs/index.php/rdbci/article/view/8648747/pdf>>. Acesso em: 10 jun. 2020.
- ROCKEMBACH, Moisés; PAVÃO, Caterina Marta Groposo. Políticas e tecnologias de

preservação digital no arquivamento da web. **RICI**: R. Ibero-Americana de Ciência da Informação. Inf., ISSN 1983-5213, Brasília, v. 11, n. 1, p. 168-182, jan. /abril.2018. Disponível em: < <https://lume.ufrgs.br/handle/10183/175153>>. Acesso em: 11 jun. 2020.

UNESCO. **Carta sobre la preservación del patrimonio digital** [Anexo N° 7], 15 out. 2003. Disponível em: <http://www.culturaydeporte.gob.es/planes-nacionales/dam/jcr:e840703d-133c-42e2-b3a2-90ab13e3bcdf/carta-sobre-la-preservaci-n-del-patrimonio-digital.pdf>. Acesso em: 20 jun. 2020.