

UNIVERSIDADE FEDERAL FLUMINENSE
INSTITUTO DE ARTE E COMUNICAÇÃO SOCIAL
DEPARTAMENTO DE CIÊNCIA DA INFORMAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO
MESTRADO EM CIÊNCIA DA INFORMAÇÃO

GABRIELA AYRES FERREIRA TERRADA

PRESERVAÇÃO DIGITAL DA *WEB*:
uma reflexão sobre políticas e práticas



NITERÓI
2022

GABRIELA AYRES FERREIRA TERRADA

PRESERVAÇÃO DIGITAL DA WEB:
uma reflexão sobre políticas e práticas

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Informação, vinculado ao Instituto de Arte e Comunicação Social da Universidade Federal Fluminense, como requisito parcial para obtenção de título de Mestre em Ciência da Informação.

Área de Concentração: Dimensões contemporâneas da informação e do conhecimento.

Linha de Pesquisa 2: Fluxos e Mediações Sociotécnicas da Informação.

Orientadora:
Prof.^a Dr.^a Regina de Barros Cianconi

Niterói
2022

Ficha catalográfica automática - SDC/BCG
Gerada com informações fornecidas pelo autor

T323p Terrada, Gabriela Ayres Ferreira
Preservação digital da web : uma reflexão sobre
políticas e práticas / Gabriela Ayres Ferreira Terrada ;
Regina de Barros Cianconi, orientadora. Niterói, 2022.
214 f.

Dissertação (mestrado)-Universidade Federal Fluminense,
Niterói, 2022.

DOI: <http://dx.doi.org/10.22409/PPGCI.2022.m.11797917706>

1. Preservação digital. 2. Patrimônio digital. 3.
Preservação da web. 4. Arquivamento da web. 5. Produção
intelectual. I. Cianconi, Regina de Barros, orientadora. II.
Universidade Federal Fluminense. Instituto de Arte e
Comunicação Social. III. Título.

CDD -

GABRIELA AYRES FERREIRA TERRADA

PRESERVAÇÃO DIGITAL DA *WEB*:
uma reflexão sobre políticas e práticas

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Informação, vinculado ao Instituto de Arte e Comunicação Social da Universidade Federal Fluminense, como requisito parcial para obtenção de título de Mestre em Ciência da Informação.

Aprovada em: 17 de maio de 2022.

BANCA EXAMINADORA

Prof.^a Dr.^a Regina de Barros Cianconi – Presidente/Orientadora
Universidade Federal Fluminense – UFF

Prof. Dr. Vitor Manoel Marques da Fonseca – Titular interno
Universidade Federal Fluminense – UFF

Prof. Dr. Moisés Rockembach – Titular externo
Universidade Federal do Rio Grande do Sul – UFRGS

Prof.^a Dr.^a Michely Jabala Mamede Vogel – Suplente interno
Universidade Federal Fluminense – UFF

Prof. Dr. Gildenir Carolino Santos – Suplente externo
Universidade Estadual de Campinas – UNICAMP

Niterói

2022

Dedico este trabalho a todas as pessoas que amo.

AGRADECIMENTOS

Em primeiro lugar, a Deus, pela benção de viver que me concede todos os dias e pela graça de poder concluir finalmente o mestrado, sobrevivendo a uma pandemia.

A toda minha família pelo incentivo e apoio durante todo o tempo que dediquei à Universidade e a esta pesquisa, principalmente à minha mãe Neuzeli e à minha irmã Juliana. Ao meu marido Alejandro, pelo seu amor, carinho, amizade, apoio (nos momentos tristes e felizes), pela paciência, enfim, por permitir fazer parte de sua vida e agora aumentando nossa família com a chegada de Teodora. Amo muito vocês!

Aos amigos da família e dos tempos do colégio que não esqueço nem esquecerei, em especial, Rovani e Ana. Aos amigos que a UFF me deu, particularmente, Thays Clarindo.

Agradeço à minha estimada orientadora Regina de Barros Cianconi pelo estímulo e colaboração para a minha formação e, sobretudo, por topar o desafio de tratar de um tema tão complexo no mestrado durante esse período de pandemia. Pela amizade e confiança no meu trabalho e pelos ensinamentos e orientação que foram fundamentais na realização desta pesquisa.

Agradeço a todos os professores da UFF que, sempre dispostos, ajudaram e contribuíram para a minha formação, em especial, aos professores do Departamento de Ciência da Informação; e aos membros da banca por aceitarem o convite, assim como pelas colaborações para com a pesquisa.

Aos colegas da UFF, que contribuíram para a minha formação, nomeadamente, Darlene, Mônica e aos colegas do mestrado Carina, Jessica Gonçalves, Jessica Nascimento, Nina pelos trabalhos realizados em grupo, pelo apoio e conselhos de Luiza e Raphael.

Aos colegas de trabalho, que colaboraram para meu crescimento profissional. Especialmente, aos meus amigos arquivistas digitais Well e Danielle Peçanha, e por toda ajuda e incentivo de Fernanda, Allan, Elen, Thais, Raissa, Daniele Cavaliere, Andrea, Ju, Sonia, Pati e colegas da FBN. *In memoriam* de Angela Monteiro Bettencourt e amizade de Liana Amadeu.

Ao apoio e estímulo de Suely Dias, coordenadora-geral do Centro de Processamento e Preservação, e equipe da BNDigital da Fundação Biblioteca Nacional.

Aos grupos de pesquisa Dríade da Rede Cariniana, Rede Sudeste e à Câmara Técnica Consultiva do Conarq.

“A sabedoria é um paradoxo.

*O homem que mais sabe é aquele que mais
reconhece a vastidão da sua ignorância.”*

Friedrich Nietzsche

RESUMO

A *web* é, atualmente, uma propagadora de informação e de conteúdos importantes em diversas áreas e setores da sociedade. Em um pouco mais de duas décadas, ela cresceu e converteu-se em um meio de informação global e instrumento de comunicação e registro da vida no século XXI. Há uma aparente impressão de solidez no ambiente digital, porém, pesquisas apontam o desaparecimento de páginas *web* em curtos períodos. A presente pesquisa teve como tema a preservação do patrimônio digital, especificamente, a preservação e o arquivamento de páginas *web*. Nesse sentido, o seguinte questionamento norteou a investigação: onde e como estão sendo preservadas as páginas *web*? Teve como objetivo geral estudar a preservação do patrimônio digital por instituições de guarda da memória das páginas *web*, que fazem parte do *International Internet Preservation Consortium* – IIPC, com base na *Carta sobre a Preservação do Patrimônio Digital* da Unesco. Adotaram-se como procedimentos metodológicos as pesquisas exploratórias e descritivas, baseadas em fontes bibliográficas e documentais. O universo da parte empírica delimitou-se nas instituições com membros fundadores e ibero-americanas do IIPC. Buscou-se conceituar e caracterizar a preservação do patrimônio digital e alguns temas relacionados, como: evolução da internet e da *web*; patrimônio digital no âmbito arquivístico, das bibliotecas nacionais e do IIPC; arquivamento da *web* e os seus atores no contexto internacional e nacional; preservação do patrimônio digital e as etapas para arquivamento e preservação da *web*. Outrossim, identificar e caracterizar as instituições de guarda da memória fundadoras do IIPC e as ibero-americanas por meio de quadros. Examinaram-se políticas, ações e critérios de preservação digital, utilizados pelas instituições ibero-americanas do IIPC com enfoque na governança institucional e identificaram-se os problemas e as barreiras à preservação das páginas *web*, com reflexões sobre as referidas instituições, alcançando dez categorias de problemas e barreiras com base nos textos da pesquisa bibliográfica e documental e nas respostas das instituições pesquisadas. Depreendeu-se que as bibliotecas nacionais localizadas no continente europeu lideram a tarefa de preservar a *web* dentro do escopo pesquisado, e dividem o trabalho com outros tipos de instituições, incluindo arquivos nacionais. Como exceção, a tarefa é realizada por organizações sem fins lucrativos ou ligadas ao ensino e à pesquisa, utilizando-se do marco de depósito legal para o recolhimento desses tipos de conteúdos, repetindo regras aplicadas nos documentos analógicos, como acesso e uso restrito ao local físico institucional. A preservação digital da *web* é um desafio tanto para os profissionais quanto para as instituições, levando-se em conta que nem sempre a teoria anda no mesmo ritmo que a prática. Concluiu-se que as instituições de memória, como bibliotecas e arquivos, devem refletir e reagir, como a Unesco bem colocou, liderando esse movimento de preservação do patrimônio digital, pois, ainda que o desafio seja imenso, deve-se considerar esses objetos complexos como parte do patrimônio nacional.

Palavras-chave: Preservação digital. Patrimônio digital. Arquivamento da *web*. Preservação da *web*.

ABSTRACT

The web is currently an important propagator of information and content in several areas and sectors of society. In a little more than two decades, it has grown to become a global information medium, a communication tool, and record of life in the 21st century. Although there are perceptions of the alleged solidity of the digital environment, surveys point to the disappearance of websites in short periods of time. This research was concerned with the preservation of digital heritage, specifically the preservation and archiving of webpages. Therefore, the following questions guided the research: where and how are the webpages being preserved? The general objective was to study digital heritage preservation by webpages memory institutions that are part of the International Internet Preservation Consortium – IIPC, based on UNESCO's Charter on Digital Heritage Preservation. The methodological procedures adopted are exploratory and descriptive research, based on bibliographic and document sources. The universe of the empirical part was delimited to the founding and Ibero-American member institutions of the IIPC. We tried to conceptualize and characterize the preservation of digital heritage and some related topics, such as the evolution of the internet and the web; digital heritage in the context of archives, national libraries, and the IIPC; web archiving and its actors in the international and national context; preservation of digital heritage and the steps for archiving and preserving the web. Furthermore, identifying and characterizing the IIPC's founding memory-keeping institutions and Ibero-American ones through frameworks. The policies, actions, and criteria for digital preservation, used by the Ibero-American institutions of the IIPC, were examined with a focus on institutional governance and the problems and barriers to the preservation of webpages were identified, with reflections on those institutions, reaching ten categories of problems and barriers based on the texts of the bibliographic and documentary research and the responses of the researched institutions. It is concluded that national libraries, located in the European continent, lead the task of preserving the web within the researched scope that shares the work with other types of institutions, including national archives. Exceptionally, the task is performed by non-profit organizations or those linked to teaching and research, using the legal deposit framework for the collection of those types of contents, repeating rules applied to analog documents, such as access and use restricted to the institutional physical location. Digital preservation of the web is a challenge for both professionals and institutions, and theory does not always move at the same pace as practice. We conclude that heritage institutions, such as libraries and archives must reflect and react, as UNESCO has well put it, leading this digital heritage preservation movement, because even though the challenge is immense, these complex objects must be considered as part of Brazil's national heritage.

Keywords: Digital preservation. Digital heritage. Web archiving. Preservation of the *web*.

LISTA DE FIGURAS

Figura 1 – Fluxo de Arquivamento da <i>web</i>	36
Figura 2 – Arquivamento da <i>web</i>	38
Figura 3 – Modelo de ciclo de vida do arquivamento da <i>web</i>	39
Figura 4 – Matriz do método de recolha	42
Figura 5 – Arquivamento lado do cliente	43
Figura 6 – Arquivamento lado do servidor.....	44
Figura 7 – Recolha de arquivo transacional	44
Figura 8 – Contexto de ação da preservação	77
Figura 9 – Prioridades para ação de preservação	78
Figura 10 – Linha do tempo Modelo OAIS.....	92
Figura 11 – Ambiente do Modelo OAIS	93
Figura 12 – Conceitos e relacionamentos do pacote de informação	94
Figura 13 – Interações entre os pacotes de informação no OAIS/SAAI.....	97
Figura 14 – Esquema do modelo funcional	98
Figura 15 – FAIR <i>versus</i> TRUST.....	100
Figura 16 – Critérios para inclusão de Metadados no contexto <i>web</i>	104
Figura 17 – Lista de Padrões de Metadados	107
Figura 18 – Partes do WARC-Type	111
Figura 19 – Definições das partes do WARC-Type.....	112
Figura 20 – Diagrama da infraestrutura geral SPAR.....	120
Figura 21 – Diagrama Funcional SPAR.....	121
Figura 22 – Funcionalidades da Ferramenta.....	124
Figura 23 – Arquivamento da <i>web</i> pelo <i>Archive-It</i>	125
Figura 24 – Volume de recolhimento anual do UKWA.....	129
Figura 25 – Pilares do Programa de Preservação Digital da LAC	141

LISTA DE QUADROS

Quadro 1 – Síntese e quantificação de termos de busca.....	23
Quadro 2 – Resumo das etapas de pesquisa.....	25
Quadro 3 – Concepção de autores sobre o processo de arquivamento da <i>web</i>	42
Quadro 4 – Os cinco tipos da <i>Preservation Description Information</i>	95
Quadro 5 – Metadados.....	101
Quadro 6 – Os elementos de metadados WAM.....	105
Quadro 7 – Campos obrigatórios WARC.....	110
Quadro 8 – Instituições Membro Fundadoras do IIPC.....	116
Quadro 9 – Instituições Ibero-Americanas do IIPC.....	117
Quadro 10 – Caracterização das Instituições Membros do IIPC.....	156
Quadro 11 – Caracterização das Instituições Ibero-Americanas do IIPC.....	160
Quadro 12 – Endereço de contato das instituições do IIPC.....	162
Quadro 13 – Enfoque na Governança Institucional.....	163

LISTA DE ABREVIATURAS E SIGLAS

AIC	<i>American Institute for Conservation</i>
AIP	<i>Archival Information Package</i>
ALA	<i>American Library Association</i>
API	<i>Application Programming Interface</i>
ARPA	<i>Advanced Research Projects Agency</i>
BAnQ	<i>Bibliothèque et Archives nationales du Québec</i>
BNC	Biblioteca Nacional do Chile
BNE	Biblioteca Nacional da Espanha
BnF	<i>Bibliothèque nationale de France</i>
CCSDS	<i>Consultative Committee for Space Data Systems</i>
CONARQ	Conselho Nacional de Arquivos
DCC	<i>Digital Curation Center</i>
DIP	<i>Dissemination Information Package</i>
DL	<i>Depósito Legal</i>
DPC	<i>Digital Preservation Coalition</i>
EUA	Estados Unidos da América
FAIC	<i>Foundation for Advancement in Conservation</i>
FAIR	<i>Findable, Accessible, Interoperable and Reusable</i>
FAPESP	Fundação de Amparo a Pesquisado Estado de São Paulo
FBN	Fundação Biblioteca Nacional
FCCN	Fundação para Computação Científica Nacional
FCT	Fundação para Ciência e Tecnologia
HTML	<i>HyperText Markup Language</i>
HTTP	<i>Hypertext Transfer Protocol</i>
ICA	<i>International Council on Archives</i>
ICOM	<i>International Council of Museums</i>
IFLA	<i>International Federation of Library Associations and Institutions</i>
IIPC	<i>International Internet Preservation Consortium</i>

INA	<i>Institut National de l'Audiovisuel</i>
IWAW	<i>International Web Archiving Workshops</i>
JISC	<i>Joint Information Systems Committee</i>
NDIIPP	<i>National Digital Information Infrastructure and Preservation Program</i>
NDSA	<i>National Digital Stewardship Alliance</i>
NISO	<i>National Information Standards Organization</i>
NUAWEB	Núcleo de Pesquisa em Arquivamento da Web e Preservação Digital
OAIS	<i>Open Archival Information System</i>
OCLC	Online Computer Library Center
PADICAT	Patrimonio Digital de Catalunya
PDI	<i>Preservation Description Information</i>
RAID	<i>Redundant Array of Inexpensive Drives</i>
RESAW	<i>Research Infrastructure for the Study of Archived Web Materials</i>
RLG	<i>Research Library Group</i>
SAA	<i>Society of American Archivists</i>
SAAI	Sistema Aberto de Arquivamento de Informação
SINPRED	Seminário Internacional de Preservação Digital
SIP	<i>Submission Information Package</i>
SPAR	<i>Système de Préservation et d'Archivage Réparti</i>
TRUST	<i>Transparency, Responsibility, User Focus and Tecnology</i>
UE	União Europeia
UFRGS	Universidade Federal do Rio Grande do Sul
UNESCO	Organização das Nações Unidas para a Educação, a Ciência e a Cultura
URI	<i>Uniform Resource Identifier</i>
URL	<i>Uniform Resource Locator</i>
W3C	<i>World Wide Web Consortium</i>
WALCM	<i>Web Archiving Life Cycle Model</i>
WAM	<i>Web Archiving Metadata Working Group</i>
WWW	<i>World Wide Web</i>

SUMÁRIO

1 INTRODUÇÃO	15
1.1 OBJETIVOS.....	19
1.1.1 Objetivo geral.....	19
1.1.1 Objetivos específicos.....	20
2 METODOLOGIA.....	22
3 EVOLUÇÃO DA INTERNET E DA WEB	26
4 ARQUIVAMENTO DA WEB	30
4.1 PROCESSOS E DESAFIOS DO ARQUIVAMENTO.....	33
4.2 ATORES DA PESQUISA SOBRE ARQUIVAMENTO NA WEB	47
4.2.1 Contexto Internacional.....	47
4.2.2 Contexto brasileiro	53
5 PATRIMÔNIO DIGITAL	59
5.1 PATRIMÔNIO ARQUIVÍSTICO DIGITAL	66
5.2 PATRIMÔNIO DIGITAL NAS BIBLIOTECAS NACIONAIS	71
6 PRESERVAÇÃO DO PATRIMÔNIO.....	75
6.1 PRESERVAÇÃO DIGITAL – CONCEITOS E PERSPECTIVAS	79
7 ETAPAS PARA ARQUIVAMENTO E PRESERVAÇÃO DA WEB	92
7.1 OPEN ARCHIVAL INFORMATION SYSTEM	92
7.2 METADADOS	100
7.3 PADRÃO DE ARQUIVAMENTO: ISO 28500:2017	109
7.4 CONSÓRCIO INTERNACIONAL DE PRESERVAÇÃO DA INTERNET	113
8 ANÁLISE E DISCUSSÃO DOS RESULTADOS	116
8.1 CARACTERIZAÇÃO DAS AÇÕES DE GUARDA E PRESERVAÇÃO NAS INSTITUIÇÕES MEMBROS FUNDADORAS DO IIPC	117
8.1.1 National Library of France e Institut National de l’Audiovisuel.....	117
8.1.2 Internet Archive.....	122
8.1.3 The British Library e The National Archives, U.K.....	126
8.1.4 Royal Danish Library.....	132
8.1.5 National Library of Finland	134
8.1.6 National Library of Sweden	136
8.1.7 National and University Library of Iceland.....	137
8.1.8 The National Library of Norway	137

8.1.9 Library and Archives Canada e National Library and Archives of Quebec	138
8.1.10 Library of Congress.....	142
8.1.11 National Library of Australia.....	144
8.2 POLÍTICAS, CRITÉRIOS E AÇÕES DE PRESERVAÇÃO DIGITAL DAS INSTITUIÇÕES IBERO-AMERICANAS DO IIPC	145
8.2.1 Arquivo.pt – Portuguese Web Archive.....	146
8.2.2 Biblioteca Nacional da Espanha.....	148
8.2.3 Biblioteca da Catalunya	150
8.2.4 Biblioteca Nacional do Chile.....	152
8.3 DISCUSSÃO DOS RESULTADOS	153
9 CONSIDERAÇÕES FINAIS.....	171
REFERÊNCIAS	176
APÊNDICE	208
APÊNDICE A – E-MAIL PARA AS INSTITUIÇÕES IBERO-AMERICANAS DO IIPC	209
ANEXO.....	210
ANEXO A – RESPOSTAS DAS INSTITUIÇÕES IBERO-AMERICANAS	211

1 INTRODUÇÃO

A internet e a interface *web* permitem a criação de vários instrumentos de comunicação síncronos, assíncronos ou multidirecionais. Esses instrumentos impulsionam a crescente produção de conteúdos, que precisam ser reunidos e armazenados de forma eficiente e dinâmica. Isso é o reflexo da necessidade atual da sociedade, que faz uso massivo de celulares e computadores, marcando presença diária nas redes sociais, comprando e vendendo por meio de *sites* de negócios, locomovendo-se por aplicativos de transporte e consumindo informações por meio de *websites*.

A rede mundial de computadores transformou-se em *input* da sociedade, proporcionando o acesso e compartilhamento das informações de forma globalizada pela *web*. Isso pode ser comprovado pelos números crescentes encontrados no *website Internet Live Stats* (<https://www.internetlivestats.com/>), com estatísticas em tempo real da quantidade de usuários da internet, de *websites* criados, de *e-mails* enviados por dia, de pesquisas diárias no Google e de usuários ativos em redes sociais (INTERNET LIVE STATS, [c2021]). Sendo assim, a internet torna-se uma ferramenta essencial nas áreas do conhecimento acadêmico-científico, nas vidas profissional e pessoal. Além disso, observa-se o uso dessa ferramenta nas organizações como instrumento estratégico na execução das atividades e comunicação com a sociedade, como é o caso dos *websites* institucionais. O cenário promovido pela pandemia demonstra na prática a importância da *internet* e da *web* no cotidiano, com o *home office* e o ensino remoto, e até na área da saúde, com a telemedicina tornando-se realidade.

O crescimento da geração de dados digitais provém principalmente da informatização dos processos no âmbito governamental e privado a partir da década de 2000 (LOTT; CIANCONI, 2018, p. 123).

Toda essa informação é gerada nato-digitalmente¹ e, assim como no suporte em papel, precisa ser tratada, guardada e preservada. No entanto, a preservação é um dos grandes desafios do século XXI – e, com o aumento da produção de informação em formato digital, tem sido questionada cada vez mais a importância de garantir sua disponibilização por longos períodos (MÁRDERO ARELLANO, 2004, p. 15-16).

Conforme Gomes, Miranda e Costa (2011), a “*Web* é extremamente efêmera”, ou seja, caracteristicamente transitória ou passageira. Além disso, os autores destacam:

¹ Documento nato-digital – documento criado originariamente em meio eletrônico (BRASIL, 2015a).

A maioria de suas informações se torna indisponível, sendo perdida para sempre após um curto período de tempo. Observou-se que 80% das páginas são atualizadas ou desaparecem após um ano. Mesmo as publicações impressas sofrem os efeitos da transitoriedade dos dados da web, porque frequentemente citam recursos on-line que se tornaram indisponíveis. Além de perder importantes informações históricas, a transitoriedade das informações publicadas na web faz com que pessoas comuns percam suas memórias como indivíduos (por exemplo, fotos compartilhadas exclusivamente pela web). Links quebrados também degradam o desempenho de aplicativos e serviços populares da Web, como favoritos compartilhados, mecanismos de pesquisa ou redes sociais, levando seus usuários à insatisfação. (GOMES; MIRANDA; COSTA, 2011, p. 408, tradução nossa)².

Em outra pesquisa, é apontado que, além dos 80% de páginas *web* não disponíveis, o número de artigos acadêmicos indisponíveis gira em torno de 13% após 27 meses. Nas redes sociais, por exemplo, no Twitter, perdem-se 11% no período de um ano (COSTA; GOMES; SILVA, 2016, p. [1]). Os autores³ salientam a importância de iniciativas de preservação para combater a dita efemeridade da *web*, mitigando o risco de perda de informações e do contexto histórico, para que seja assegurada a acessibilidade das páginas *web* para as gerações futuras (GOMES; MIRANDA; COSTA, 2011, p. 408, tradução nossa). À medida que as tecnologias *web* evoluem, as técnicas estabelecidas de arquivamento e preservação da *web* enfrentarão desafios⁴ (XIE *et al.*, 2013, p. 1, tradução nossa). Como já mencionado, há problemas em recuperar conteúdos ou até mesmo trabalhos acadêmicos, podendo ocorrer o chamado erro 404 (*error 404 – not found* ou página não encontrada), pois a *web* não foi pensada em sua criação em termos de preservação digital, assim como grande parte das tecnologias (ROCKEMBACH, 2021a).

Toda essa revolução tecnológica trouxe bônus e ônus, visto que a ação de preservar passou do suporte em papel para um documento digital; portanto, com contextos de produções diversos, o ofício de preservar em meio digital é repleto de reveses.

O valor social da *web* torna importante “reflexões sobre os documentos nato-digitais e sua preservação, sendo necessário garantir que as gerações futuras irão ter acesso a um registro igualmente significativo e relevante da produção informativa contemporânea”

² Most of its information becomes unavailable and is lost forever after a short period of time. It was observed that 80% of the pages are updated or disappear after 1 year. Even printed publications suffer from the effects of web data transience because they frequently cite online resources that became unavailable. Besides losing important scientific and historical information, the transience of the information published on the web causes common people to lose their memories as individuals (e.g. photos shared exclusively through the web). Broken links also degrade the performance of popular web applications and services, such as shared bookmarks, search engines or social networks, leading their users to dissatisfaction (GOMES; MIRANDA; COSTA, 2011, p. 408).

³ The web needs preservation initiatives to fight ephemerality. It must be ensured that the information besides being accessible World Wide Web, prevails across time to transmit knowledge for future generations. Web archives are innovative systems that acquire, store and preserve information published on the web (GOMES; MIRANDA; COSTA, 2011, p. 408).

⁴ As web technologies evolve, established web archiving techniques face challenges (XIE *et al.*, 2013, p. 1).

(PINTO, 2011, p. 12; FERREIRA, M., 2006, p. 11). Preservar a *web* se faz necessário, pois parte da nossa memória social agora nasce no ambiente digital. Também é difícil não a imaginar em nosso cotidiano, já que os sistemas globais de governança, comunicação, ciência e cultura estão completamente entrelaçados e dependentes da *web* (SUMMERS, 2018; MELO, 2020).

Ademais, realizar a preservação da *web* ou seu arquivamento é um desafio ainda mais complexo no escopo da preservação digital, que envolve diferentes tecnologias, linguagens, formatos, integridade das páginas, acessibilidade e usabilidade em navegadores, além das questões éticas e legais. No artigo “Arquivamento da Web no contexto das humanidades digitais”, Rockembach (2019) salienta que:

Pensar em como preservar a web ou realizar o seu arquivamento é uma área de pesquisa que envolve as tecnologias utilizadas para criar páginas web e a variedade de conteúdos disponibilizados, diferentes plataformas, modalidades de software, linguagens e formatos utilizados e como garantir a reprodução dessa informação retrospectivamente. O papel das humanidades dentro dessas pesquisas é fundamental, pois complementam a compreensão dos fenômenos contemporâneos, muitos dos quais acontecem mediados tecnologicamente. (ROCKEMBACH, 2019, p. 132).

Para preservar a *web* como um artefato cultural, é necessário utilizar diferentes padrões estruturais, pois ela apresenta um estilo dinâmico de organização das informações (MASANÈS, 2006, p. 18).

Ciente dos desafios da preservação no ambiente digital, a Organização das Nações Unidas para a Educação, a Ciência e a Cultura (Unesco), em 2003, publicou a *Carta sobre a Preservação do Patrimônio Digital*, na qual ressalta a preocupação com a criação, a manutenção, a preservação e o acesso ao patrimônio digital às futuras gerações. Entre os objetos digitais citados na referida carta, encontram-se as páginas *web*, compondo o patrimônio digital.

A Unesco, em 2015 reafirmou a preocupação com os desafios no âmbito digital publicando a *Recomendação sobre a preservação e o acesso ao patrimônio documental, inclusive em formato digital*⁵. No documento, é apontado que a preservação do patrimônio digital deve considerar a rápida evolução da tecnologia e os desafios de estabelecer modelos e processos de preservação de objetos complexos, como páginas *web*.

As páginas *web* são compreendidas como recurso acessível utilizando o Protocolo de Transferência de Hipertexto (*Hypertext Transfer Protocol* – HTTP) e reproduzido por um

⁵ Recommendation concerning the preservation of, and access to, documentary heritage including in digital form (UNESCO, 2016b).

navegador, podendo ser estáticas, com conteúdo fixo, ou dinâmicas, com conteúdo variável, derivado de uma base de dados. Os *websites* são coleções de páginas *web* relacionadas por conteúdo ou domínio, com caráter dinâmico (SOCIETY OF AMERICAN ARCHIVISTS, c2021b).

A partir da minha vivência profissional como bibliotecária e arquivista, designada para compor o “Grupo de Estudos sobre Preservação Digital no âmbito da Biblioteca Nacional do Brasil⁶”, participando do Grupo de Pesquisa Estudos e Práticas de Preservação Digital⁷, conhecido como Dríade – após a adesão da BNDigital à Rede Brasileira de Serviço e Preservação Digital – Rede Cariniana, com o “Curso de Repositórios Digitais em Bibliotecas IBEREX 2017” realizado na Espanha, corroborando o curso oferecido pelo Arquivo Nacional dos Países Baixos “Winter School: Managing Archives – Challenges and Chances for the 21st.-2019⁸” realizado em Haia, surgiu a ideia de estudar a preservação das páginas *web*, com indagações no campo profissional e pessoal, pois o volume informacional gerado nas duas áreas é exponencial. Atualmente, como membro da Comissão Permanente de Preservação Digital da Fundação Biblioteca Nacional (FBN)⁹, o tema também é uma preocupação institucional, mencionado no escopo da política de preservação digital da FBN. Além disso, há escassez de documentos e estudos que abordem a temática do arquivamento e preservação da *web* no Brasil.

Na Biblioteca Nacional da Espanha – BNE, tomei ciência do tema preservação da *web* do domínio (.es) garantido por legislação naquele país. No Brasil, ainda não temos nenhuma instituição realizando o trabalho de preservação de páginas *web*, apenas pesquisas desenvolvidas pelo Núcleo de Pesquisa em Arquivamento da Web e Preservação Digital (Nuaweb), criado em 2017, e pelo Grupo de Pesquisa Dríade – Rede Cariniana – Subgrupo Arquivamento da Web, originado em 2020. Além disso, foi aprovado na Câmara dos Deputados o parecer do Projeto de Lei nº 2.431/2015 que “dispõe sobre o patrimônio público digital institucional inserido na rede mundial de computadores e dá outras providências”

⁶ FUNDAÇÃO BIBLIOTECA NACIONAL. *Portaria FBN/PRESI nº 34 de 27 de agosto de 2019*. Disponível em: https://sei.bn.gov.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador 0027239 e o código CRC F152785B. Acesso em: 23 ago. 2021.

⁷ CONSELHO NACIONAL DE DESENVOLVIMENTO CIENTÍFICO E TECNOLÓGICO – CNPq. [Grupo de pesquisa] Estudos e práticas de preservação digital. [S. l.], c2021. Disponível em: dgp.cnpq.br/dgp/espelhogrupo/3997875180380796. Acesso em: 15 maio 2021.

⁸ FUNDAÇÃO BIBLIOTECA NACIONAL. Servidoras da BN participam de curso sobre atuais desafios do gerenciamento de arquivos. *BN Notícias*, dez. 2019. Disponível em: <https://www.bn.gov.br/acontece/noticias/2019/12/servidoras-bn-participam-curso-sobre-atuais-desafios>. Acesso em: 23 ago. 2021.

⁹ FUNDAÇÃO BIBLIOTECA NACIONAL. *Portaria FBN nº 8, de 8 de fevereiro de 2021*. Disponível em: <https://www.in.gov.br/web/dou/-/portaria-fbn-n-8-de-8-de-fevereiro-de-2021-304722636>. Acesso em: 23 ago. 2021.

(BRASIL, 2015b, p. 1), que, se aprovado, tornará as ações de preservação do ambiente *web* obrigatórias no território brasileiro. Como é uma área incipiente no contexto brasileiro, acredita-se que a preservação desse tipo de informação seja relevante para as pesquisas futuras das ciências humanas e sociais.

Além disso, justifica-se pesquisar a temática por causa de todo tipo de riscos, como ataques *hackers*/ciberataques, conflitos e guerras. Prova disso é que, em 2022, cerca de 1.300 profissionais do patrimônio cultural, bibliotecários, arquivistas, pesquisadores e programadores estão trabalhando como voluntários para identificar e arquivar *websites* em risco de instituições de patrimônio cultural ucraniano enquanto o país está sob ataque, por meio do *Saving Ukrainian Cultural Heritage Online (SUCHO)*¹⁰. Até o momento, foram salvos mais de 30 TB de documentos digitalizados, obras de arte e outros materiais digitais, em mais de 3.500 *websites* de museus, bibliotecas e arquivos ucranianos (DOMBROWSKI; KIJAS; MAJSTOROVIC, 2022; VERMA, 2022).

Espera-se, com a presente pesquisa, trazer reflexões teóricas e práticas sobre a importância da preservação do patrimônio digital, isto é, das páginas *web*, para a sociedade brasileira e para as instituições de guarda de memória, além de estimular que estas identifiquem esse tipo de material como parte de seus acervos.

Dessa forma, a presente pesquisa teve como tema preservação do patrimônio digital, especificamente, a preservação e o arquivamento das páginas *web*, com o propósito de responder ao seguinte questionamento: onde e como estão sendo preservadas as páginas *web*?

1.1 OBJETIVOS

A pesquisa tem como objetivos:

1.1.1 Objetivo geral

Estudar a preservação do patrimônio digital por instituições de guarda da memória das páginas *web*, que fazem parte do *International Internet Preservation Consortium – IIPC*, com base na *Carta sobre a Preservação do Patrimônio Digital* da Unesco.

¹⁰ <https://www.sucho.org/>

1.1.1 Objetivos específicos

- A. Conceituar e caracterizar a preservação do patrimônio digital, arquivamento da *web*, preservação digital e preservação do patrimônio arquivístico digital;
- B. Identificar e caracterizar as instituições de guarda da memória fundadoras do IIPC;
- C. Identificar as políticas, as ações e os critérios de preservação digital utilizados pelas instituições ibero-americanas do IIPC;
- D. Identificar os problemas e as barreiras à preservação das páginas *web*.

Esta dissertação está dividida nas seguintes seções, além desta introdução:

- 2. *Metodologia* – a pesquisa foi realizada em duas fases: uma parte teórica e uma parte empírica, apresentando conceituação e autores importantes da área;
- 3. *Evolução da Internet e da web* – introduz a temática do ambiente digital e apresenta o surgimento da internet e da *web*, bem como a forma que ambas foram inseridas ao longo dos anos no cotidiano da sociedade, tornando-se instrumentos importantes de comunicação;
- 4. *Arquivamento da web* – apresentação dos conceitos de arquivamento da *web* e arquivos da *web*, salientando a importância da preservação das páginas *web* para as futuras gerações;
- 5. *Patrimônio digital* – aborda a importância da *Carta sobre a Preservação do Patrimônio Digital* e outros documentos produzidos pela Unesco no cenário da preservação, incentivando políticas institucionais no âmbito da preservação. Tal documento incluiu as páginas *web* como parte do patrimônio digital, tornando-se, então, elemento-chave para o tema;
- 6. *Preservação do patrimônio* – introduz o conceito de preservação, ainda no ambiente analógico, e mostra as semelhanças com a preservação digital, pois ambos os conceitos são parte integrante das missões das instituições de memória e patrimônio;
- 7. *Etapas para arquivamento e preservação da web* – apresenta o modelo de referência *Open Archival Information System* (OAIS), princípios *Findable, Accessible, Interoperable and Reusable* (FAIR) e *Transparency, Responsibility, User Focus and Technology* (TRUST), correlacionando-os com padrões de metadados e arquivamento, introduz o IIPC;
- 8. *Análise e discussão dos resultados da parte empírica*
- 8.1. *Caracterização das Ações de Guarda e Preservação nas instituições membros fundadoras do IIPC* – trata-se da parte empírica da pesquisa, que identifica e caracteriza as instituições membros fundadoras do IIPC, a saber: *Bibliothèque nationale de France* (BnF), *Internet Archive*, *The British Library*, *Royal Danish Library*, *National Library of Finland*,

National Library of Sweden, National and University Library of Iceland, Library and Archives Canada, Library of Congress, National Library of Norway, National Library of Australia e National Library of Italy;

8.2. Políticas, Critérios e Ações de Preservação Digital das Instituições Ibero-Americanas do IIPC;

9. Considerações finais – apresentação das conclusões da pesquisa.

2 METODOLOGIA

Em termos metodológicos, trata-se de uma pesquisa exploratória e descritiva; com procedimentos técnicos baseados em fontes bibliográficas e documentais (primárias, secundárias e terciárias)¹¹, como artigos, livros, teses e dissertações, normas, relatórios, bibliografias, bases de dados, páginas *web*, vídeos, eventos e seminários.

As pesquisas de caráter exploratório “têm como principal finalidade desenvolver, esclarecer e modificar conceitos e ideias”, e aquelas de cunho descritivo “têm como objetivo primordial a descrição das características de determinada população ou fenômeno ou o estabelecimento de relações entre variáveis” (GIL, 1989, p. 44-45). O autor afirma que a pesquisa bibliográfica “utiliza-se fundamentalmente das contribuições dos diversos autores sobre determinado assunto” e que a pesquisa documental necessita de “fontes que não receberam tratamento analítico” (GIL, 1989, p. 73).

Para atingir o objetivo específico A – conceituar e caracterizar a preservação do patrimônio digital, arquivamento da *web*, preservação digital e preservação do patrimônio arquivístico digital –, foi feita uma revisão de literatura por meio de pesquisa no Portal de Periódicos da Capes. Foram selecionadas bases de dados da área de Ciência da Informação, como *Wiley Online Library* (com textos completos), *Scopus* (referencias com resumos), *LISTA Ebsco* (referenciais com resumos e textos completos), *Science Direct* (textos completos), *Web of Science* (referenciais com resumos), *SAGE Journals* (textos completos), *Springer books* (livros), *Brapi* (textos completos) e *Google Scholar*. O tema escolhido tem maior desenvolvimento na literatura estrangeira, abrangendo livros, artigos de periódicos, artigos de jornais, normas, relatórios e revistas especializadas e *sites* de organismo internacionais. Portanto, na estratégia de busca, foram utilizados termos em inglês (“*web archiving*” AND “*web archives*” AND “*digital preservation*” AND “*digital heritage*”) e em português (“arquivamento da web” AND “arquivos web” AND “preservação digital” AND “patrimônio digital”).

Embora tenha sido recuperado um grande número de documentos com a combinação de palavras-chave, muitos documentos eram de acesso restrito (mediante pagamento/compra) e não foi possível acessar o texto completo, sendo apenas recuperados conteúdos em português nas bases Brapi e Springer. Optou-se por recuperar os documentos de acesso

¹¹ Fontes primárias – são aquelas que contêm informações originais ou pelo menos novas interpretações de fatos ou ideias; Fontes secundárias – têm a função de facilitar o uso do conhecimento disperso nas fontes primárias, ou seja, apresenta informações filtradas e analisadas; Fontes terciárias – têm a função de guiar o pesquisador para as fontes primárias e secundárias (SANTOS, G., *et al.*, 2005).

aberto. A seguir, o quadro 1 sintetiza os termos e os números de textos recuperados durante essa busca inicial para fundamentar a parte teórica:

Quadro 1 – Síntese e quantificação de termos de busca

Bases	Termos	Números
SCOPUS	("web archiving" AND "web archives" AND "digital preservation" AND "digital heritage")	21
ScienceDirect	web archiving "web archives" "digital preservation"	43
	("web archiving" AND "web archives" AND "digital preservation" AND "digital heritage")	6
SpringLink	("arquivamento da web" AND "arquivos web" AND "preservação digital" AND "patrimônio digital")	1
	("web archiving" AND "web archives" AND "digital preservation" AND "digital heritage")	226
Wiley Online Library	"web archiving AND web archives AND digital preservation"	300
	"web archiving AND web archives AND digital preservation AND Digital heritage"	31
Web of Science	ALL=(web archiving* and web archives* and digital preservation* and digital heritage)	5
	ALL=(web archiving* and web archives* and digital preservation*)	57
SAGE	("web archiving" AND "web archives" AND "digital preservation" AND "digital heritage")	5
LISTA EBSCO	("web archiving" AND "web archives" AND "digital preservation" AND "digital heritage")	8
Google	web archiving "web archives" digital preservation	844
	("arquivamento da web" AND "arquivos web" AND "preservação digital" AND "patrimônio digital")	1
	("web archiving" AND "web archives" AND "digital preservation" AND "digital heritage")	207
BRAPCI	ARQUIVAMENTO DA WEB* PRESERVACAO DIGITAL**	9
	ARQUIVAMENTO DA WEB AND ARQUIVO WEB AND PRESERVACAO DIGITAL AND PATRIMONIO DIGITAL**	1
Total		1765

Fonte: Elaborado pela autora.

Ainda que o quadro 1 mostre um número relevante de publicações, foram priorizados alguns textos e referências bases das bibliografias disponibilizadas pelo IIPC, Rede Cariniana, instituições ligadas à preservação digital e eventos relacionados com o tema.

A partir da fase exploratória de busca por conteúdos relevantes, construiu-se a parte teórica, que engloba as seções 3 a 7, sendo o ponto de partida a seção 3, *Evolução da Internet e da web*, na qual foi introduzida a temática do ambiente digital e apresentado o surgimento da internet e da *web*. A seção 4, *Arquivamento da web*, visou assimilar, por intermédio da subseção 4.1, os processos e desafios do arquivamento da *web*, e conhecer, por meio da subseção 4.2, atores da pesquisa sobre arquivamento na *web*, no cenário internacional e nacional. Na seção 5, *Patrimônio digital*, foi abordada a importância da *Carta sobre a Preservação do Patrimônio Digital* e outros documentos produzidos pela Unesco no cenário

da preservação, que foi subdividido nas subseções 4.1, *Patrimônio arquivístico digital*, e 4.2 *Patrimônio digital nas bibliotecas nacionais*. Tais subdivisões serviram de alicerce para o desenvolvimento das seguintes, bem como para melhor conceituar e compreender os aspectos do arquivamento da *web* e preservação digital. Por intermédio da seção 6, *Preservação do patrimônio*, buscou-se entender as diferentes conceituações e abordagens da preservação digital; já a seção 7, *Etapas para arquivamento e preservação da web*, teve como foco os aspectos técnicos e de governança.

Noronha e Ferreira (2000) relatam que o pesquisador faz uso da revisão de literatura na busca por conhecimentos e atualização sobre o tema em fontes variadas, e que as “revisões podem ser classificadas conforme seu propósito, abrangência, função e tipo de análise desenvolvida”. Para a presente pesquisa, o propósito foi de base – “ser de apoio científico”, servindo de respaldo teórico às questões levantadas; a abrangência foi temática – pelo recorte na preservação de páginas *web*; a função de atualização –, “permitindo a identificação de informações para o desenvolvimento corrente do conhecimento”; e quanto ao tipo de análise, foi bibliográfica e documental – “como subsídio para comparação das abordagens nos diferentes trabalhos e documentos, permitindo a seleção daqueles de maior interesse” (NORONHA; FERREIRA, 2000, p. 191-194).

O universo da parte empírica limitou-se às instituições que são membros fundadores e ibero-americanos do IIPC, devido às restrições de tempo. Para atender aos objetivos específicos B, C e D, a pesquisa teve como base a literatura produzida sobre a temática e os documentos disponibilizados pelas instituições pesquisadas.

Para cumprir o objetivo B – identificar e caracterizar as instituições de guarda da memória fundadoras do IIPC –, foi consultado o *website* do IIPC (<https://netpreserve.org/about-us/members>) e os *websites* institucionais.

Para atender ao objetivo C – identificar as políticas, as ações e os critérios de preservação digital utilizados pelas instituições ibero-americanas do IIPC –, foi consultado o *website* do IIPC e os *websites* das instituições. Além de ser analisado o conteúdo da documentação disponibilizada, enviaram-se *e-mails* solicitando informações que não foram localizadas nos *websites* das instituições ibero-americanas.

Para alcançar o último objetivo específico D – identificar os problemas e as barreiras à preservação das páginas *web* – foram realizadas reflexões sobre as referidas instituições de memória com base nos textos da pesquisa bibliográfica e documental. O quadro 2 resume as etapas da pesquisa:

Quadro 2 – Resumo das etapas de pesquisa

Questão da Pesquisa	Onde e como estão sendo preservadas as páginas <i>web</i> ?		
Objetivo geral	Estudar a preservação do patrimônio digital por instituições de guarda da memória das páginas <i>web</i> , que fazem parte do <i>International Internet Preservation Consortium – IIPC</i> , com base na <i>Carta sobre a Preservação do Patrimônio Digital</i> da Unesco		
Objetivos Específicos	Etapas	Resultados	Fundamentação Teórica
(A) Conceituar e caracterizar a preservação do patrimônio digital e alguns temas relacionados, como: arquivamento da <i>web</i> ; preservação digital; preservação do patrimônio arquivístico digital	Pesquisa bibliográfica e documental; construção das seções teóricas	Seções: 3. Evolução da Internet e da <i>web</i> ; 4. Arquivamento da <i>web</i> ; 5. Patrimônio digital; 6. Preservação do patrimônio e 7. Etapas para arquivamento e preservação da <i>web</i>	ALENCAR-BRAYNER (2016); BANOS <i>et al.</i> (2013); BROWN (2006); BRÜGGER (2005, 2011, 2018); GARRET (2019); GRÁCIO (2012); GOMES <i>et al.</i> (2011, 2014, 2016); IIPC (2016, 2020, 2021); ISO (2009, 2012, 2017, 2021), LOTT, CIANCONI (2018); MÁRDERO ARELLANO (2004, 2008); MASANÈS (2006); MELO (2020); OWENS (2018); PENNOCK (2013); ROCKEMBACH (2018, 2020, 2021); REYES AYALA (2013); UNESCO (1972, 1997, 2002, 2004, 2012, 2015, 2016, 2021), entre outros
(B) Identificar e caracterizar as instituições de guarda da memória fundadoras do IIPC	Consulta ao <i>site</i> do IIPC (https://netpreserve.org/about-us/members) e os <i>websites</i> institucionais	Seção 8: Análise e discussão dos resultados	Membros Fundadores do IIPC
(C) Identificar as políticas, as ações e os critérios de preservação digital utilizados pelas instituições ibero-americanas do IIPC	<i>Websites</i> institucionais e análise de conteúdos da documentação disponibilizada e envio de <i>e-mail</i> solicitando mais informações que não foram localizadas em seus <i>websites</i> para todas as instituições ibero-americanas	Seção 8: Análise e discussão dos Resultados	Arquivo.pt (Portuguese Web Archive); National Library of Spain; Biblioteca de Catalunya (PADICAT) e Chilean National Library; BARDIN (1977); GLUSHKO (2020); GRÁCIO (2012); OWENS (2018) PÉREZ; SERRA (2010) SCHISLER (2020)
(D) Identificar os problemas e as barreiras à preservação das páginas <i>web</i>	Foi desenvolvido por meio das reflexões das referidas instituições de memória com base nos textos da pesquisa bibliográfica e documental	Seções 8, 9	Membros Fundadores do IIPC; Arquivo.pt (Portuguese Web Archive); National Library of Spain; Biblioteca de Catalunya (PADICAT) e Chilean National Library, IIPC; NUNES (2021); MASANÈS; MAJOR; GOMES (2021); RAMOS JUNIOR, CIANCONI (2019)

Fonte: Elaborado pela autora.

3 EVOLUÇÃO DA INTERNET E DA WEB

A “*Internet* é um sistema global de redes de computadores interligados que utilizam um conjunto de protocolos (*Internet Protocol Suite* ou TCP/IP) com o propósito de conectar usuários ao redor do mundo” (INTERNET, [2021]). O embrião da Internet pode ser encontrado na Arpanet, uma rede de computadores criada pela *Advanced Research Projects Agency* (ARPA) em 1969, formada por pesquisadores e pelo Departamento de Defesa dos Estados Unidos da América – EUA (RIBEIRO, 1998; GOETHALS; AGUIAR; ALMEIDA, 2000; CASTELLS, 2003).

Durante a década de 1970, a Arpanet começou a interagir com outras redes, “isto porque no ano de 1972, tinha-se iniciado na ARPA a investigação do conceito ‘*internetworking*’ forma de interligação de redes” (GOETHALS; AGUIAR; ALMEIDA, 2000, p. 5). Robert Kahn e Vint Cerf foram os criadores do protocolo padronizado de comunicação TCP/IP, que viabilizou e tornou-se a rede mundial com a versão IPv4, desenvolvida na década de 1970. Atualmente, continua sendo o protocolo padrão em uso da internet, mas, desde 2004, a versão utilizada é IPv6, suportada pela maioria dos equipamentos (LEINER *et al.*, 1997; LIVING INTERNET, [2021]).

Em 1980, essa interligação de redes já era realidade, misturando redes militares, científicas e universitárias (RIBEIRO, 1998; GOETHALS; AGUIAR; ALMEIDA; 2000). “A Arpanet tornou-se ARPA-INTERNET, e ficou dedicada à pesquisa. Em fevereiro de 1990, a Arpanet, já tecnologicamente obsoleta, foi retirada de operação libertando a *internet* de seu ambiente militar” (CASTELLS, 2003, p. 17).

A internet então começou a ser comercializada e distribuída por provedores de serviços que estabeleceram suas próprias redes e portas de comunicação. Para Castells:

A partir de então, a Internet cresceu rapidamente como uma rede global de redes de computadores. O que tornou isso possível foi o projeto original da Arpanet, baseado numa arquitetura em múltiplas camadas, descentralizada, e protocolos de comunicação abertos. Nessas condições a Net pôde se expandir pela adição de novos nós e a reconfiguração infinita da rede para acomodar necessidades de comunicação. (CASTELLS, 2003, p. 18).

No entanto, o que permitiu à internet conectar o mundo com uma interface amigável foi a *World Wide Web* (WWW), primeiro navegador (*browser*) da *web*, posteriormente chamado de Nexus (SLAVIN, 2016; BOULTON, 2014). Criada por Tim Berners-Lee em 1991, com objetivo de troca e compartilhamento de informações entre os cientistas da

Organização Europeia para Pesquisa Nuclear, conhecida como CERN (*Conseil Européen pour la Recherche Nucléaire*), apesar de não ser um projeto oficial da instituição, rapidamente se tornou um meio de publicação e comunicação para o formato digital (WORLD WIDE WEB FOUNDATION, [c2021]).

Três tecnologias são fundamentais e permanecem como base da *web* de hoje, encontradas em partes do navegador. A *HyperText Markup Language* (HTML) – a linguagem de marcação de hipertexto serve para a formatação da *web*. O *Uniform Resource Identifier* (URI), identificador uniforme de recursos, é uma espécie de “endereço” único, utilizado para identificar cada recurso na *web*, sendo comumente chamado de *Uniform Resource Locator* (URL). O *Hypertext Transfer Protocol* (HTTP) – permite a recuperação de recursos vinculados por hipertexto, em toda a *web* (WORLD WIDE WEB FOUNDATION, c2021, tradução nossa)¹².

O mesmo Bernes-Lee fundou, na década de 1990, o *World Wide Web Consortium* (W3C), com a “missão de levar a WWW ao seu potencial máximo, desenvolvendo protocolos e diretrizes que garantam seu crescimento a longo prazo” (W3C ESCRITÓRIO BRASIL, 2021). Nesse consórcio internacional, com organizações filiadas, uma equipe, em tempo integral, e o público trabalham juntos para desenvolver padrões para a *web*. O W3C tem como princípios a “*Web* para todos: onde o valor social da *web* está nas novas possibilidades de comunicação humana, comércio e compartilhamento de conhecimentos” e “*Web* em todas as coisas: significa o número de diferentes tipos de dispositivos que podem acessar”. Um dos principais objetivos do W3C é tornar esses “benefícios disponíveis para todas as pessoas, independente do *hardware* que utilizam, *software*, infraestrutura de rede, idioma, cultura, localização geográfica ou capacidade física e mental” (W3C ESCRITÓRIO BRASIL, 2021).

Em 1992, foi criado o *ViolaWWW*. Em 1993, Marc Andreessen elaborou o *Mosaic*, que deu origem ao *Netscape*, lançado em 1994, que se tornou o mais conhecido e usado na época (*VIOLAWWW*, [2021]).

Posteriormente, outras empresas começaram a adquirir licenças para adaptar o código do *Mosaic* e, a partir de 1995, criaram seus próprios *browsers*, dando início a uma disputa que se estende até os dias de hoje, sendo um dos mais conhecidos o *Internet Explorer* (atual *Edge*), da *Microsoft*. Em 1995, a *Netscape* decidiu tornar seu código aberto e criou a organização

¹² Tim had written the three fundamental technologies that remain the foundation of today’s web (and which you may have seen appear on parts of your web browser): HTML: HyperText Markup Language. The markup (formatting) language for the web; URI: Uniform Resource Identifier. A kind of “address” that is unique and used to identify to each resource on the web. It is also commonly called a URL; HTTP: Hypertext Transfer Protocol. Allows for the retrieval of linked resources from across the web (WORLD WIDE WEB FOUNDATION, c2021).

sem fins lucrativos Mozilla, que elaborou e lançou o Firefox em 2002; e a Google lançou o Google Chrome em 2008 (GUERRA..., [2021]; MOZILLA.ORG, 2021).

Ainda que a internet tenha seu início na demanda dos cientistas da computação, com a criação do Protocolo TCP/IP na década de 1970, a partir dos anos de 1990, a internet começa a se popularizar, com a expansão da *web* para a sociedade em geral (CASTELLS, 2003). Lott e Cianconi (2018) apontam que a popularização dos computadores pessoais e a ampliação da área de cobertura dos serviços de acesso à internet foram precursores do surgimento da *web* social ou *Web 2.0*, fase em que a *web* passou a ser, de fato, colaborativa, “permitindo maior interação dos usuários entre si”, podendo ser considerada como “o embrião das redes sociais como Twitter, Instagram, YouTube e Facebook, entre outras” (LOTT; CIANCONI, 2018, p. 123-124).

Com a perspectiva de promover e democratizar a internet, Vint Cerf e Robert Kahn, criadores do protocolo TCP/IP, fundaram, em 1992, a *Internet Society*. Ela configura-se como uma organização global sem fins lucrativos, que capacita as pessoas para manter a internet aberta, globalmente conectada, segura e confiável. A *Internet Society* tem como missão promover o desenvolvimento aberto da internet para todas as pessoas em todo o mundo, além desempenhar um papel importante de informar sobre a história da internet (LEINER *et al.*, 1997).

A WWW desempenha um papel de veículo disseminador de informações, em que as organizações fazem sua comunicação oficial e institucional. Esse espaço também passou a ser de construção da memória institucional que se refere “ao conjunto de atributos, histórias, momentos, relações externas e trajetórias realizadas pelas instituições”, atualmente migrado para o ambiente digital (MARCIAL; VIEIRA, 2021, p. 151).

A velocidade com que a web se tornou parte da vida cotidiana é sem precedentes: foram necessárias pouco mais de duas décadas para que a web crescesse de um serviço relativamente pequeno utilizado principalmente por cientistas, para um meio de informação global. É agora não só um centro de comunicações, mas também um registo único da vida no século XXI. No entanto, a própria velocidade a que mesma se desenvolve constitui uma ameaça à nossa memória cultural digital, ao seu legado técnico, à sua própria evolução e à nossa história social. (PENNOCK, 2013, p. 3, tradução nossa).

Desse modo, torna-se importante realizar estudos para melhor compreensão e estímulo a respeito do arquivamento e da preservação de páginas *web*, para que as futuras gerações tenham acesso aos conteúdos já apagados ou com diferentes versões e, dessa maneira, tentem

resgatar os valores primários¹³ (prova) e secundários¹⁴ desses registros documentais, memória que pode ser considerada parte do patrimônio digital, seja ele institucional, nacional ou mundial, sendo conceitos importantes no campo arquivístico.

¹³ Valor primário – valor atribuído a documento em função do interesse que possa ter para a entidade produtora, levando-se em conta a sua utilidade para fins administrativos, legais e fiscais (ARQUIVO NACIONAL, 2005, p. 171).

¹⁴ Valor Secundário – valor atribuído a um documento em função do interesse que possa ter para a entidade produtora e outros usuários, tendo em vista a sua utilidade para fins diferentes daqueles para os quais foi originalmente produzido (ARQUIVO NACIONAL, 2005, p. 172).

4 ARQUIVAMENTO DA WEB

Como uma das principais razões para o arquivamento de páginas *web*¹⁵, destacam-se a transitoriedade e a efemeridade dos *websites*¹⁶. Apesar de ser um espaço ubíquo, a *web* tem sido alvo de preocupações da comunidade do patrimônio cultural, pois os conteúdos podem ser modificados ou apagados em um prazo relativamente curto. Pennock afirma que

[...] vários estudos analisaram a duração média de vida de páginas *web*, com resultados que vão de 44 dias a 75 ou 100 (Kahle, 1998; Lawrence et al., 2001; Weiss, 2003). O conteúdo específico de uma página pode desaparecer ainda mais frequentemente, particularmente em páginas e sítios de redes sociais. Em alguns casos, o conteúdo pode ser inacessível 'arquivado' ou relocado quando os sítios são redesenhados¹⁷. (PENNOCK, 2013, p. 3, tradução nossa).

A autora ressalta ainda que os *websites* “podem simplesmente desaparecer por *links* quebrados, resultando em erros 404 de página não encontrada” e faz uma comparação com o livro perdido dentro de uma biblioteca (PENNOCK, 2013, p. 3, tradução nossa). O documento *Collecting and preserving the World Wide Web*, publicado em 2003, já enfatizava “a necessidade de iniciativas de arquivamento da *Web* para ajudar a preservar o valor informativo, cultural e probatório da *World Wide Web* ou de alguns subconjuntos específicos da mesma” (DAY, 2003b, p. 1). Ou seja, o enfoque do arquivamento poderia ser seletivo, já que a preservação em longo prazo requer investimento contínuo.

O IIPC considera arquivamento da *web* como “o processo de recolha de porções da *World Wide Web*, preservando as coleções num formato de arquivo, e depois oferecendo para acesso e utilização” (INTERNATIONAL INTERNET PRESERVATION CONSORTIUM, 2021a, tradução nossa).

Niu (2012) define “o arquivamento da *web* como um processo de reunir dados que foram registrados na *World Wide Web*, armazenando-os, garantindo que os dados sejam preservados em um arquivo e disponibilizando os dados coletados para pesquisas futuras” (NIU, 2012, n.p., tradução nossa).

¹⁵ Web page – A resource accessed using the Hypertext Transfer Protocol (HTTP) and rendered by a browser for display or printing (SAA, c2021c). Disponível em: <https://dictionary.archivists.org/entry/web-page.html>

¹⁶ Websites – A collection of web pages related by content or domain (SAA, c2021e). Disponível em: <https://dictionary.archivists.org/entry/website.html>.

¹⁷ Various studies have looked into the average lifespan of web pages, with results ranging from 44 days to 75 or 100 (KAHLE, 1998; LAWRENCE *et al.*, 2001; WEISS, 2003). Specific content on a page can disappear even more frequently, particularly on newsdriven and social media websites. In some cases, content can be inaccessibly ‘archived’ or relocated when sites are re-designed (PENNOCK, 2013, p. 3).

Nesta pesquisa, considera-se que *web archive*, ou arquivo da *web*, é aquele que “contém materiais publicados na *web*, pelos quais uma organização aceitou a responsabilidade em longo prazo, tanto pela preservação como pelo acesso” (MURRAY; HSIEH, 2008, p. 5, tradução nossa).

A *Library of Congress* usa como termo arquivo *web* não somente “para descrever toda a coleção de arquivos *web*”, mas também como “um grupo de *URLs* iniciais arquivadas e descritas pela biblioteca, representando uma organização ou pessoa. Um arquivo da *web* pode estar associado a uma ou mais coleções, e pode ter um ou mais *URLs* a ele associados” (LIBRARY OF CONGRESS, [2021a]).

O documento *Web at Risk: preserving our nation's cultural heritage*, produzido por Murray e Hsieh (2006), descreve que um arquivo *web*, cujo conteúdo é composto por materiais publicados na *web*, compartilha semelhanças com um arquivo tradicional. Infere-se que, na designação ‘arquivo’, estão as responsabilidades da instituição de guarda pela preservação dos conteúdos e, conseqüentemente, a integridade ao longo do tempo, além da disponibilização do acesso em consonância com a legislação. Tal como os arquivos tradicionais, o conteúdo dos arquivos da *web* inclui uma gama de materiais, desde objetos digitais simples, como documentos textuais, até objetos digitais complexos, com *links* de relacionamento com outras páginas *web* (MURRAY; HSIEH, 2006).

O livro *The Archived Web: Doing History in the Digital Age* (BRÜGGER, 2018) questiona, no capítulo “*Is Web Archiving?*”, o fato de o arquivo da *web* ser realmente um arquivo. O autor tece que “o termo arquivo da *web* tem sido utilizado para descrever qualquer recolha da *web*, e conseqüentemente, para representar o ato de recolher e preservar a *web* e dar acesso”. Todavia, Brügger fala que é necessário refletir sobre a terminologia utilizada, pois as instituições de patrimônio possuem diferenças significantes na forma de recolha ou aquisição dos documentos.

Talvez o termo arquivo não seja a melhor definição para diferenciar as instituições de patrimônio cultural. Terrada e Fonseca (2019, p. 386) ponderam “que hoje, na literatura arquivística, a palavra arquivo é polissêmica, ou seja, tem vários significados” e apresenta algumas acepções pelo *Dicionário Brasileiro de Terminologia Arquivística* do Arquivo Nacional do Brasil (ARQUIVO NACIONAL, 2005, p. 27 e 96) e por Paes (2004, p. 20):

- Conjunto de documentos produzidos e acumulados por uma entidade coletiva, pública ou privada, pessoa ou família, no desempenho de suas atividades, independentemente da natureza do suporte;
- Fundo que é um conjunto de documentos de uma mesma proveniência;

- Instituição ou serviço que tem por finalidade a custódia, processamento técnico, conservação e o acesso a documentos;
- Instalações onde funcionam arquivos, [isto é, o local propriamente dito];
- Móvel destinado à guarda de documentos.
- Outros significados atribuídos à palavra são: arquivos de computador ou mídias digitais e em títulos de periódicos, geralmente no plural, devido às influências inglesa e francesa.

Pode-se observar que a palavra arquivo serve tanto para documentos como para instituições ou arquivos digitais. Dessa forma, o arquivamento da *web* pode ser entendido no sentido amplo. Além disso, Belloto (2014) esclarece que

A existência dos arquivos na sociedade justifica-se pela necessidade que sempre tiveram as comunidades humanas, desde a mais remota antiguidade, de registrar, em suportes inteligíveis, as suas normas, ações, transações, direitos, deveres etc. de modo a preservar os testemunhos necessários ao andamento das relações entre governantes e governados, tanto quanto dos membros dessa mesma sociedade entre si. (BELLOTO, 2014, p. 132).

O livro *O documento arquivístico ante a realidade digital*, publicado por Rondinelli (2013), produzido a partir de sua tese de doutorado em 2011, apresenta conceitos importantes, como o de documento, informação, documento arquivístico e documento arquivístico digital. A autora ressalta convergências e divergências entre os termos ao longo do texto, de forma sistematizada em diversas épocas, com base nos autores da ciência da informação e da arquivologia. Apreende-se o quão desafiadoras são a compreensão e a conceptualização dos referidos termos, quando o trabalho arquivístico é transportado para o ambiente digital.

Por outro lado, Brügger (2018, p. 78) diz que “se a terminologia for ser consistente com a distinção comumente utilizada entre bibliotecas e arquivos, aquilo a que agora se chama de um arquivo *web*, deve chamar-se uma *webrary* (*web+library*)”. Então, Brügger (2018) constata que a terminologia “arquivos *web*” tem sido amplamente utilizada pelas instituições de patrimônio “com coleções nacionais e transnacionais como exemplo do *Internet Archive* e, que por essas razões, esta terminologia um tanto confusa e enganosa provavelmente deve ser aceita”. No livro *The Archived Web: Doing History in the Digital Age* (BRÜGGER, 2018), o autor utiliza-se do termo arquivo da *web* para indicar a atividade de recolha, preservação e disponibilização da *web*, independentemente de qual parte da *web* está em questão (público ou privado); e o arquivo *web* é utilizado para os resultados dessa atividade. Para Brügger, espera-se que os arquivos da *web* sejam arquivos no sentido tradicional da palavra, é preciso ter clara a terminologia (BRÜGGER, 2018, p. 78).

Ademais das questões terminológicas do arquivo da *web* ou arquivamento da *web*, no artigo “*Archiving Web-published materials: A needs assessment of librarians, researchers, and content providers*”, Murray e Hsieh (2006) observaram as questões e os desafios enfrentados por participantes de uma pesquisa subsidiada pela *Library of Congress*, etapa do projeto *Web-at-risk*¹⁸ que faz parte do “Programa Nacional de Infraestrutura e Preservação da Informação Digital”. O objetivo era levantar as necessidades de arquivamento na *web* e a preservação. Assim, os autores ponderaram a necessidade de colaboração entre bibliotecas e entidades governamentais para realizar a preservação.

Nessa conjuntura, Brügger (2011, p. 25) reforça que as soluções para efetuar o arquivamento da *web* devem prever a dinamicidade dos conteúdos para guarda e preservação, ou seja, pensar nas características efêmeras, diferentes dos procedimentos realizados com o formato analógico. Afirma que “o armazenamento casual da *web*, não pode ser considerado arquivamento e o ato de preservar deve permear um grau de reflexão”,

[...] sobre a razão pela qual o arquivamento é realizado – seja por diversão; com vista a preservar a antiga página da família; a fim de documentar como uma determinada página da web era vista num determinado momento; com um projeto de investigação em mente; ou com vista a preservar o patrimônio cultural de uma nação ou de uma cultura. (BRÜGGER, 2011, p. 26, tradução nossa¹⁹).

É apontado que, “assim como os espaços habituais de preservação da memória, os arquivos da *web* também se constituem como fonte para pesquisas e podem se consolidar como espaços fundamentais para a salvaguarda de informações de uma época” (MELO, 2020, p. 32).

4.1 PROCESSOS E DESAFIOS DO ARQUIVAMENTO

De acordo com Vlassenroot e colaboradores (2019), o arquivamento da *web* “nasce entre os anos 1990 do movimento da preservação digital, comandado por instituições de memória, com objetivo de capturar e preservar registros” e as memórias que a sociedade produzia naquele momento (VLASSENROOT *et al.*, 2019, tradução nossa).

¹⁸ Disponível em: https://digitalpreservation.gov/partners/documents/webatrisk_cpg_final_31may2006.pdf.

¹⁹ And in order to be designated as web archiving, the casual storing of web material is not sufficient. In addition, the act of preserving must be accompanied by reflection on why the archiving is carried out – be that for fun; with a view to preserving the old family home page; in order to document how a certain webpage looked at a certain point in time; with a research project in mind; or with a view to preserving the cultural heritage of a nation or a culture (BRÜGGER, 2011, p. 26).

As questões técnicas são apontadas como desafios do arquivamento da *web*. Segundo Pennock (2013, p. 4, tradução nossa), “capturar²⁰ e arquivar um *site* individual pode ser relativamente simples, isto é baixá-lo no servidor e armazená-lo de maneira *off-line*. No entanto, o arquivamento da *web* em larga escala e a longo prazo é um assunto mais complicado”. As principais abordagens de arquivamento da *web* são por depósito legal, comumente em biblioteca nacionais, por colheita automática, por seleção ou de forma combinada (DAY, 2003a, p. 3).

No livro *Archiving Websites: general considerations and strategies*, Brügger (2005) menciona dois tipos de abordagem de arquivamento, o micro e o macro. O “*microarchiving* é efetuado em pequena escala tanto no que diz respeito ao espaço (um número limitado de *websites*) como tempo (um período limitado e isolado)”, e pode ser realizado “por indivíduos que não têm à sua disposição um computador consideravelmente potente e capacidade de armazenamento, cujos conhecimentos técnicos de arquivamento ou de o tratamento subsequente ou é inexistente ou a nível amador”, muitas vezes, com “base numa necessidade imediata, aqui e agora, de preservar um objeto de estudo” (BRÜGGER, 2005, p. 10).

Já o “*macroarchiving* é realizado em grande escala, de um grande número de *websites* e, em princípio, sem definição de tempo e tamanho de forma contínua”. Pode ser feito “por instituições (públicas/privadas, nacionais/internacionais) que têm consideráveis potência tecnológica, capacidade de armazenamento, e conhecimentos técnicos profissionais à sua disposição”, com intuito de “arquivar (parte do) patrimônio cultural (inter) nacional” (BRÜGGER, 2005, p. 11), ou seja, o patrimônio produzido em meio digital.

Hockx-Yu (2011, p. [2]) frisa que, “independentemente da abordagem existe um conjunto de processos essenciais ao arquivamento da *web* que precisam ser executados e geridos, para assegurar a adequação a qualquer sistema de arquivo da *web*”. Os processos que envolvem o arquivamento da *web* são a seleção, o *harvesting*, o armazenamento, o acesso e a preservação digital.

Segundo o autor, a **seleção** é o processo de decisão que determina quais os sítios *web* serão arquivados e quais incluir como parte de uma coleção de arquivos da *web*. Ainda, podem ser adicionados metadados descritivos para descrever o conteúdo selecionado, e para as instituições que não têm o mandato legal para recolher sítios *web*, o pedido de autorização também faz parte do fluxo de trabalho (HOCKX-YU, 2011).

²⁰ Captura/Coleta/Rastreamento/Recolha/*Crawlers*: são termos sinônimos para presente pesquisa, utilizados em arquivamento da *web*.

Já quanto ao *harvesting*, ou recolha, Hockx-Yu (2011) afirma que se refere ao processo automatizado de descarregar as cópias de sítios *web* selecionados, geralmente utilizando *software* de rastreamento (*crawlers*) da *web*. Começa, geralmente, a partir de uma lista de URLs (sementes), visitando-os e descarregando-os, antes de identificar todas as hiperligações dentro das páginas visitadas. O autor ressalta que a verificação da garantia de qualidade, manual ou automatizada, faz, frequentemente, parte do processo de recolha.

O **armazenamento** consiste no processo de retenção de sítios *web* arquivados de forma segura e confiável. Os formatos de arquivo comumente utilizados para sítios *web* são ARC e WARC. Ambos os formatos foram desenvolvidos especificamente para arquivos da *web* (HOCKX-YU, 2011; ISO, 2009, 2017). Conforme a norma “ISO 28500: 2017 *Information and documentation – WARC file format*”:

O formato WARC oferece uma forma padrão de estruturar, gerir e armazenar os milhares de milhões de recursos recolhidos a partir da *web* e de outros locais. É utilizado para construir aplicações para a colheita, gestão, acesso, mineração e troca de conteúdos. Embora represente o formato padrão único dos arquivos *web*, tem sido adotado para além da comunidade de arquivamento da *web* para armazenar materiais digitais nascidos ou digitalizados. (ISO, 2017, p. 1, tradução nossa).

O formato “WARC é uma extensão do formato do arquivo ARC que tem sido tradicionalmente utilizado para armazenar ‘*crawls*’ como sequências de blocos de conteúdo colhidos na *World Wide Web*”. Na norma ISO 28500:2017, é mencionado que “além do conteúdo primário registrado nos ARC, o formato estendido WARC acomoda o conteúdo secundário”, tais como “metadados atribuídos”, por exemplo (ISO, 2017).

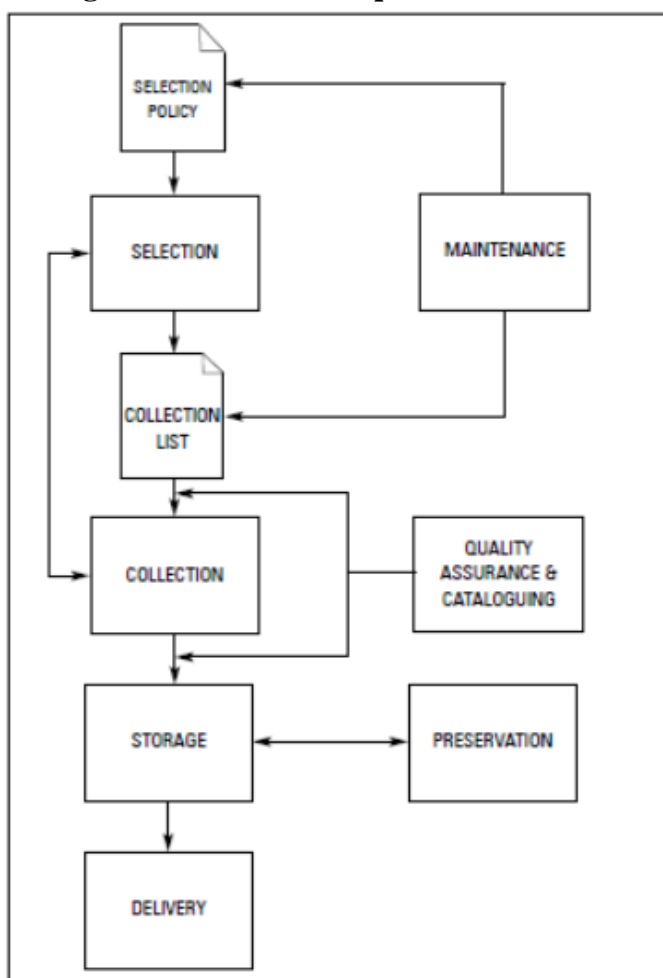
O **acesso** refere-se “à reprodução e disponibilização do acesso, aos sítios *web* arquivados para os usuários nos arquivos da *web*”. Já a **preservação digital** “está relacionada com as normas, melhores práticas e tecnologias, que em conjunto são necessárias para assegurar o acesso aos arquivos da *web* ao longo do tempo” (HOCKX-YU, 2011, p. [3]).

Os processos apresentados por Hockx-Yu (2011) explicitam a complexidade operacional do arquivamento da *web*, e Masanès (2006) corrobora as questões técnicas do arquivamento afirmando que:

[...] os arquivos da *web* necessitam primeiro separar o conteúdo da publicação permanente de seu criador original e, depois em segundo lugar, garantir que o conteúdo possa ser resistente ao fracasso e evolução da *web* atual. O primeiro requer a cópia e o arquivamento do conteúdo em uma infraestrutura separada. E o último envolve a preservação ativa do conteúdo da *web* para remover a dependência dos vários componentes do sistema (protocolos, formatos digitais, aplicativos, etc.) e sua inevitável obsolescência tecnológica. (MASANÈS, 2006, p. 15, tradução nossa).

Com intuito de ilustrar o fluxo de arquivamento, apresenta-se a figura do “processo de *web archiving*” retirada do livro *Archiving websites – a practical guide for information management professional* (BROWN, 2006, p. 6), com os nove elementos do arquivamento da *web*: política de seleção, seleção, lista de recolha, coleção, manutenção, garantida da qualidade e catalogação, armazenamento, acesso e preservação, conforme figura 1.

Figura 1 – Fluxo de Arquivamento da *web*



Fonte: Brown (2006, p. 6).

A figura 1 mostra o processo de arquivamento da *web*, no qual os recursos precisam ser selecionados, recolhidos, preservados e entregues aos usuários. Tal figura foi traduzida no artigo “Arquivamento da web: estudos de caso internacionais e o caso brasileiro” por Rockembach (2018). A política de seleção configura-se como uma tomada de decisão importante, ou seja, é o início do processo com a definição do escopo do arquivo, assim

“como o contexto, o método e os critérios de seleção” (BROWN, 2006, p. 27). Masanès destaca alguns benefícios da construção de uma política de seleção:

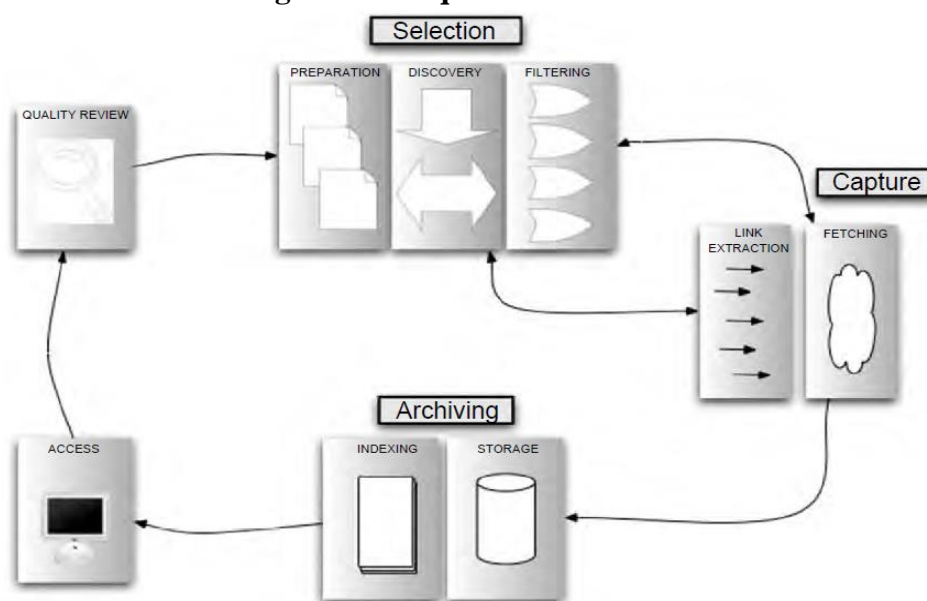
[...] evitará preconceitos pessoais ou alterações relativas a uma conjuntura específica, proporcionando assim continuidade, definindo prioridades e permitindo o planejamento, posicionando as coleções num contexto de arquivo mais vasto para facilitar a cooperação. (MASANÈS, 2006, p. 72).

A política de seleção será essencial para todo o processo de arquivamento da *web* – “uma vez que a política de seleção foi estabelecida e aprovada, resultará na seleção de recursos específicos da *web* para recolha”. O produto final será uma “lista de itens a serem recolhidos baseados no processo de seleção, que fornecerá a base para o recolhimento dos recursos digitais, contendo os critérios de seleção do conteúdo, a extensão, o tempo e frequência apropriados para coleção” (BROWN, 2006, p. 28-33). A manutenção representada na figura 1 envolve um ciclo contínuo de seleção e recolha, apoiado na política e na lista de recolha que contém a definição de limites, frequência e tempo, para que a coleção obtenha qualidade de arquivamento e preservação. “Claramente a política de seleção quando definida e atualizada é a pedra angular de qualquer programa de arquivamento da *web*, e constitui-se num pré-requisito essencial para a construção de uma coleção significativa” (BROWN, 2006, p. 40, tradução nossa).

Lisiane Ferreira (2019) afirma que a “seleção compreende três etapas consecutivas: preparação, descoberta e filtragem”. O objetivo da fase de preparação “é definir a meta de coleta, a política de captura e as ferramentas para implementá-la”. A fase de descoberta tem como objetivo “determinar a lista de pontos de entrada que serão usados para a captura, bem como a frequência e o escopo dessa captura”. Já a fase de filtragem “pode ser feita manualmente ou automaticamente, sendo que a manual é necessária quando os critérios usados para a seleção necessitam de interpretação humana” (FERREIRA, L., 2019, p. 29; MASANÈS, 2006, p. 71).

Para Masanès, o processo de arquivamento da *web* ocorre de maneira cíclica de seleção (preparação, descoberta e filtragem), com captura (busca e extração de *links*), arquivamento (armazenamento e indexação), além do acesso e revisão da qualidade, conforme na figura 2:

Figura 2 – Arquivamento da *web*



Fonte: Masanès (2006, p. 71).

Percebe-se, na figura 2, a ligação do processo de filtragem com a busca, enquanto o de descoberta está ligado ao de extração de *links*, promovendo um ciclo entre a seleção e a captura.

Para Banos e colaboradores (2013), o processo de arquivamento da *web* inclui: a identificação (*identification*), a avaliação (*appraisal*) e a seleção (*selection*), somados à aquisição (*acquisition*), à inclusão (*ingest*), à organização (*organization*), ao armazenamento (*storage*), à descrição (*description*) e ao acesso (*access*). O autor não menciona a ordem dos processos ou se acontecem de forma simultânea.

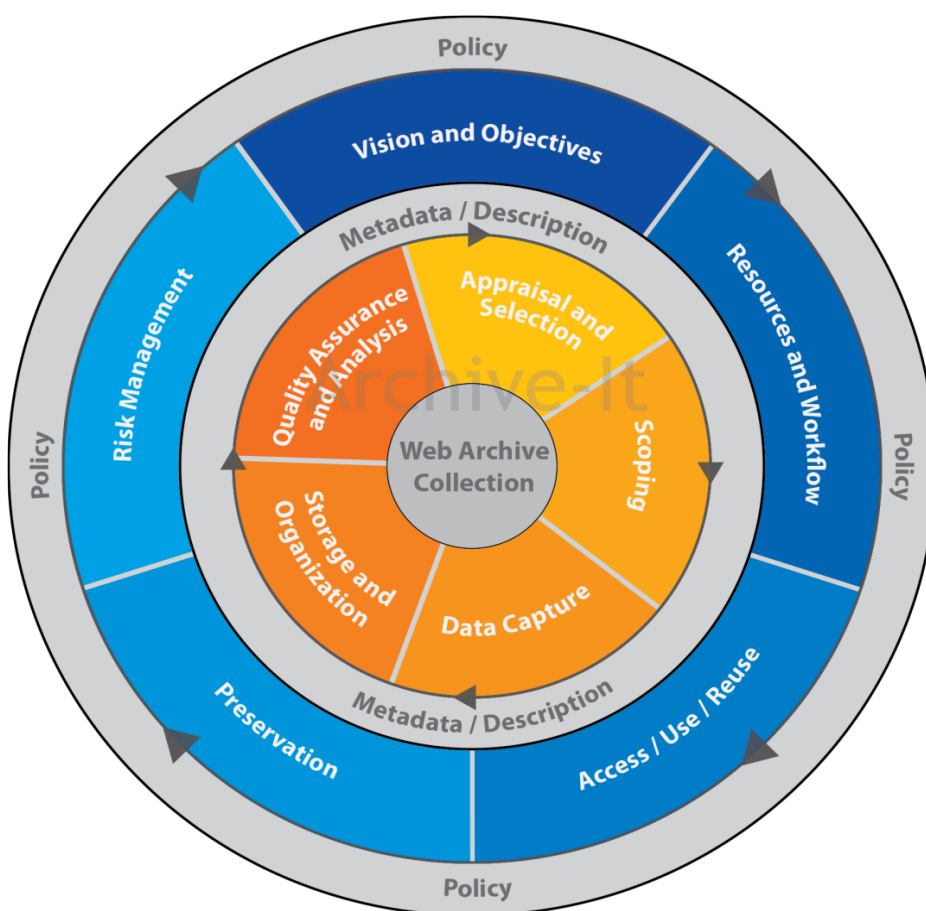
Conforme Banos e colaboradores (2013), o arquivamento da *web* tem dois aspectos-chave: o organizacional e o técnico. O aspecto organizacional “envolve a entidade responsável pelo processo, a sua governança, o financiamento, a viabilidade em longo prazo e o pessoal responsável pelas tarefas de arquivamento”, e o aspecto técnico abrange “as tarefas operacionais do arquivamento do conteúdo da *web*” (BANOS *et al.*, 2013, p. 1).

No mesmo ano da publicação de Banos e colaboradores, a equipe do *Internet Archive*, especificamente ligada ao *Archive-It*, publicou *The Web Archiving Life Cycle Model* (WALCM), o modelo do ciclo de vida do arquivamento da *web*, tendo como autores principais Molly Bragg e Kristine Hanna. O documento foi produzido com base nas experiências da equipe e dos parceiros associados (BRAGG; HANNA, 2013).

Segundo as autoras, o “WALCM é uma tentativa de incorporar os braços tecnológicos e programáticos do arquivamento da *web* em uma estrutura que será relevante para qualquer

organização que pretenda arquivar *web*”. Ou seja, por meio de “representação dos fluxos de trabalho comuns às organizações, poderá servir de base para o desenvolvimento ou aprimoramento de programas de arquivamento *web*”. É salientado que “embora o modelo seja dividido em etapas individuais, cada ação não é solitária, pois foi construído em forma de círculo, desta maneira as etapas e fases estão relacionadas, com uma quantidade significativa de sobreposição entre elas” (BRAGG; HANNA, 2013, p. 1-2). A figura 3 ilustra o Modelo de ciclo de vida do arquivamento da *web*:

Figura 3 – Modelo de ciclo de vida do arquivamento da *web*



Fonte: Bragg e Hanna (2013, p. 3).

Conforme a figura 3, o WALCM apresenta-se dividido por áreas e cores; o nível mais externo do ciclo de vida, na cor cinza, é a área das decisões organizacionais, no qual são produzidas as políticas. Segundo as autoras, é um fator importante, “pois quase todos os aspectos do arquivamento da *web* envolvem algum tipo de decisão política” (BRAGG; HANNA, 2013, p. 2).

Em seguida, tem-se as etapas de construção da política. Referem-se às decisões fundamentais, compostas por: visões e objetivos (*Vision and Objectives*) institucionais no cerne do arquivamento da *web*; recursos e fluxos de trabalho (*Resources and Workflow*), etapa que visa refletir sobre os recursos disponíveis para execução do trabalho, incluindo orçamento e equipe; acesso / uso / reutilização (*Access / Use / Reuse*), que fornece acesso e mecanismos de monitoramento dos usuários; preservação (*Preservation*), que é a parte tomada de decisão sobre como preservar os conteúdos recolhidos, incluindo arquivos de dados e metadados; por último, gerenciamento de risco (*Risk Management*), etapa na qual se reflete sobre o risco de um programa de arquivamento da *web*, assim como sobre as questões de direitos autorais e acesso (BRAGG; HANNA, 2013, p. 4). As autoras afirmam que “englobar as etapas do ciclo de vida com uma área dedicada à política, era com objetivo de representar visualmente a formulação de políticas” (BRAGG; HANNA, 2013, p. 2).

Uma segunda área do círculo, também cinza, “representa os metadados e a descrição”. Escolheu-se “incorporar os metadados como uma área ao invés de um segmento da roda para enfatizar que criar, importar e exportar metadados é um processo contínuo que ocorre em conjunto com uma série de outras atividades no ciclo de vida” do arquivamento (BRAGG; HANNA, 2013, p. 2-3).

O círculo interno, laranja, descreve as tarefas diárias envolvidas no processo de arquivamento, ou seja, “diz respeito às tarefas operacionais de arquivamento da *web*” (ROCKEMBACH, 2018). Inicia-se com a “avaliação e seleção” (*Appraisal and Selection*), refere-se à decisão da escolha dos *sites* para recolhimento; seguida da definição do “escopo” (*Scoping*), parte que está relacionada com a extensão e a profundidade de recolha; intimamente ligadas à “captura dos dados” (*Data Capture*), isto é, a frequência e o tempo da recolha. Depois vem a etapa de “armazenamento e organização” (*Storage and Organization*), que “inclui um plano de armazenamento temporário ou de longo prazo para os dados arquivados, e em algumas instituições, a fase de armazenamento e organização do ciclo de vida também pode constituir suas atividades de preservação”. O ciclo laranja é finalizado com a etapa “garantia e análise de qualidade” (*Quality Assurance and Analysis*), na qual “ocorre a revisão das coleções arquivadas e verificação do cumprimento das metas pré-estabelecidas” (BRAGG; HANNA, 2013, p. 4-5).

No centro do ciclo de vida do arquivamento da *web*, cinza, encontra-se a coleção em si, o conteúdo da *web* arquivado. Vale ressaltar que o formato adotado pelos arquivos da *web* desenvolveu-se a partir do trabalho do *Archive-It*. A norma ISO 28500:2009 – *Information*

and documentation — WARC file format, atualizada em 2017(E), especifica o formato *Web ARChive*, ou WARC, como um padrão de arquivo a ser utilizado (ISO, 2009, 2017). Para os autores, capturar e preservar as coleções de dados está no centro de todas as atividades de arquivamento da *web* e, portanto, é o centro do modelo (BRAGG; HANNA, 2013; ROCKEMBACH, 2018).

Melo (2020, p. 112) criou um mapa mental para arquivamento de *websites* governamentais contendo seis etapas: 1 – o escopo do que deseja arquivar; 2 – conhecer a estrutura do *website*; 3 – preparar o arquivamento que inclui a frequência de coleta, metadados e política de preservação digital; 4 – o arquivamento com a captura/recolha dos conteúdos; 5 – garantia de qualidade, parte que abrange as conformidades com os padrões; e 6 – o acesso, uso e reuso dos conteúdos. O autor baseou-se no modelo de ciclo de vida do *Arquivamento da Web* (BRAGG; HANNA, 2013) e da *Abordagem sistemática para a preservação da web* (KHAN; RAHMAN, 2019), esta última voltada para preservação da *web*, que, segundo a tradução de Melo (2020, p. 72) compreende-se em dez etapas: 1 – Definição do escopo do arquivo *web*; 2 – Compreensão da estrutura da *web*; 3 – Identificação dos recursos da *web*; 4 – Identificação da comunidade designada; 5 – Priorização dos recursos *web*; 6 – Capturar os recursos da *web*; 7 – Política de seleção; 8 – Identificação dos metadados; 9 – Formato de armazenamento; e 10 – Mecanismos de disseminação de informações.

Khan e Rahman (2019, p. 71) descrevem o processo de preservação da *web* de forma ampla, combinando os pontos fortes das diferentes técnicas observadas nos estudos para a preservação da *web*, em que foram identificadas diferentes abordagens e possíveis técnicas de implementação que podem ser adoptadas no arquivamento digital.

Compreende-se como processos essenciais do arquivamento da *web*: seleção, recolha/coleta, armazenamento e acesso (HOCKX-YU, 2011; BROWN, 2006; MASANÈS, 2006), com adição de outros elementos como preservação, manutenção e qualidade (BROWN, 2006; MASANÈS, 2006; BRAGG; HANNA, 2013), o que consiste em algumas acepções completamente distintas das adotadas por Banos e colaboradores (2013) e por Bragg e Hanna (2013). O quadro 3 condensa as percepções dos autores sobre o processo de arquivamento da *web*:

Quadro 3 – Concepção de autores sobre o processo de arquivamento da web

Processos de Arquivamento da web (<i>webarchiving process</i>)				
Brown, 2006	Masanès, 2006	Hockx Yu, 2011	Banos e colaboradores, 2013	Bragg e Hanna, 2013
Política de Seleção Seleção Lista a recolha Recolha Manutenção Armazenamento Garantia de qualidade Acesso Preservação	Seleção – preparação, descoberta e filtragem; Captura – busca e extração de <i>links</i> Arquivamento – armazenamento e indexação Acesso Revisão de Qualidade	Seleção Recolha (Harversting) Armazenamento Acesso Preservação Digital	Identificação Avaliação Seleção Aquisição Inclusão Organização Armazenamento Descrição Acesso	Política Preservação, Acesso/uso/reuso, Gerenciamento de Risco Metadados e Descrição Avaliação e Seleção Escopo Captura dos dados Armazenamento e Organização Garantia de qualidade

Fonte: Elaborado pela autora.

Nota-se que Brown (2006) e Bragg e Hanna (2013) apresentam o processo com mais detalhes que os outros autores, inserindo a política, a manutenção, a qualidade dos recursos digitais e a preservação como aspectos importantes do arquivamento da *web*. Dessa maneira, destacam-se os quatro métodos para recolha do conteúdo *web*, que Brown (2006) chamou de “*The Matrix of collection methods*”, conforme figura 4 abaixo:

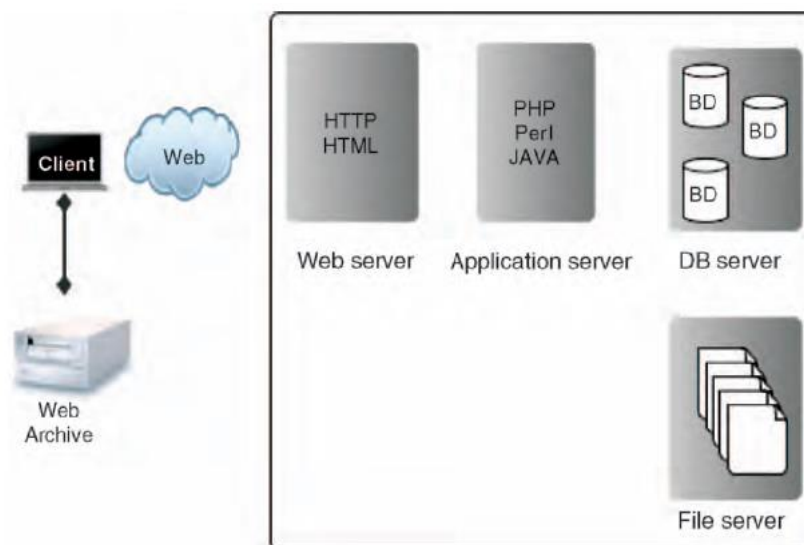
Figura 4 – Matriz do método de recolha

	Content-driven	Event-driven
Client-side	Remote harvesting	No method available
Server-side	Direct transfer Database archiving	Transactional archiving

Fonte: Brown (2006, p. 46).

A matriz do método de recolha pode ser por meio do “cliente” (*Client-side*) com orientação para conteúdo (*Content-driven*). Nesse caso, o tipo de recolha é por intermédio de “*harvesting* remoto que envolve o uso de *software* rastreador, robôs para coletar conteúdo de servidores remotos da *web*”. No caso “cliente” orientado para eventos (*Event-driven*), o autor não menciona o tipo de método na figura 4 (BROWN, 2006, p. 46-50). De acordo com Masanès (2006, p. 23), no “arquivamento – lado do cliente, a instituição está na posição de cliente para recolher conteúdo do servidor *web*. O servidor pode gerar conteúdo a partir de vários outros servidores (aplicação, base de dados, servidores de ficheiros/arquivos)”, como ilustrado na figura 5 abaixo:

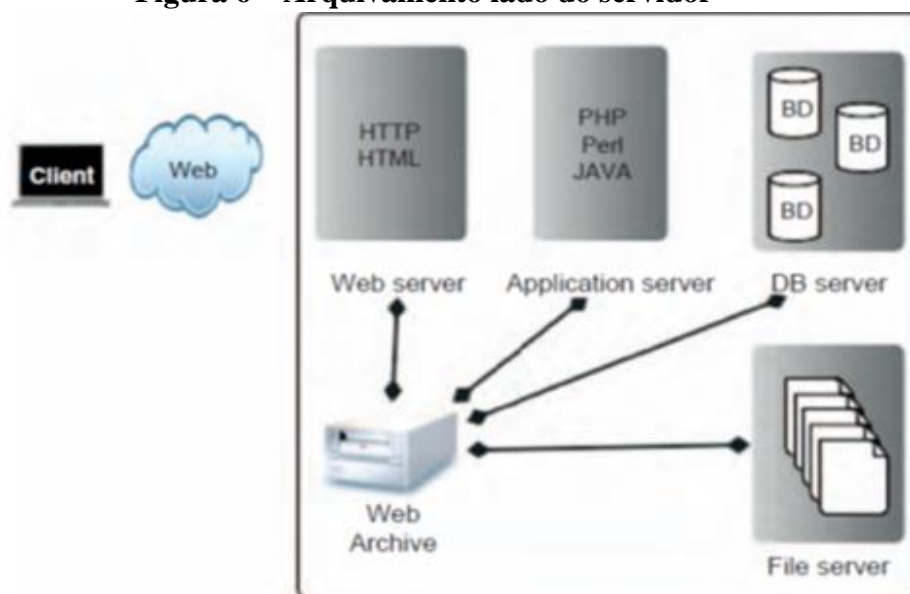
Figura 5 – Arquivamento lado do cliente



Fonte: Masanès (2006, p. 23).

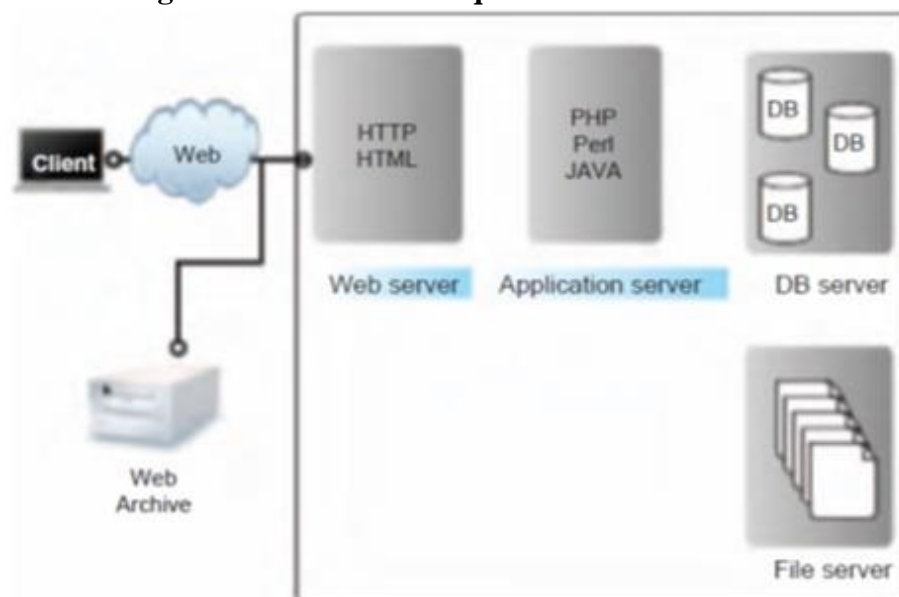
Já o método enfocando o “servidor” (*Server-side*), no âmbito do conteúdo (*Content-driven*), seria por meio de transferência direta. Usos assim são “indicados para *sites* estáticos ou [com] poucos recursos dinâmicos, [...] [sendo modalidade] mais simples de recolha de recursos *web*, na qual é solicitada uma cópia do conteúdo do *website* diretamente da fonte original, ou seja, com o criador.” (BROWN, 2006, p. 46), ou por meio de arquivamento do banco de dados que “envolve três etapas: primeiro o repositório define um modelo de dados padrão e o formato para arquivos bancos de dados; em seguida, cada banco de dados de origem é convertido para esse formato padrão; e, finalmente, uma interface de acesso padrão é fornecida aos bancos de dados arquivados” (BROWN, 2006, p. 59).

Masanès (2006, p. 27) ressalta que “no arquivamento do lado do servidor, diferentes grupos de informações são arquivadas diretamente dos servidores, gerando uma versão funcional do conteúdo arquivado e não apenas uma cópia de segurança dos ficheiros/arquivos, este é o desafio deste método”. Esse modelo está descrito conforme a figura 6 a seguir:

Figura 6 – Arquivamento lado do servidor

Fonte: Masanès (2006, p. 27).

No contexto do arquivamento “lado-servidor” (*Server-side*), aquele orientado para eventos (*Content-driven*) seria arquivo transacional, que “concentra-se em coletar as transações reais que ocorrem entre os servidores da *web* e o navegador, em vez do conteúdo que está disponível para ser entregue a partir do servidor da *web*” (BROWN, 2006, p. 61-62). Esse modelo está descrito conforme ilustrado pela figura 7:

Figura 7 – Recolha de arquivo transacional

Fonte: Masanès (2006, p. 26).

Brown (2006) afirma que o “arquivamento transacional por ocorrer no lado do servidor, requer a cooperação ativa do proprietário do site” (BROWN, 2006, p. 65).

O autor assevera que os criadores dos *websites* podem otimizar e facilitar o arquivamento seguindo alguns requisitos técnicos, tornando os *sites* mais amigáveis com arquivabilidade²¹ e permitindo o acesso aos arquivos pelos rastreadores. O artigo “Arquivabilidade de websites para preservação digital: estudo a partir da área da saúde” ressalta que se pode aplicar o método CLEAR (*Credible Live Evaluation of Archive Readiness*) que é “um conjunto de métricas para quantificar o nível de arquivabilidade de qualquer *site*” e a ferramenta *ArchiveReady* (<http://archiveready.com/>) *on-line* de avaliação de *websites*, ambos desenvolvidos por Banos e colaboradores (2013), para a melhoria do quesito arquivabilidade (BROWN, 2006; BANOS *et al.*, 2013; MELO; ROCKEMBACH, 2020).

No quesito URL e nomes de caminho, é necessário o uso de URLs persistentes e de nomes relativos, ao invés de absolutos, para *hyperlinks*. Deve-se manter a documentação da arquitetura do *website* atualizada, incluindo *hardware* e *software*. Além de todo conteúdo do *site*, este deverá conter metadados completos e precisos, como, por exemplo, utilizar-se dos elementos do padrão Dublin Core (BROWN, 2006, p. 66-67).

No artigo “*Archiving The World Wide Web*”, Lyman tece alguns problemas relativos ao tema, como: “problema cultural, pois parte da nossa herança cultural perde-se, aliada ao ritmo de mudanças tecnológicas”; “problemas técnicos de *hardware* e *software* necessários para visualização de documentos antigos; problemas econômicos (quem tem a responsabilidade de fazer o arquivamento e preservação da *web*?)” (LYMAN, 2002, p. 38-39). Lyman ainda afirma que

[...] não é a preservação que constitui uma ameaça econômica, é o acesso aos arquivos que possam prejudicar novos mercados. Encontrar um equilíbrio entre a preservação e o acesso é o problema mais urgente a ser resolvido, porque se a Web de hoje não for salva, não existirá no futuro. O acesso é um problema político, bem como um problema jurídico. A resposta ao problema de acesso, tal como as respostas a todos os problemas políticos, reside em estabelecer um processo de negociação entre as partes interessadas. Quem são as partes interessadas, e quais são as apostas, na construção de um arquivo Web? (LYMAN, 2002, p. 40, tradução nossa)²².

²¹ A arquivabilidade do *site* é definida como a extensão em que ele atende às condições para a transferência segura de seu conteúdo para um arquivo da *web* para fins de preservação (MELO; ROCKEMBACH, 2020, p. 532).

²² And yet it is not preservation that poses an economic threat, it is access to archives that might damage new markets. Finding a balance between preservation and access is the most urgent problem to be solved, because if today's Web is not saved it will not exist in the future. Access is a political as well as a legal problem. The answer to the access problem, like the answers to all political problems, lies in establishing a process of negotiation among interested parties. Who are the stakeholders, and what are the stakes, in building a Web archive? (LYMAN, 2002, p. 40).

Além de apresentar os problemas do âmbito político, organizacional, ou do autor, Lyman (2002) cita os cinco problemas técnicos essenciais da preservação digital, baseando-se no artigo “*Digital Longevity*”, de Howard Besser (2000). São eles: os problemas de visualização (*The Viewing Problem*), de codificação (*The Scrambling Problem*), de inter-relação (*The Inter-relation Problem*), de custódia (*The Custodial Problem*) e de tradução (*The Translation Problem*)²³. Todos esses fatores são de extrema importância para ações de arquivamento da *web*.

No entanto, os *websites* podem ser capturados e arquivados, mas será a preservação digital que fará com que eles sejam acessíveis em longo prazo. Vlassenroot e colaboradores (2019) ainda destacam as implicações legais em torno do arquivamento da *web*, como:

1. Legislação sobre direitos autorais;
2. A proteção de dados pessoais com arquivamento da *web*, provavelmente implicará em processo de dados pessoais. Entretanto, é importante ter em mente que o Regulamento Geral de Proteção de Dados (RGPD) autoriza derrogações legais aos direitos dos sujeitos e dos dados, quando os dados pessoais são para fins históricos ou científicos e para fins de arquivamento no interesse público;
3. No âmbito legal, com a devida autenticidade e integridade dos conteúdos, os arquivos da *web* poderiam ser utilizados nos tribunais por razões probatórias;
4. A questão do conteúdo ilegal que viola as políticas públicas e seu potencial interesse para os pesquisadores devido à natureza automática das ferramentas de arquivamento da *web*;
5. Delimitando legalmente o escopo nacional de competência em um contexto de arquivamento da *web*, com fronteiras digitais pouco claras. De fato, em relação a potencial sobreposição entre a legislação, como a lei sobre depósito legal e sobre registros públicos, é importante ter critérios claros para determinar no país que instituição de patrimônio nacional será responsável pelo arquivamento de um determinado *website*;
6. Legislação relativa à reutilização de informações do setor público. (VLASSENROOT *et al.*, 2019, p. 3, tradução nossa²⁴).

²³ BESSER, H. Digital longevity. In: SITTS, M. (ed.). *Handbook for digital projects: a management tool for preservation and access*. Andover, MA: Northeast Document Conservation Center, 2000. Disponível em: <http://besser.tsoa.nyu.edu/howard/Papers/sfs-longevity.html>. Acesso em: 10 fev. 2021.

²⁴ 1. Copyright legislation.

2. Personal data protection as web archiving is likely to imply personal data processing. However, it is important to keep in mind that the General Data Protection Regulation (GDPR) authorises legal derogations from the rights of the data subjects when personal data are processed for historical or scientific purposes and for archiving purposes in the public interest.

3. The legal framework on authenticity and integrity of online content as web archives could be used before courts for probative reasons.

4. The issue of illegal contents violating public policy and their potential interest for researchers due to the automatic nature of web archiving tools.

5. Legally delimiting the national scope of competence in a web archiving context with unclear digital boundaries. Indeed, regarding potential overlap between legislation on legal deposit and on public records, it is important to have clear criteria to determine the country and the national heritage institution in charge of the archiving of a particular website.

6. Legislation concerning reuse of public sector information (VLASSENROOT *et al.*, 2019, p. 3).

Observa-se que o processo de arquivamento envolve diversos atores, como profissionais e instituições, assim como os aspectos técnicos, organizacionais e jurídicos. A tomada de decisão e o tipo de abordagem irão influenciar na salvaguarda do patrimônio digital em longo prazo, assim como na preservação e no acesso aos conteúdos *web*.

4.2 ATORES DA PESQUISA SOBRE ARQUIVAMENTO NA WEB

Essa perspectiva de salvaguardar os registros produzidos no ambiente digital como fontes de pesquisas de uma época tornou-se objeto de estudos para muitos pesquisadores e instituições.

4.2.1 Contexto Internacional

Entre 2001 e 2004, a BnF organizou o *International Web Archiving Workshops* (IWAU)²⁵ em conjunto com *European Conferences on Digital Libraries* (ECDL). O objetivo dos *workshops* era reunir investigadores, profissionais, estudantes de pós-graduação e desenvolvedores de tecnologia da informação com conhecimentos e interesse na construção de arquivos da *web*. Os *workshops* tinham o intuito de proporcionar um fórum de interação entre bibliotecários, arquivistas, pesquisadores acadêmicos e investigadores industriais interessados em estabelecer métodos eficazes e desenvolver soluções para o arquivo da *web* (IWAU, 2006). A última menção ao evento foi ao 10º IWAU²⁶, no ano de 2010, que aconteceu junto com *iPRES 2010 – International Conference on Digital Preservation*, em Viena, Áustria. Acredita-se que o citado *workshop* proporcionou a construção de uma comunidade de pesquisadores envolvidos com arquivamento e preservação da *web*.

Na mesma época, em 2004, foi criada a *European Archive Foundation*, organização sem fins lucrativos com objetivo de arquivar e preservar a *web*, com sede em Amsterdam. O primeiro diretor foi Julien Masanès (BnF), o mesmo precursor do IWAU e do IIPC. Entre os anos de 2010 e 2018, a *European Archive Foundation* passou a se chamar *Internet Memory Foundation*, com sede em Paris, “apoando ativamente a preservação da *Internet* como um novo meio de comunicação social” (INTERNET MEMORY FOUNDATION, 2015).

²⁵ INTERNATIONAL WEB ARCHIVING WORKSHOPS. [S. l.]: BnF, 2006. Disponível em: <http://bibnum.bnf.fr/ecdl/>. Acesso em: 31 maio 2021.

²⁶ INTERNATIONAL WEB ARCHIVING WORKSHOPS. IWAU 2010: the 10th Intl Web Archiving Workshop. [S. l.: s. n.], [2010]. Disponível em: <http://www.wikicfp.com/cfp/servlet/event.showcfp?eventid=9651>. Acesso em: 31 maio 2021.

A *Internet Memory Foundation* criou alguns projetos no cenário europeu entre 2008 e 2012, “como parte do desenvolvimento e promoção de tecnologias de arquivamento da *web* em larga escala”, os quais foram financiados pela Comissão Europeia por três anos. Um deles foi o projeto interno *Archivethe.Net*²⁷ – AtN, uma plataforma de arquivamento da *web*, para “qualquer instituição cultural disposta a empenhar-se em arquivar *web*, para preservar o patrimônio cultural”. O segundo, com foco em redes sociais e memória, foi o *Collect-All ARchives to COmmunity MEMories* – ARCOMEM, visando às “instituições de memória como arquivos, museus, e bibliotecas na era da *web* social, contribuiu para o desenvolvimento de tecnologias e coleções, para então explorar novas possibilidades de seleção e preservação automatizada de conteúdos”. Ocorreu também o *SCAlable Preservation Environments* – SCAPE, já pensando na preservação digital, com “desenvolvimento de serviços escaláveis para planejamento e execução de estratégias de preservação institucional, numa plataforma de fonte aberta que faz a gestão dos fluxos de trabalho de maneira semiautomatizada, para coleções heterogêneas em grande escala”, bem como o *Living Web Archives* – LiWA, que levou o “arquivamento da *web* para além da abordagem caracterizada por instantâneos estáticos, para uma abordagem dinâmica e de inter-relações do conteúdo da *web*” (INTERNET MEMORY FOUNDATION, 2014).

Além desses, a *Internet Memory Foundation* desenvolveu o projeto *Longitudinal Analytics of Web Archive data* – LAWA, “com objetivo de criar um banco de ensaio experimental baseado na *Internet*, para análise de dados em larga escala”, e o *LivingKnowledge* – LK, que “considerava a diversidade como um bem e tinha a pretensão de torná-la rastreável, compreensível e explorável, com o objetivo de melhorar a navegação e a pesquisa em grandes conjuntos de dados” (INTERNET MEMORY FOUNDATION, 2014). Todos os projetos contaram com a participação de diversas instituições de pesquisa e do patrimônio cultural.

Em 2010, o *Joint Information Systems Committee* (JISC), com sede no Reino Unido, constituiu-se como uma empresa sem fins lucrativos do setor de educação superior que oferece serviços e soluções digitais. Financiou dois relatórios ligados com pesquisa e arquivos da *web*, no ano de 2010: “*Researcher Engagement with Web Archives: State of the Art*” e “*Researcher Engagement with Web Archives: Challenges and Opportunities for Investment*”. O primeiro resumiu o estado da arte do arquivamento da *web*, em relação aos pesquisadores e às necessidades de investigação, “com uma série de recomendações relativas ao

²⁷ INTERNET MEMORY FOUNDATION. *Archivethe.net*. 2015. Disponível em: <https://web.archive.org/web/20150206044401/http://internetmemory.org/en/index.php/projects/atn>. Acesso em: 31 maio 2021.

desenvolvimento de capacidades adicionais para o arquivamento da *web* e para a investigação de arquivos da *web*. Estas recomendações estão agrupadas em três temáticas: construir uma comunidade, ferramentas, recursos e práticas” (DOUGHERTY *et al.*, 2010, p. 5). Já o segundo, “vai além do estado dos arquivos da *web*, naquele momento e dos usos que deles são feitos, para expandir e para apontar algumas das importantes oportunidades que existem para os organismos de financiamento. Traz consigo novas oportunidades de inovação na tecnologia e na prática do arquivamento” (THOMAS *et al.*, 2010, p. 5).

No contexto latino-americano, foi publicado pelo *Center for Research Libraries-Global Resources Network*, em fevereiro de 2019, o relatório produzido por Jeffrey Garrett, *An Evaluation of Web Archiving Programs in the US Relevant to International and Area Studies: The Example of Latin America and the Caribbean*, que avalia “os esforços em torno do arquivamento da *web* no Caribe e América Latina para o uso no futuro por pesquisadores” (GARRET, 2019, p. 5). Alguns aspectos apontam um problema na questão da historiografia da *web*: como os pesquisadores farão uso dessas informações se elas não forem preservadas? “A importância histórica, cultural e intelectual do arquivamento e da preservação da *web* tem sido amplamente reconhecida. Atualmente, países com uma elevada taxa de penetração da *Internet* estão criando iniciativas de arquivamento do conteúdo da *web*²⁸” (XIE *et al.*, 2013, p. 1, tradução nossa).

Conforme Gomes e Costa (2014), os arquivos da *web* “permitem aos pesquisadores uma gama de informações com perspectivas diferentes, além de possibilitar o fornecimento de contribuições valiosas, aplicando seu conhecimento de campo para selecionar e organizar informações da *web*”. Os autores salientam que os pesquisadores podem exercer um papel importante na preservação, pois têm o potencial de “identificar os *sites* com informações significantes a serem preservadas e sugerir-las para preservação” e, assim adiciona-se também uma função de curadoria aos pesquisadores (GOMES; COSTA, 2014, p. 110, tradução nossa).

Em 2012, foi criada uma comunidade de pesquisa do tema, a “*Research Infrastructure for the Study of Archived Web Materials – RESAW*²⁹, que tem como objetivo promover uma infraestrutura de pesquisa colaborativa europeia para o estudo de materiais arquivados da *web*”, coordenada por Niels Brügger (Aarhus University, NetLab e o Centro de Estudos da *Internet*). Em 2016, a comunidade RESAW serviu de base para candidatura no *Horizon*

²⁸ The historical, cultural, and intellectual importance of archiving the web has been widely recognized. Today, all countries with high Internet penetration rate have established high-profile archiving initiatives to crawl and archive the fast-disappearing web content for long-term use (XIE *et al.*, 2013, p. 1).

²⁹ <https://resaw.eu/>

2020³⁰, o maior programa de financiamento em pesquisa e inovação da União Europeia (UE), com quase € 80 bilhões disponíveis para pesquisa e desenvolvimento, entre os anos de 2014-2020. A comunidade de pesquisadores ainda realiza a *RESAW Conference*, evento que acontece a cada dois anos. A *Fourth RESAW Conference 2021*³¹ aconteceu entre os dias 17 e 18 de junho, com sede virtual na Universidade de Luxemburgo (RESAW, [2021]).

Outro evento sediado em Luxemburgo foi a *IIPC General Assembly and the Web Archiving Conference*, que ocorreu também, de forma virtual, entre 14 e 16 de junho de 2021, na *National Library of Luxembourg* (INTERNATIONAL INTERNET PRESERVATION CONSORTIUM, c2021c). A *Web Archiving Conference* contou com a participação do pesquisador brasileiro Moisés Rockembach, no dia 15 de junho, na mesa dedicada aos “Domínios Nacionais” (*National Domains*), expondo a pesquisa “*Brazilian elections web archive*”, desenvolvida no âmbito do Nuaweb/UFRGS (INTERNATIONAL INTERNET PRESERVATION CONSORTIUM, [c2021c]; ROCKEMBACH, 2021b; UFRGS, [c2021]).

A UNT Digital Library³² está preservando a coleção “*International Internet Preservation Consortium*”, com cerca de 360 itens, abarcando os anos 2000-2021.

Outra comunidade de pesquisadores é *Engaging with Web Archives – EWA*³³, que tem como objetivo “aumentar a conscientização para o uso de arquivos da *web* e dos seus conteúdos para pesquisa e educação em uma ampla gama de profissões e disciplinas, como nas artes, humanidades, ciências sociais, ciência política, estudos de mídia, ciência da informação, ciência da computação”, entre outras. Outrossim, busca “fomentar colaborações entre iniciativas de arquivamento da *web*, pesquisadores, educadores e profissionais de TI” (EWA, c2021). Promove o evento *EWA Virtual Conference*, que, em 2020, foi transmitido pelo canal YouTube (EWA, 2020). Outra conferência com o título *Engaging with Web Archives 4 Digital Humanities (#EWA4DH)* foi realizada em agosto de 2021.

A comunidade de pesquisa é crescente em torno dos arquivos da *web* e usos do patrimônio digital. O *CollEx-Persée*, uma espécie de consórcio entre instituições francesas, estabelece-se como “uma infraestrutura de investigação criada em 2017, no qual reúne 21

³⁰ Horizon 2020 is the biggest EU Research and Innovation programme ever with nearly €80 billion of funding available over 7 years (2014 to 2020) – in addition to the private investment that this money will attract. It promises more breakthroughs, discoveries and world-firsts by taking great ideas from the lab to the market. EUROPEAN COMMISSION. Horizon 2020. Disponível em: <https://ec.europa.eu/programmes/horizon2020/en/what-horizon-2020>. Acesso em: 20 mar. 2022.

³¹ <https://www.resaw2021.net/>

³² INTERNATIONAL INTERNET PRESERVATION CONSORTIUM. UNT Digital Library. University of North Texas Libraries. Disponível em: <https://digital.library.unt.edu/explore/partners/IIPC/>. Acesso em: 7 jan. 2022.

³³ (#EWAVirtual), 21-22 Sept. 2020, County Kildare, Ireland. Disponível em: <https://ewaconference.com/>. Acesso em: 31 maio 2021.

bibliotecas de investigação/universitárias, 4 operadores nacionais (Persée para digitalização; ABES para dados bibliográficos; CTLes para armazenamento e preservação de documentos; INIST para mineração de dados) e a Biblioteca Nacional da França (*Bibliothèque nationale de France – BnF*). A organização é financiada pelo Ministério Francês do Ensino Superior e Investigação e tem como objetivo facilitar o acesso ao patrimônio digital, salvaguardados pelas bibliotecas aos pesquisadores (COLLEX-PERSÉE, [c2021a]).

Em maio de 2021, foi lançado pelo *CollEx-Persée*, em evento virtual, o projeto ResPaDon – *Réseau de Partenaires pour l’analyse et l’exploration de données numériques* (Rede de Parceiros para Análise e Exploração de Dados Digitais) – “rede com intuito de desenvolver e diversificar os usos dos arquivos da *web* recolhidos e preservados pela *BnF*”, desenvolvido pela Universidade de Lille na França, tendo entre as parceiras a BnF. Além disso, reúne profissionais da informação científica e investigadores com experiências diversas e considera importante ter momentos de discussão, como seminários, conferências internacionais e capacitação, para aprofundamento das pesquisas. O ResPaDon inclui dois objetivos: o primeiro, a prototipagem de acesso remoto ao arquivo *web* da BnF, atualmente restrito ao local, nas bibliotecas da Universidade de Lille; o segundo, a experimentação colaborativa em torno da aplicação de métodos de exploração da *web* ao vivo para arquivos da *web*, ou seja, um trabalho de curadoria. O projeto possui o orçamento de € 180.450, com período de execução de abril de 2021 a abril de 2023 (COLLEX-PERSÉE, [c2021b]).

Ademais, a comunidade arquivística também atua em pesquisas sobre o arquivamento da *web*. Desde 2013, a *Society of American Archivists* (SAA) dispõe de uma seção arquivamento da *web*, exclusiva para tratar do assunto. Tem como “missão aumentar a defesa e a consciência das questões que os arquivistas encontram relacionadas com a seleção, avaliação, recolha, gestão, preservação e acesso aos conteúdos da *web*” (SOCIETY OF AMERICAN ARCHIVISTS, c2021a).

As metas e os objetivos da SAA – *Web archiving section* – são fornecer ao SAA uma voz para questões relacionadas com o arquivamento da *web* e melhorar o arquivamento da *web* por meio da criação de melhores práticas e normas baseadas na comunidade; almejando dar a direção e a liderança da SAA e o conhecimento a respeito do tema arquivamento da *web*. Além disso, pretende cultivar a necessidade de refletir sobre arquivos e coleções especiais, para a criação de novas linhas de trabalho ou para a reorientação das já existentes, de modo a incluir novos componentes ao arquivamento da *web*. Deseja também educar os membros da SAA, aumentando a sua compreensão e capacidade de implementar práticas de arquivamento da *web* nas suas próprias instituições. Por último, pretende servir de ponte para outras

organizações, em nível nacional e internacional, envolvidas no arquivamento da *web*, como a Reunião dos parceiros do *Archive-It* e do IIPC (SOCIETY OF AMERICAN ARCHIVISTS, c2021c).

Por intermédio de iniciativas como essas e a criação de grupos/fóruns, percebe-se a preocupação dos arquivistas com a temática, mostrando que o arquivamento e a preservação da *web* são também campos de estudos da Arquivologia.

Assim sendo, os temas “preservação digital e arquivamento da *web*” se tornaram base para reflexão crescente para autores no mundo entre 1990 e 2000, conforme alguns dos artigos publicados nesse período: “*The Cyberarchive: a look at the storage and preservation of web sites*” (CASEY, 1998); “*Preserving the Fabric of Our Lives: a survey of web preservation initiatives*” (DAY, 2003a); “*Archiving Websites: General Considerations and Strategies*” (BRÜGGER, 2005); e “*Archiving Web-published materials: A needs assessment of librarians, researchers, and content providers*” (MURRAY; HSIEH, 2008). Destaca-se também como publicação importante o livro “*Web Archiving*” (MASANÈS, 2006), cujo autor evidencia a importância da preservação da *web*, sentenciando que “é preciso fazer uma revisão radical nas práticas tradicionais de preservação³⁴”.

Ressaltam-se dois trabalhos de autores da temática encontrados no Google Acadêmico, produzidos em 2011: “*A survey on web archiving initiatives* (GOMES; MIRANDA; COSTA, 2011.)”, citado 101 vezes, em que os autores apresentam uma visão geral sobre as iniciativas de arquivamento da *web* em todo o mundo, e como resultado, foi criado um verbete na Wikipédia³⁵ com a lista de iniciativas; e “*Web Archiving – between Past, Present, and Future*” (BRÜGGER, 2011), citado 74 vezes, e que coloca em discussão a importância de inserir o arquivamento da *web* na agenda de pesquisa.

Em diferentes investigações, observaram-se, com destaque, os documentos “*Archiving Websites: A Practical Guide for Information Management Professionals*” (BROWN, 2006); “*The Archived Web: Doing History in the Digital Age*” (BRÜGGER, 2018); “*Web archiving bibliography 2013*” (REYES AYALA, 2013) e a bibliografia disponibilizada no *website* do IIPC³⁶. A *Online Computer Library Center – OCLC*, entidade sem fins lucrativos fundada em 1967, configura-se como uma comunidade de bibliotecas com membros de 107 países.

³⁴ Web preservation is for this reason a cultural and historical necessity. But the Web is also different from the previous publication systems to necessitate a radical revision of traditional preservation practices. (MASANÈS, 2006, p. 1).

³⁵ LIST of web archiving initiatives. In: WIKIPEDIA: a enciclopédia livre. [São Francisco, CA: Wikimedia Foundation, 2021]. Disponível em: https://en.wikipedia.org/wiki/List_of_Web_archiving_initiatives. Acesso em: 10 fev. 2021.

³⁶ <https://netpreserve.org/web-archiving/bibliography/>

Publicou, em 2018, três livros, tendo como foco metadados em *webarchiving*. O primeiro livro é o *Review of Harvesting Tools*; o Segundo, *Recommendations of the OCLC Research Library Partnership Web Archiving Metadata Working Group*; e o terceiro, *Literature Review of User Needs*. As três publicações são parte do trabalho do *OCLC Research Library Partnership Web Archiving Metadata Working Group* (WAM) e têm o objetivo de ajudar as instituições a melhorarem a consistência e as práticas relacionadas com os metadados das páginas arquivadas (OCLC, [c2021]).

Ainda em 2021, foi lançado o livro *The Past Web: Exploring Web Archives* (GOMES *et al.*, 2021), produzido por pesquisadores de Portugal, Reino Unido e Alemanha, que visa oferecer “informações práticas sobre arquivos da *web*, com exemplos inspiradores para arquivistas da *web*, levantando novos desafios e compartilhando resultados de pesquisas recentes sobre métodos de acesso para explorar informações do passado, preservadas por arquivos da *web*”³⁷.

4.2.2 Contexto brasileiro

No cenário brasileiro, foram levantadas no artigo “Arquivamento da Web: legislação correlata” (SANTOS, V., 2020, p. 2) “as legislações e proposições que abordassem os termos ‘*internet*’, ‘*páginas web*’, ‘*depósito legal*’ e ‘*preservação digital*’, para servir de subsídios para o grupo de pesquisa Dríade”, no eixo Arquivamento da *web*. O autor concluiu que:

Embora não exista uma legislação específica sobre a proteção, captura e preservação de sítios web de instituições públicas ou privadas, entende-se que é possível, à luz da legislação atual, defender a necessidade de que esse serviço seja prestado por instituições de memória e de acordo com a legislação arquivística. (SANTOS, V., 2020, p. 8).

Vanderlei Santos (2020, p. 8) considera importante realizar uma “análise mais detida sobre o tratamento que deve ser dado aos sítios institucionais, por exemplo, se no viés arquivístico ou no bibliográfico, visto que não é um entendimento pacificado entre os profissionais de preservação digital”.

Dantas (2015) realizou “uma investigação teórico-prática, de uma coleção digital de páginas das ferramentas de busca do Brasil, entre 1997 e 2013” – com apoio do *Internet Archive*, criou-se a coleção “Buscas.br”, por meio da plataforma *Archive-It*, facilitando a

³⁷ GOMES, D. *et al.* (ed.). *The past web: exploring web archives*. Suíça: Springer, 2021. Disponível em: <https://www.springer.com/gp/book/9783030632908>. Acesso em: 31 maio 2021.

curadoria e a pesquisa por palavras dos documentos *web* do Brasil, coletados pelo *Internet Archive* (DANTAS, 2015; ROCKEMBACH, 2018). A autora relata que, entre outras reflexões, encontrou algumas barreiras tecnológicas e culturais associadas à preservação e ao acesso à memória digital, e que, no contexto brasileiro, os registros natos digitais não têm sido alvo de uma política ativa de preservação (DANTAS, 2015, p. 147).

Um ponto significativo acerca dessa discussão é que, recentemente, foram defendidas cinco dissertações e uma tese do Programa de Pós-Graduação em Comunicação e Informação da Universidade Federal do Rio Grande do Sul (UFRGS), tendo como tema o arquivamento da *web*. São elas: Lisiane Ferreira (2019), “Arquivamento da web e mídias sociais: preservação digital de vídeos da campanha presidencial brasileira de 2018”, com enfoque na plataforma Facebook; Martins (2019), “Mapeamento de públicos estratégicos para iniciativas brasileiras universitárias de arquivamento da web no âmbito acadêmico: uma projeção para Universidade Federal do Rio Grande do Sul e seu Programa de Pós-graduação em Comunicação”, com ênfase nas universidades; Melo (2020), “Arquivamento dos websites do governo federal brasileiro: preservação do domínio gov.br”, sobre as possibilidades de arquivamento das páginas oficiais do governo federal, a qual recebeu o Prêmio Olga Gallego de Investigación en Archivos 2021³⁸, da *Fundación Olga Gallego*, da Espanha, destacando-se como trabalho inovador; Nunes (2021), “Arquivamento da web: aspectos éticos e legais no acesso e uso da informação”; e Luz (2021) “Comunicação pública e memória comunicacional”.

Todos os autores enfatizaram a importância de iniciativas e práticas de arquivamento da *web* (FERREIRA, L., 2019; MARTINS, 2019; MELO, 2020; NUNES, 2021) – “para que seja possível recuperar esses conteúdos ao longo do tempo, é preciso o estabelecimento de comunidades de pesquisa e práticas de arquivamento da *web*” (ROCKEMBACH, 2019, p. 137). Segundo Rockembach (2019), tal preocupação é algo que necessita ser incentivado, para que tenhamos um fomento da preservação dos conteúdos *web* nacionais e locais (ROCKEMBACH, 2019, p. 137).

Rezende e Martins (2018) publicaram o artigo “Iniciativas científicas de arquivamento e preservação de conteúdos em mídias sociais: panorama atual”, no qual fazem “um mapeamento de experiências científicas que realizam coleta e arquivamento de conteúdos de mídias sociais”, particularmente do Twitter, e são evidenciados vários desafios no cerne do

³⁸ FUNDACION OLGA GALLEGU. *Fallo da IV edición do Premio Olga Gallego de Investigación en Archivos*. 08 mar. 2021. Disponível em: <https://www.fundacionolgagallego.gal/actualidade/fallo-da-iv-edicion-do-premio-olga-gallego-de-investigacion-en-archivos.htm>. Acesso em: 08 jun. 2021.

arquivamento, da preservação e curadoria desses tipos de conteúdos digitais. (REZENDE; MARTINS, 2018, p. 219).

Alguns trabalhos apresentados em eventos são pontos de reflexão sobre o tema. Na 10ª Conferência Luso-Brasileira de Ciência Aberta – ConfOA, em 2019, foi apresentado o trabalho “*Web Archiving como ferramenta de preservação digital: colaboração da Library of Congress, Rio Office no projeto dos Jogos Olímpicos de 2016*”. Segundo Fazano e Fonseca (2020), o trabalho “é parte de uma pesquisa em andamento, no curso de Mestrado Profissional em Bens Culturais e Projetos Sociais, da Fundação Getúlio Vargas” e tem a “pretensão de destacar a importância de discussões nessa seara no Brasil como forma de preservar o que se é produzido hoje em sítios da *internet* no país para que seja garantido o acesso às futuras gerações” (FAZANO; FONSECA, 2020).

Percebe-se, entre os autores e pesquisadores do assunto, a valorização de iniciativas no campo, principalmente no âmbito das instituições de memória e patrimônio. Para Alencar Brayner (2016), é fundamental a guarda e a preservação das páginas *web*, pois se configuram em um material expressivo do patrimônio cultural. O autor salientou a falta de políticas nesse sentido no contexto brasileiro e apresenta a experiência do Arquivo Nacional do Reino Unido, no artigo “Programa do arquivo de páginas *web* no Reino Unido: Uma breve história de oportunidades e desafios” (ALENCAR BRAYNER, 2016).

No Brasil, as produções acadêmicas referentes ao tema, em sua maioria, são provenientes das pesquisas desenvolvidas no Nuaweb, criado em 2017, coordenado pelo professor Moisés Rockembach, com “objetivo de investigar características do arquivamento da *web* por meio de iniciativas nacionais e internacionais, lidando tanto com as políticas, quanto as tecnologias envolvidas no processo” (UFRGS, [2020]). Integrantes do Nuaweb, os professores Moisés Rockembach e Caterina Pavão (2018) afirmaram no artigo “Políticas e tecnologias de preservação digital no arquivamento da *web*” serem “três grandes áreas de pesquisa que investigam e publicam sobre o tema de arquivamento da *web* e estão concentradas nas Ciências Sociais, Ciências da Computação e Engenharia, seguido de Artes e Humanidades, *Decision Sciences* e Negócios, Administração e Contabilidade” (ROCKEMBACH; PAVÃO, 2018, p. 173). Segundo os autores, muitos utilizam como palavra-chave o termo “*Digital Preservation*”.

O Brasil sediou alguns eventos relacionados ao tema arquivamento da *web*. No ano de 2019, o 9º Fórum da Internet do Brasil em Manaus, em 2 de outubro de 2019, teve a mesa

“Preservação do Conteúdo Web Brasileiro”³⁹, com o objetivo de “defender a criação de políticas públicas para preservação do conteúdo *web* brasileiro”. Uma conclusão unânime entre os palestrantes foi a urgência de estabelecer modelos de arquivamento do conteúdo *on-line*. No mesmo, ano aconteceu a aula aberta “O Uso dos Arquivos da Web na pesquisa acadêmica”⁴⁰, com o curador digital Ricardo Basílio, do Arquivo.PT (Portugal), promovido pela UFRGS em 29 de novembro de 2019.

Em 2020, a Universidade de São Paulo – USP, em parceria com a Federação Brasileira de Associações de Bibliotecários, Cientistas de Informação e Instituições – FEBAB, organizou o *VI Congresso Internacional em Tecnologia e Organização da Informação (TOI): I Simpósio Políticas e Estratégias de Preservação de Conteúdo na Web*⁴¹, com intuito de promover o debate em temas centrais sobre arquivamento da *web*, preservação e políticas públicas para o patrimônio digital nacional.

O tema também foi abordado no *World Digital Preservation Day 2020*⁴² promovido no Brasil pela Rede Cariniana (Ibict), com a participação do professor Moises Rokembach, com o trabalho “Preservação da web na UFRGS”.

Em 2021, o grupo de pesquisa Dríade – Estudo e Práticas de Preservação Digital, ligado à Rede Cariniana e ao Instituto Brasileiro de Informação em Ciência e Tecnologia (Ibict) – realizou o *webinar Arquivamento e Preservação da web – trocando experiências*⁴³, com intuito de promover a troca de experiências sobre o tema, contando com pesquisadores brasileiros e catalães. Ademais, o V Seminário Internacional de Preservação Digital – Sindpred, em 2021, realizado virtualmente pela Rede Cariniana e Universidade Estadual de Campinas – Unicamp, com a colaboração da Biblioteca Nacional do Brasil, contou com a mesa 8 – *Preservação de Conteúdo Web* –, com o palestrante internacional Michael Day, da *British Library*, que mostrou o trabalho realizado por aquela instituição. Ao final do V Sinpred, foi lançada a “Bibliografia sobre preservação digital: um levantamento nos diversos

³⁹ COMITÊ GESTOR DA INTERNET NO BRASIL – CGI.BR. [Workshop] Preservação do conteúdo web brasileiro. 2 out. 2019. In: FÓRUM DA INTERNET NO BRASIL, 9., 1-4 out. 2019, Manaus. [S. l.]: CGI.BR, 2019. Disponível em: <https://forumdainternet.cgi.br/2019/programacao/detalhe/2/1577/>. Acesso em: 01 jun. 2021.

⁴⁰ BASÍLIO, R. *O uso dos arquivos da web na pesquisa acadêmica*. 2019. Disponível em: <https://www.ufrgs.br/nuaweb/wp-content/uploads/2019/11/divulgacao-29-11.png>. Acesso em: 26 maio 2022.

⁴¹ SIMPÓSIO POLÍTICAS E ESTRATÉGIAS DE PRESERVAÇÃO DE CONTEÚDO NA WEB, 1., 2020. [S. l.]: FEBAB, 11 ago 2020. Publicado no canal FEBAB. Disponível em: <https://www.youtube.com/watch?v=X6SmBmbD7fg>. Disponível em: 01 jun. 2021

⁴² REDE CARINIANA. *World Digital Preservation Day 2020*. 22 out. 2020. Disponível em: <https://cariniana.ibict.br/index.php/noticias/414-world-digital-preservation-day-2020>. Acesso em: 01 jun. 2021.

⁴³ REDE CARINIANA. *Arquivamento e preservação da web*. 2 fev. 2021. Disponível em: <https://youtu.be/QPZuLwWR9y8>. Acesso em: 01 jun. 2021.

suportes informacionais”, contendo, entre os temas, uma seção sobre arquivamento de páginas *Web*⁴⁴.

Nesse mesmo evento, foi apresentado o trabalho intitulado “Ensaio sobre arquivamento de páginas web: foco na experiência do Portal de Periódicos da UNICAMP, utilizando o Conifer (Rhizome)”, que tem como objetivo “a preservação e arquivamento das páginas *web* do Portal de Periódicos Científicos da UNICAMP” (SANTOS, G., 2021). Nota-se que a abordagem utilizada foi a de “*microarchivig*”, mencionada por Brügger (2005), visto que o arquivamento é realizado por um indivíduo com a intenção de preservar um determinado objeto de estudo.

Outro artigo lançado foi “A web brasileira na Covid-19: arquivamento da web e preservação digital”. A pesquisa “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” (ROCKEMBACH, 2021a, p. 1).

Ao longo de 2021, alguns avanços foram realizados no contexto brasileiro, o Conselho Nacional de Arquivos (Conarq) criou a Câmara Técnica Consultiva de Preservação de *websites* e mídias sociais⁴⁵, instituída pela Portaria nº 131, de 9 de novembro de 2021, conforme aprovação do Plenário do Conarq em sua reunião extraordinária de 27 de outubro de 2021, a qual realizou as primeiras reuniões de trabalho nos meses de novembro e dezembro. A Câmara Técnica Consultiva tem prazo de um ano para elaborar uma proposta de resolução com recomendações de elementos mínimos de uma política de preservação e de requisitos mínimos de preservação para *websites* e mídias sociais para apreciação do Conarq.

Aconteceu, de forma virtual, o *II Simpósio Políticas e Estratégias de Preservação de Conteúdo na Web*⁴⁶ – um dos eixos temáticos do *VII Congresso Internacional em Tecnologia e Organização da Informação* – tendo como proposta promover o debate em temas centrais sobre arquivamento da *web*, preservação e políticas públicas para o patrimônio digital nacional, contando com pesquisadores brasileiros, portugueses e chilenos.

Recentemente, foi publicado o artigo de Formento e Gracioso (2022) “Padrões de metadados no arquivamento da web: recursos tecnológicos para a garantia da preservação

⁴⁴ REDE CARINIANA. V *SINPRED*. 2021. Disponível em: <https://www.youtube.com/channel/UC7pJ4CrMcF7WVYsWjic-nsg>. Acesso em: 8 jun. 2021.

⁴⁵ CONSELHO NACIONAL DE ARQUIVOS – CONARQ. Câmara Técnica Consultiva do Conarq sobre Preservação de websites e mídias sociais inicia seus trabalhos. *CONARQ Notícias*, 8 dez. 2021. Disponível em: <https://www.gov.br/conarq/pt-br/assuntos/noticias/camara-tecnica-consultiva-do-conarq-sobre-preservacao-de-websites-e-midias-sociais-inicia-seus-trabalhos>. Acesso em: 07 jan. 2022.

⁴⁶ II SIMPÓSIO POLÍTICAS E ESTRATÉGIAS DE PRESERVAÇÃO DE CONTEÚDO NA WEB, TOI 2021 ONLINE, 15 dez. 2021. 1 vídeo (160 min). Publicado pelo canal TOI 2021 ONLINE. Disponível em: <https://www.youtube.com/watch?v=3KhAu7OczSA>. Acesso em: 23 dez. 2021.

digital de websites arquivados”, em que os autores analisaram diferentes padrões de metadados para o arquivamento e preservação da *web*.

Percebe-se que, nos últimos anos, os pesquisadores e as instituições no Brasil vêm fazendo esforços para a promoção do tema arquivamento e preservação da *web*, mas são tentativas embrionárias frente à conjuntura internacional que será apresentada.

5 PATRIMÔNIO DIGITAL

A Unesco, na Conferência Geral em Paris, de 17 de outubro a 21 de novembro de 1972, assinou a Convenção para a Proteção do Patrimônio Mundial, Cultural e Natural em 16 de novembro de 1972, na qual apresenta, em seu art. 1º, a definição de patrimônio cultural: primeiro “*os monumentos*. – obras arquitetônicas, de escultura ou de pintura monumentais, elementos de estruturas de caráter arqueológico, inscrições, grutas e grupos de elementos com valor universal excepcional do ponto de vista da história, da arte ou da ciência”; segundo “*os conjuntos*. – grupos de construções isoladas ou reunidas que, em virtude da sua arquitetura, unidade ou integração na paisagem tem valor universal excepcional do ponto de vista da história, da arte ou da ciência”; e, por último, “*os locais de interesse*. – obras do homem, ou obras conjugadas do homem e da natureza, e as zonas, incluindo os locais de interesse arqueológico, com um valor universal excepcional do ponto de vista histórico, estético, etnológico ou antropológico” (UNESCO, 1972, p. [2], grifo no original). Para a Unesco, a definição de patrimônio é ramificada nesses três grandes grupos.

Cunha e Cavalcanti (2008) conceituam patrimônio cultural como “conjunto de bens móveis e imóveis existentes no país cuja conservação seja de interesse público quer por sua vinculação a fatos memoráveis, quer por seu excepcional valor” (CUNHA; CAVALCANTI, 2008, p. 277). Nos documentos elaborados pela Unesco, o termo “patrimônio é definido também como herança do passado, nosso patrimônio atual é o que deixamos de legado às gerações futuras. O patrimônio é, ou deveria ser algo que é passado de geração em geração porque é valorizado” (UNESCO, c2021b).

De maneira complementar, é fundamental destacar que patrimônio é um tema importante para a execução desta pesquisa. De acordo com Crippa (2011, p. 53), está vinculado à construção do Estado-Nação, tornando-se elemento articulador das entidades nacionais, discutido por arquitetos e urbanistas, historiadores, sociólogos, por exemplo. Jardim e Fonseca (2008) afirmam que:

A partir do séc. XIX, no interior do projeto de Estado Nacional, desenham-se concepções de memória e inventam-se tradições para uma *nação* que reserve um passado comum aos seus integrantes. A noção de *patrimônio histórico-cultural* insere-se neste processo pelo qual o Estado se organiza mediante a criação de um patrimônio comum e uma identidade própria. Os arquivos, bibliotecas e museus nacionais são produtos da construção desse patrimônio. (JARDIM; FONSECA, 2008, p. 127, grifo no original).

O patrimônio produzido pela sociedade atual no cenário das comunicações, seja escrito, visual ou oral, em sua maioria, é em ambiente digital. Tendo em conta que os documentos digitais devem permanecer acessíveis por longo prazo, a Unesco, na 32ª Conferência Geral, em 17 de outubro de 2003, publicou a *Carta sobre a Preservação do Patrimônio Digital*:

Reconhecendo que os recursos de informação e expressão criativa se elaboram, distribuem, utilizam e conservam cada vez mais em forma eletrônica, e que ele dá lugar ao um novo tipo de legado: o patrimônio digital, **Consciente** de que o acesso ao dito patrimônio brindará maiores oportunidades de criação, comunicação e intercâmbio de conhecimento entre todos os povos, **Entendendo** que este patrimônio digital se encontra em perigo de desaparecer, e que sua preservação em benefício das gerações reais e futuras é uma preocupação urgente no mundo inteiro. (UNESCO, 2004, p. 79, grifo no original, tradução nossa)⁴⁷.

O teor da carta indica, no art. 1º⁴⁸, que as páginas *web* são parte da “herança digital, como recursos únicos resultado dos conhecimentos ou expressões de seres humanos” e que o patrimônio digital inclui “os recursos culturais, educativos, científicos ou administrativos, informação técnica, jurídica, médica e outras”, que “são gerados diretamente em formato digital ou convertidos para formato digital a partir de material analógico existente. Os produtos de ‘origem digital’ não existem em nenhum outro formato que não seja eletrônico”. Além disso, apresenta como objetos digitais “textos, bases de dados, imagens fixas ou em movimento, gravações de som, gráficos, *software* ou páginas *web*, entre muitos outros formatos possíveis, num vasto e cada vez mais diversificado repertório”. Frisa que a efemeridade da *web* é um dos maiores desafios do trabalho de preservação no que diz respeito aos “processos de produção, manutenção e gestão” (UNESCO, 2004, p. 80). O texto da carta

⁴⁷ **Reconociendo** que esos recursos de información y expresión creativa se elaboran, distribuyen, utilizan y conservan cada vez más en forma electrónica, y que ello da lugar a un nuevo tipo de legado: el patrimonio digital, **Consciente** de que el acceso a dicho patrimonio brindará mayores oportunidades de creación, comunicación e intercambio de conocimientos entre todos los pueblos, **Entendiendo** que este patrimonio digital se encuentra en peligro de desaparición, y que su preservación en beneficio de las generaciones actuales y futuras es una preocupación urgente en el mundo entero (UNESCO, 2004, p. 79, grifo no original).

⁴⁸ El patrimonio digital consiste en recursos únicos que son fruto del saber o la expresión de los seres humanos. Comprende recursos de carácter cultural, educativo, científico o administrativo e información técnica, jurídica, médica y de otras clases, que se generan directamente en formato digital o se convierten a éste a partir de material analógico ya existente. Los productos “de origen digital” no existen en otro formato que el electrónico. Los objetos digitales pueden ser textos, bases de datos, imágenes fijas o en movimiento, grabaciones sonoras, material gráfico, programas informáticos o páginas Web, entre otros muchos formatos posibles dentro de un vasto repertorio de diversidad creciente. A menudo son efímeros, y su conservación requiere un trabajo específico en este sentido en los procesos de producción, mantenimiento y gestión. Muchos de esos recursos revisten valor e importancia duraderos, y constituyen por ello un patrimonio digno de protección y conservación en beneficio de las generaciones actuales y futuras. Este legado en constante aumento puede existir en cualquier lengua, cualquier lugar del mundo y cualquier campo de la el saber humanos (UNESCO, 2004, p. 80).

afirma que grande parte dos objetos digitais têm “valor e significado” de longo prazo, “constituindo assim um patrimônio digno de proteção e preservação”, independentemente de área do conhecimento, idioma ou lugar do mundo (UNESCO, 2004, p. 80).

O documento reconhece que o patrimônio digital corre o risco de desaparecer, caso não haja a preocupação de preservá-lo para as futuras gerações (UNESCO, 2004, p. 80, tradução nossa). É ressaltado, no art. 3º, o perigo da perda:

O patrimônio digital do mundo está em perigo de se perder para a posteridade. Os fatores que contribuem para isso incluem a rápida obsolescência do hardware e software que lhe dá vida, incertezas sobre os recursos, responsabilidade e métodos para a sua manutenção e preservação, e a falta de legislação para apoiar estes processos. A mudança comportamental tem ficado para trás do progresso tecnológico. A evolução da tecnologia digital tem sido tão rápida e dispendiosa que os governos e as instituições têm sido incapazes de desenvolver estratégias de conservação adequadas e fundamentadas. Não foi compreendida completamente a ameaça ao potencial econômico, social, intelectual e cultural do patrimônio, sobre o qual o futuro está construído. (UNESCO, 2004, p. 80, tradução nossa)⁴⁹.

Observa-se que a *Carta sobre a Preservação do Patrimônio Digital* também trabalha a noção de patrimônio no ambiente. Cavalcante (2007) realizou a pesquisa de pós-doutorado, na qual fez uma reflexão, com base na referida carta da Unesco, intitulada *Patrimônio digital e informação: política, cultura e diversidade*. O objetivo era “apresentar algumas reflexões sobre os desafios de salvaguardar e dar acesso ao patrimônio, em meio eletrônico, com ênfase nas questões culturais, econômicas e políticas que envolvem a temática” (CAVALCANTE, 2007, p. 152). A autora afirma que a Unesco e os documentos produzidos por ela “apoiam cotidianamente ações culturais, sociais e políticas, também por meio de programas de digitalização de acervos ao redor do mundo, o que a torna um dos agentes mais importantes e decisivos para o processo de patrimonialização digital, que tende a efetivar-se mundialmente” (CAVALCANTE, 2007, p. 154).

Concomitante à publicação da *Carta sobre a Preservação do Patrimônio Digital* em 2003, a Biblioteca Nacional da Austrália (*National Library of Australia*) produziu as “Diretrizes para a preservação do patrimônio digital” (*Guidelines for the preservation of digital heritage*) para a Unesco. O documento “fornece orientações gerais e técnicas para a

⁴⁹ El patrimonio digital del mundo corre el peligro de perderse para la posteridad. Contribuyen a ello, entre otros factores, la rápida obsolescencia de los equipos y programas informáticos que le dan vida, las incertidumbres existentes en torno a los recursos, la responsabilidad y los métodos para su mantenimiento y conservación y la falta de legislación que ampare estos procesos.

Los cambios en las conductas han ido a la zaga del progreso tecnológico. La evolución de la tecnología digital ha sido tan rápida y onerosa que los gobiernos e instituciones no han podido elaborar estrategias de conservación oportunas y bien fundamentadas. No se ha comprendido en toda su magnitud la amenaza que pesa sobre el potencial económico, social, intelectual y cultural que encierra el patrimonio, sobre el cual se edifica el porvenir. (UNESCO, 2004, p. 80).

preservação e acesso contínuo do crescente patrimônio digital mundial”, e foi elaborado a partir de “reuniões regionais, nas quais houve a participação de 175 especialistas de 86 países, representando uma variedade de instituições e interesses, incluindo bibliotecas, arquivos, fornecedores de serviços de *internet*, organismos nacionais de normalização, fabricantes de *hardware* e *software*, jornalistas, advogados, universidades e agências governamentais” (NATIONAL LIBRARY OF AUSTRALIA, 2003, p. 5). Compreende-se como o patrimônio digital:

[...] bens de informação de valor duradouro, dignos de preservação para as gerações futuras, de diferentes comunidades, indústrias, setores e regiões. Nem todos os objetos digitais têm um valor duradouro, mas se este for o caso, a preservação deve ser abordada de forma proativa com pretensão de manter a continuidade do patrimônio digital. (NATIONAL LIBRARY OF AUSTRALIA, 2003, p. 29, tradução nossa).

A Biblioteca Nacional da Austrália apresentou e definiu no documento os tipos de patrimônio digital; e entre os tipos mais difíceis de definir, estão as páginas *web*. Ainda é mencionado que, embora haja iniciativas em diversos países para preservação do patrimônio digital, inclusive das páginas *web*, era necessária a adoção de normas internacionais, ressaltando como pontos-chave da preservação do patrimônio digital a cooperação e o compartilhamento entre as instâncias nacionais e internacionais.

Em 1992, a Unesco criou o programa internacional “Memória do Mundo”, com a visão “de que o patrimônio documental mundial pertence a todos e deve ser completamente preservado e protegido por todos [...] e [estar] permanentemente acessível”. O programa tem como missão “facilitar a preservação, utilizando das técnicas mais adequadas, promover o acesso e aumentar a consciência mundial sobre a existência e a importância do patrimônio documental” (UNESCO, [c2021b]). O Brasil⁵⁰ conseguiu dez registros internacionais ao longo dos anos.

A Unesco, na 38ª Conferência Anual, em 30 de julho de 2015, apresentou o projeto “Salvaguardar a Memória do Mundo, com recomendações relativas à preservação do patrimônio documental, incluindo o patrimônio digital e o acesso ao mesmo”. Considera-se no projeto de resolução que há a necessidade de normativas que abarquem o tema, que poderiam servir de instrumentos de cooperação internacional, para assim sensibilizar a comunidade sobre a temática da preservação do patrimônio documental (UNESCO, 2015). O

⁵⁰ UNESCO. *Memory of the world*. [c2017]. Disponível em: <http://www.unesco.org/new/en/communication-and-information/memory-of-the-world/register/access-by-region-and-country/br/>. Acesso em: 1 jun. 2021.

documento final, aprovado em 17 de novembro de 2015, intitulado de *Recommendation concerning the Preservation of, and Access to, Documentary Heritage Including in Digital Form*, considera como desafiador estabelecer modelos e processos para preservar objetos do patrimônio digital, em particular, os objetos complexos, tais como obras multimídia interativa, diálogos na internet, dados dinâmicos, conteúdo de dispositivos móveis e formatos que possam surgir no futuro (UNESCO, 2016b).

Ainda foi ressaltado que a Unesco criou, em 1992, o Programa Memória do Mundo para aumentar o conhecimento e a proteção do patrimônio documental mundial e para torná-lo universal e permanentemente acessível, inclusive tendo em conta a rápida evolução da tecnologia e o desafio de estabelecer modelos e processos para preservar objetos do patrimônio digital, principalmente os objetos complexos, tais como obras na internet (UNESCO, 2016b). Observa-se que a preocupação com patrimônio digital proveniente da *web* persistiu nas recomendações da Unesco.

Ademais, menciona como definição importante no contexto do documento as “instituições de memória”, que incluiriam os “arquivos, bibliotecas, museus e outras organizações educacionais, culturais e de investigação”, definindo os seus papéis dentro do desafio da preservação do patrimônio, afirmando que “as instituições de memória, em coordenação com a sociedade civil, devem estabelecer políticas, mecanismos e critérios para selecionar, adquirir e descartar o patrimônio documental” (UNESCO, 2015):

[...] tendo em conta não só os documentos mais importantes, mas também o seu material contextual, incluindo os meios de comunicações sociais. Os critérios de seleção devem ser de natureza não discriminatória e claramente definidos. Além disso, a escolha deve ser neutra e equilibrada com respeito aos campos do conhecimento, expressões artísticas e períodos históricos. Dada a sua natureza inerentemente temporária, as decisões relativas à preservação de materiais digitais podem ser tomadas no momento, ou antes, da sua criação. (UNESCO, 2015, p. 3, tradução nossa).

A Unesco demonstra, com os documentos publicados, a dimensão do trabalho e o cuidado que as instituições precisam ter para com a preservação do patrimônio. Por conseguinte, realizou uma Conferência Internacional em Vancouver, Canadá, entre 26 e 28 de setembro de 2012, com o título “A Memória do Mundo na Era Digital: Digitalização e Preservação”. A conferência foi aberta com o discurso⁵¹ de Janis Karklins, Subdiretor-Geral de Comunicação e Informação da Unesco, e o evento tinha o intuito de explorar as principais

⁵¹ UNESCO TV. *The Memory of the World in the Digital age: digitization and preservation*. [S. l.], 2012. Speaker: Janis 2012. Disponível em: <https://www.unesco.org/archives/multimedia/document-2460>. Acesso em: 01 jun. 2021.

questões que afetam a preservação do patrimônio documental digital, a fim de desenvolver estratégias para contribuir com a proteção dos bens digitais e ajudar a definir uma metodologia de implementação apropriada aos países em desenvolvimento (UNESCO, 2012).

Na Conferência, mais de 500 participantes de 110 países contribuíram para a produção da Declaração Unesco/UBC Vancouver (*UNESCO/UBC Vancouver Declaration*) e discutiram pontos-chave que afetam os dois principais aspectos dos registros, documentos e dados no ambiente digital: o primeiro, os “assuntos pertinentes à digitalização de material; e o segundo os assuntos pertinentes à continuidade, ao acesso e à preservação de materiais digitais autênticos, confiáveis e exatos” (UNESCO, 2012).

Dessa maneira, a organização publicou, em 2016, as “*Directrices UNESCO/PERSIST sobre selección del patrimonio digital para su conservación a largo plazo*” (Diretrizes Unesco/PERSIST sobre a seleção do patrimônio digital para sua preservação em longo prazo), com o apoio dos organismos internacionais de profissionais, como o *International Council on Archives* – ICA (Conselho Internacional de Arquivos), a *International Federation of Library Associations and Institutions* – IFLA (Federação Internacional de Associações e Instituições de Bibliotecas) e o *International Council of Museums* – ICOM (Conselho Internacional dos Museus), visando “ajudar os profissionais e suas instituições na tarefa de preservar o patrimônio digital de maneira sustentável a longo prazo (UNESCO, 2016a).”.

O “programa PERSIST serviu como um catalisador para o trabalho de preservação digital da UNESCO, como parte do Plano de Ação para o Fortalecimento do Programa Memória do Mundo (adotado pelo 191º Conselho Executivo, objetivo 4a)” (UNESCO PERSIST, [c2021c]). No texto introdutório das Diretrizes Unesco/PERSIST, é afirmado que o patrimônio e as responsabilidades são distintos conforme o contexto:

[...] O patrimônio digital pode diferir consideravelmente entre comunidades, regiões e países, sua preservação requer o compromisso e a cooperação do setor público, do setor privado e dos criadores de conteúdo. Assim como as instituições públicas têm uma responsabilidade jurídica intrínseca de gerenciar coleções de patrimônio, o setor privado deve da mesma forma, enfrentar o desafio de preservar e garantir o acesso às suas informações digitais⁵². (UNESCO, 2016a, p. 4, tradução nossa).

⁵² Dado que el patrimonio digital puede diferir considerablemente entre comunidades, regiones y países, su preservación requiere del compromiso y la cooperación del sector público, del sector privado y de los creadores de contenido. Al igual que las instituciones públicas tienen la responsabilidad legal intrínseca de gestionar las colecciones patrimoniales, el sector privado debe, del mismo modo, hacer frente al desafío de preservar y asegurar el acceso a su información digital (UNESCO, 2016a, p. 4).

É apontado o papel das instituições de âmbito nacional a respeito do patrimônio digital, considerando que elas “devem desempenhar um papel fundamental na liderança das comunidades patrimoniais nas questões de seleção e preservação digital”. O documento frisa que “em muitos países, as instituições nacionais designadas para tal missão, possuem legislação em vigor para depósito legal relativo ao patrimônio cultural”. Afirmar, por exemplo, “que geralmente as publicações dos governos ou documentos oficiais, exigem guarda permanente e nos casos em que a legislação apenas abarque o material físico, a mesma deverá ser adaptada para cobrir o material digital”. Sugere que as instituições nacionais “trabalhem de forma colaborativa para desenvolver estratégias e padrões para recolher, organizar e preservar informações digitais, contando com apoio de organismos internacionais como IFLA, ICA e ICOM, para este grande desafio global de preservar o patrimônio digital” (UNESCO, 2016a). Além disso, afirma que o trabalho institucional é:

Adquirir e coletar o patrimônio digital, por meio de diferentes canais e plataformas, requer esforços e recursos significativos. Por exemplo, os domínios da *web* nacionais podem hospedar milhares a milhões de páginas da *web* nas quais vários milhões ou bilhões de arquivos são publicados, atualizados ou excluídos diariamente. Grande parte deste patrimônio digital vale a pena preservar para o futuro, mesmo quando estiver *off-line* (por exemplo, dados brutos de pesquisas, documentos governamentais, arquivos digitais privados de organizações e/ou indivíduos). A abrangência desse desafio torna essencial que as grandes instituições nacionais tenham um papel de liderança, estabelecendo suas próprias políticas e sistemas de coleta e gerenciamento de material digital, ou liderando redes colaborativas para a adoção de modelos comuns de seleção e preservação. (UNESCO, 2016a, p. 5, tradução nossa).

Para Cavalcante (2007, p. 167), “o patrimônio digital é, sobretudo, uma construção social, na qual os seres humanos e suas ações no tempo e no espaço constituem a grande riqueza documentada”, que, além de envolver o processo de guarda, compreende também a produção de conhecimento, preservação e acesso.

Assim como a Unesco, a União Europeia (UE) desenvolve e promove ações para salvaguardar o patrimônio cultural dos países-membros. A Comissão Europeia (*European Commission*) é o organismo que ajuda a definir a estratégia geral da UE, propondo novos atos legislativos e políticas, assim como a execução e o gerenciamento de orçamento, bem como desempenha um papel importante no apoio ao desenvolvimento internacional e na prestação de ajuda humanitária. Possui a área denominada de “*Culture and Creativity*” em seu *website*; e em uma das subseções, está alocada a “*European digital heritage*”, que faz parte de um conjunto significativo de políticas públicas no âmbito cultural (EUROPEAN COMMISSION, 2021).

Uma das responsabilidades da Comissão Europeia “é a política, a coordenação e o financiamento de ações para apoiar a digitalização e o acesso em linha dos Estados-Membros ao material cultural e à preservação digital”, competência da Direção-Geral das Redes de Comunicações, Conteúdo e Tecnologia (EUROPEAN COMMISSION, 2021). Além disso, apresenta mais informações sobre o projeto Europeana – agregador europeu de conteúdos salvaguardados em instituições patrimoniais; também sobre digitalização e preservação de documentos, entre outros ligados a cultura com aporte tecnológico. Constatase que a comunidade europeia desenvolve várias ações de preservação de patrimônio cultural, com recomendações e apoio financeiro, inclusive projetos e iniciativas de *web archiving*.

Para Dodebei (2006, p. [8]), “a web vista como espaço mítico da memória social cria um fértil terreno de pesquisas sobre o comportamento e as propriedades dos meios de produção do conhecimento, quer sejam eles de natureza histórica, artística ou técnica”, ressaltando que “a web pode ser um grande centro virtual da memória do mundo”. É salientado pela autora que “o patrimônio digital pode ser estudado sob dois enfoques não exclusivos mutuamente”, sendo que:

[...] o primeiro o considera como processo em constante produção, seja dos objetos digitalizados ou dos objetos já nascidos digitais. O segundo é representado pelo patrimônio como produto de uma escolha, quer dizer um conjunto isolado de objetos escolhidos para representar simbolicamente as ações sociais. Este último pressupõe constante atenção quanto à salvaguarda de sua integridade singular. A esses dois enfoques correspondem modelos teórico-metodológicos que podemos, por analogia, referir à Memória e à História. (DODEBEI, 2006, p. [8]).

Observam-se os diferentes desafios de salvaguardar o patrimônio digital, o qual envolve diversas instâncias e atores para a preservação, em especial, dos objetos digitais complexos como as páginas *web*, patrimônio que poderá ter um tratamento distinto conforme a instituição de guarda, por exemplo, arquivos e bibliotecas.

5.1 PATRIMÔNIO ARQUIVÍSTICO DIGITAL

Os arquivos nacionais são instituições de Estado responsáveis pelo recolhimento, salvaguarda, preservação e acesso dos documentos gerados pela administração pública de um país, bem como por orientar, por meio das políticas de gestão e normativas, os arquivos públicos.

Para Fonseca (2005, p. 40), dessa forma, “o Estado reconhece sua responsabilidade em relação ao cuidado devido ao patrimônio documental do passado e aos documentos

produzidos ao logo de suas atividades”. Atualmente, o patrimônio documental foi transportado do papel para o ambiente binário, provocando muitas mudanças nas responsabilidades e na missão institucional dos arquivos.

A *Carta sobre a Preservação do Patrimônio Digital* da Unesco (2003) foi adaptada pelo Conselho Nacional de Arquivos no Brasil (Conarq) em 2005, com o título *Carta para a Preservação do Patrimônio Arquivístico Digital*. O trabalho foi realizado pelos membros da Câmara Técnica de Documentos Eletrônicos, do Conarq, enfatizando os documentos arquivísticos.

A referida carta considera que “a informação arquivística produzida, recebida, utilizada e conservada em sistemas informatizados vem constituindo um novo tipo de legado: o patrimônio arquivístico digital” (CONARQ, 2005, p. 1), e ainda:

Considerando que esse patrimônio arquivístico digital se encontra em perigo de desaparecimento e de falta de confiabilidade, e que sua preservação em benefício das gerações atuais e futuras é uma preocupação urgente no mundo inteiro;

Considerando que a Carta para a Preservação do Patrimônio Digital da UNESCO manifesta a necessidade de os Estados-membros, incluindo o Brasil, estabelecerem políticas e ações para proteger o patrimônio digital;

Considerando que o Conselho Internacional de Arquivos estabeleceu entre seus princípios que os arquivos devem facilitar o estabelecimento de políticas, procedimentos, sistemas, normas e práticas que levem os produtores de documentos a criar e manter documentos arquivísticos fidedignos, autênticos, preserváveis e acessíveis;

Considerando que o Conselho Nacional de Arquivos tem por finalidade definir política nacional de arquivos públicos e privados e exercer orientação normativa visando à gestão documental e à proteção especial aos documentos de arquivo, independente do suporte em que a informação está registrada; o Conselho Nacional de Arquivos, em sua 34ª reunião plenária, realizada em 6 de julho de 2004, no Rio de Janeiro, aprova a presente Carta. (CONARQ, 2005, p. 1, grifo no original).

É ressaltado na carta que “os documentos arquivísticos são gerados e mantidos por organizações e pessoas para registrar suas atividades e servirem como fontes de prova e informação”, e preservar esses documentos tem “por objetivo garantir a autenticidade e a integridade da informação, enquanto o acesso depende de que os mesmos estejam em condições de serem utilizados e compreendidos” (CONARQ, 2005, p. 2).

A *Carta para a Preservação do Patrimônio Arquivístico Digital* (2005) explicita como parte do patrimônio as páginas *web*, assim como a Carta da Unesco (2003):

[...] documentos arquivísticos exclusivamente em formato digital, como textos, bases de dados, planilhas, mensagens eletrônicas, imagens fixas ou em movimento, gravações sonoras, material gráfico, **sítios da internet**, dentre muitos outros formatos e apresentações possíveis de um vasto repertório de diversidade crescente. (CONARQ, 2005, p. 1, grifo nosso).

Comparando o texto da Carta do Conarq (2005) com a Carta da Unesco (2003), percebe-se que a versão brasileira, substituiu a expressão “objetos digitais/materiais digitais” por “documentos arquivísticos”.

No Glossário de Documentos Arquivísticos Digitais, de 2020, o conceito de patrimônio digital é o “conjunto de objetos digitais que possuem valor suficiente para serem conservados, a fim de que possam ser consultados e utilizados no futuro”. Tal conceito está associado ao termo objeto digital, “unidade de informação em formato digital composta de uma ou mais cadeia de bits e de metadados que as identificam e descrevem suas propriedades” (CONARQ, 2020, p. 37-39). Já o documento digital é conceituado como a “informação registrada, codificada em dígitos binários, acessível e interpretável por meio de sistema computacional”, e o documento arquivístico digital configura-se como “documento digital reconhecido e tratado como um documento arquivístico” (CONARQ, 2020, p. 25).

A *Carta para a Preservação do Patrimônio Arquivístico Digital* (2005), em seu texto, aponta a dependência das instituições públicas e privadas ao ambiente digital; conseqüentemente, torna-se desafiadora a preservação digital devido à rápida obsolescência tecnológica e à ineficiência dos sistemas de informação em assegurar e garantir a preservação em longo prazo. Além de envolver uma multiplicidade de atores, não apenas ligados à tecnologia, a preservação abrange toda a sociedade com “questões administrativas, legais, políticas, econômico-financeiras e, sobretudo, de descrição dessa informação através de estruturas de metadados que viabilizem o gerenciamento da preservação digital e o acesso no futuro” (CONARQ, 2005, p. 2-3). Assim como a Carta da Unesco (2003), afirma que “preservar exige compromissos de longo prazo entre os vários segmentos da sociedade”:

[...] poderes públicos, indústria de tecnologia da informação, instituições de ensino e pesquisa, **arquivos e bibliotecas nacionais** e demais organizações públicas e privadas. Reconhecida a instabilidade da informação arquivística digital, é necessário o estabelecimento de políticas públicas, diretrizes, programas e projetos específicos, legislação, metodologias, normas, padrões e protocolos que minimizem os efeitos da fragilidade e da obsolescência de hardware, software e formatos e que assegurem, ao longo do tempo, a autenticidade, a integridade, o acesso contínuo e o uso pleno da informação a todos os segmentos da sociedade. (CONARQ, 2005, p. 3, grifo nosso).

No teor da carta do Conarq (2005), é salientada a necessidade de elaborar estratégias e políticas no âmbito da gestão de documentos digitais, promovendo a criação de uma infraestrutura tecnológica nas instituições, apoiadas por recursos humanos especializados para a preservação do patrimônio arquivístico digital. A cooperação entre instituições por meio de consórcios e a participação dos especialistas em projetos do governo são pontos importantes.

No entanto, é primordial seguir as normativas dos organismos internacionais, como o Conselho Internacional de Arquivos e a Unesco, com a adoção de normas, padrões, protocolos e metadados que sejam reconhecidos pela comunidade arquivística, visando também garantir a segurança da informação. Em suma, é necessária uma agenda de promoção de conhecimento, pois o trabalho de preservar em meio digital é contínuo, e o compartilhamento de informações entre profissionais, pesquisadores e instituições torna-se fundamental em longo prazo (CONARQ, 2005, p. 4-5).

Rondinelli (2013) menciona a visão de Yeo (2008)⁵³, que busca “ampliar a definição de documento arquivístico, como forma de libertá-lo de uma perspectiva restritiva de evidência ou informação” (YEO, 2008 *apud* RONDINELLI, 2013, p. 197). Rondinelli (2013) sintetiza o conceito de documento arquivístico por intermédio de quadros e aponta as convergências entre os autores clássicos e contemporâneos a respeito do conceito, que são a natureza jurídica e a organicidade (RONDINELLI, 2013, p. 171-200).

Dessa maneira, um *website* ou páginas *web* podem ser entendidos “como um documento arquivístico, mas também como uma ferramenta de gestão corporativa, ou mesmo uma manifestação de cultura contemporânea”, segundo o exemplo oferecido por Yeo (2008), e ainda diz que “o mundo dos documentos arquivísticos é diversificado e multifacetado” (YEO, 2008, p. 121-143 *apud* RONDINELLI, 2013, p. 197).

Em relação aos documentos digitais, Rondinelli (2013) diz que estes podem ser classificados como estáticos, “não permitindo alteração de forma e conteúdo, além das oferecidas pelas tecnologias”, ou interativos, que “permitem alteração de forma e/ou conteúdo por meio de regras fixas ou variáveis”. A autora subdivide os documentos interativos em “documento digital interativo não dinâmico, fornecendo o exemplo de catálogos de venda on-line” e o “documento digital interativo dinâmico”, como, por exemplo, os “serviços de previsão do tempo ou cotação de moeda cujos conteúdos estão sempre mudando” (RONDINELLI, 2013, p. 248).

Para ela, o “documento arquivístico digital perpassa pelo [sic] conceito de documento⁵⁴, o qual, segundo a diplomática, pressupõe as características de forma física e conteúdo estável” (RONDINELLI, 2013, p. 259), acrescentando-se mais quatro características, como “relação orgânica, contexto identificável, ação e cinco pessoas (autor, redator, destinatário, originador e produtor) que concedem ao documento digital estático e não

⁵³ YEO, G. Concepts of record (2): prototypes and boundary objects. *The American Archivist*, v. 71, n. 1, p. 118-143, apr. 2008. DOI: <https://doi.org/10.17723/aarc.71.1.p0675v40tr14q6w2>.

⁵⁴ Documento – Unidade de registro de informações, qualquer que seja o formato ou o suporte (ARQUIVO NACIONAL, 2005, p. 73).

dinâmico status arquivístico”. Segundo Rondinelli (2013), “os documentos digitais dinâmicos necessitam adquirir as características de forma fixa e conteúdo estável para se tornarem arquivísticos” (RONDINELLI, 2013, p. 260).

Por outro lado, Flores (2021), na palestra apresentada no *I Encontro Nacional de Memória do Poder Judiciário*, do Conselho Nacional de Justiça – CNJ⁵⁵, com o título “Preservação de Páginas Web e Redes Sociais em Cadeia de Custódia: identificação, seleção e arquivamento”, declara que o documento arquivístico digital caracteriza-se como “complexo e específico”. Outrossim, para sua utilização como “fonte de prova, evidência, testemunho, memória e patrimônio”, ele deve ser “mantido autêntico em sua transmissão, preservação e custódia, ou seja, dotados [sic] de componentes de identidade e integridade”. Precisa ser “confiável e possuir (completeza, forma e controles de produção) em uma Cadeia de Custódia Digital, sendo este original (nativo digital) ou um representante digital (digitalização)” (FLORES, 2021, p. 12).

Flores (2021) cita a Resolução nº 37, de 19 de dezembro de 2012, do Conarq, que aprova as diretrizes para a presunção de autenticidade de documentos arquivísticos digitais, e afirma que “a presunção de autenticidade dos documentos arquivísticos sempre fez parte do processo tradicional de avaliação desses documentos e é fortemente apoiada na análise de sua forma e de seu conteúdo”, [que] “nos documentos não digitais estão inextricavelmente ligados ao suporte – isto é, forma, conteúdo e suporte são inseparáveis” (CONARQ, 2012, p. 1 *apud* FLORES, 2021).

Além disso, essa presunção baseia-se na confirmação da existência de uma cadeia de custódia ininterrupta⁵⁶, desde o momento da produção do documento até a sua transferência para a instituição arquivística responsável pela sua preservação no longo prazo. Caso essa cadeia de custódia seja interrompida, o tempo em que os documentos não estiveram sob a proteção do seu produtor ou sucessor pode causar muitas dúvidas sobre a sua autenticidade. (CONARQ, 2012, p. 1 *apud* FLORES, 2021, informação verbal).

Peter Hirtle (2000), no artigo “*Archival Authenticity in a Digital Age*” (Autenticidade Arquivística na Era Digital), afirma que “um verdadeiro arquivo é um corpo orgânico de evidências contextualmente baseado, não uma coleção de informações diversas” e sua

⁵⁵ FLORES, D. Preservação de páginas web e redes sociais em cadeia de custódia: identificação, seleção e arquivamento. In: ENCONTRO NACIONAL DE MEMÓRIA DO PODER JUDICIÁRIO, 1., 18 maio 2021, Brasília, DF. 1 vídeo (160 min). Publicado pelo canal Conselho Nacional de Justiça (CNJ). Disponível em: <https://www.youtube.com/watch?v=RkR2RjaSjC0>. Acesso em: 01 jun. 2021.

⁵⁶ Cadeia de custódia ininterrupta: linha contínua de custodiadores de documentos arquivísticos (desde o seu produtor até o seu legítimo sucessor) pela qual se assegura que esses documentos são os mesmos desde o início, não sofreram nenhum processo de alteração e, portanto, são autênticos (CONARQ, 2012, p. 1.)

“existência está intrinsecamente ligada a uma cadeia de custódia ininterrupta desde a criação até sua transferência aos arquivos” (HIRTLE, 2000, p. 10-12, tradução nossa).

Refletindo sobre os conteúdos apresentados por Flores (2021) e Hirtle (2000), o produtor do documento deve prover um ambiente seguro desde a produção do documento até o momento de transferência para instituição arquivística responsável pela salvaguarda e preservação em longo prazo. Observa-se que são pontos que convergem com a posição de Rondinelli (2013), da natureza jurídica e orgânica do documento arquivístico, que deverá ser “submetido aos procedimentos de gestão e preservação, preconizados pela arquivologia” (RONDINELLI, 2013, p. 267). Todavia, o caráter enigmático e dinâmico das páginas *web* torna desafiador o trabalho do arquivista no âmbito da preservação digital.

Todos os conceitos arquivísticos apresentados são importantes para a execução da presente pesquisa, assim como para a compreensão do patrimônio digital, do patrimônio arquivístico digital e do patrimônio digital salvaguardado nas bibliotecas nacionais.

5.2 PATRIMÔNIO DIGITAL NAS BIBLIOTECAS NACIONAIS

As bibliotecas nacionais destacam-se pelas suas coleções/acervos imponentes; e, tais como os arquivos nacionais, são instituições mantidas pelo Estado.

Campello (2006) mostra um novo perfil de bibliotecas nacionais a partir de debates promovidos pela *International Federation of Library Associations* (IFLA) e a Unesco, entre 1980 e 1990, especialmente pensando nos aspectos tecnológicos (CAMPELLO, 2006, p. 25). Notam-se três diferentes dimensões nas funções das bibliotecas nacionais: a função de depositária – da herança cultural de um país; a de infraestrutura – coordenando e orientando os serviços das bibliotecas públicas de um país; e a de serviço nacional abrangente – promovendo o acesso aos cidadãos (LOR; SONNEKUS, 1997, p. 5; CAMPELLO, 2006, p. 27).

A primeira função de depositária da herança cultural de um país, vinculada à preservação do patrimônio, revela a necessidade de salvaguardar um acervo representativo daquela nação, geralmente utilizando-se de mandato legal.

Conforme Larivière (2000), no “*Guidelines for legal deposit legislation*” (Legislação sobre depósito legal: diretrizes), documento financiado pela Unesco/IFLA, “a legislação de depósito legal é claramente de interesse público, garantindo a aquisição, registro, preservação e disponibilidade do patrimônio de obras publicadas de um país” e “uma coleção nacional é sem dúvida um dos principais componentes da política cultural de cada país e deve também

ser vista como pilar da política nacional em matéria de liberdade de expressão e acesso à informação”, sendo o papel do depósito legal:

[...] assegurar a criação de uma coleção nacional de material publicado em diferentes formatos. Deve também encorajar a compilação e publicação de uma bibliografia nacional para assegurar o controle bibliográfico abrangente. Finalmente, uma legislação eficaz de depósito legal assegurando que os cidadãos e investigadores no país e no estrangeiro tenham acesso a coleções de materiais publicados, a nível nacional. (LARIVIÈRE, 2000, p. 4, tradução nossa).

O autor destaca que as “publicações eletrônicas dinâmicas em linha (ou seja, aquelas que são atualizadas de forma contínua) devem também estar sujeitas ao depósito legal”. Diz que “a primeira versão de uma publicação eletrônica dinâmica em linha deve ser sempre depositada”, e as bases de dados sem algum tipo de organização não devem ser consideradas como objeto de depósito legal (LARIVIÈRE, 2000, p. 46).

Segundo a “*IFLA Statement on Legal Deposit*” (Declaração da IFLA sobre Depósito Legal), esse depósito é “fundamental para a preservação e o acesso ao patrimônio documental de uma nação, independentemente do formato ou tecnologia”. As diretrizes da IFLA definem depósito legal como “uma obrigação legal que exige que qualquer organização, comercial ou pública, e qualquer indivíduo que produza qualquer tipo de documentação em múltiplas cópias, seja obrigado a depositar uma ou mais cópias em uma instituição nacional reconhecida” (IFLA, 2011, 2013, tradução nossa).

No documento, é mencionado que, muitas vezes, essa obrigação é exigida por meio de legislação de abrangência nacional, mas pode ser também realizada de maneira voluntária. Ressalta que “os países devem contar com um sistema robusto e eficiente de depósito legal, procurando estreitar a cooperação entre os responsáveis pelos depósitos (entrega do material), isto é, os editores ou criadores de conteúdos e os designados para guardar e preservar o patrimônio”, que, no caso das legislações de depósito legal, geralmente são missões das bibliotecas nacionais, podendo ser compartilhada com outras instituições de acordo com o país (IFLA, 2011, 2013, tradução nossa).

Para Larivière (2000), o futuro do depósito legal está profundamente ligado ao ambiente digital, cabendo questionar a viabilidade dele, visto que:

[...] como a própria natureza do material muda, os documentos já não são ‘publicados’, mas ‘pesquisáveis’ em redes. As ‘cópias’ já não são ‘vendidas’; ao invés disso, os utilizadores devem “pagar assinaturas” para terem “acesso” ao material. A maioria dos sistemas de depósito legal tem sido capaz de integrar publicações eletrônicas off-line sem grandes problemas, principalmente porque têm um meio físico e realizam uma gestão quase análoga ao do material impresso. Mas o

ambiente de acesso em linha, e mais especificamente o caso das publicações eletrônicas dinâmicas, está a criar sérios problemas jurídicos de depósito legal. Claramente, o problema mais grave é o fato da informação ser constantemente atualizada em tempo real, e consequentemente os dados e a informação desaparecem automaticamente em poucos segundos e deixam de existir como documentos históricos. O maior desafio para o depósito legal num ambiente eletrônico ou digital é a questão da preservação digital para as gerações futuras. (LARIVIÈRE, 2000, p. 46, tradução nossa).

O depósito dos documentos digitais destaca-se como promissor e, ao mesmo tempo, desafiador, sendo alguns contratempos do meio digital relatados por especialistas britânicos no livro *Electronic Legal Deposit*⁵⁷: *shaping the library collections of the future* (Depósito Legal eletrônico: moldando as coleções da biblioteca do futuro). Lançado em 2020, a obra foi dividida em duas partes: a primeira apresenta exemplos de instituições e a implementação de depósito em meio digital; e a segunda parte trata sobre os usuários e seus contextos, sendo o capítulo 8, “*Giving with one click, taking with the other: electronic legal deposit, web archives and researcher access*”, dedicado ao depósito legal eletrônico, arquivos da *web* e acesso aos pesquisadores. A publicação fornece informações sobre os impactos sociais e institucionais do depósito legal eletrônico no século XXI e, ao mesmo tempo, que esse novo *modus operandi* em meio digital poderá afetar o futuro das coleções e de seus usuários.

Embora as bibliotecas nacionais utilizem tecnologias para o cumprimento de sua missão, “as regras de disponibilização de publicações em geral permanecem nos moldes analógicos, ou seja, a consulta aos documentos em locais pré-estabelecidos e não em qualquer parte” (PINTO, 2011, p. 90). Todavia, o acesso *on-line* parece ser uma exceção; e, conforme Pinto (2011, p. 90), é necessária a reflexão “sobre as formas de disseminação das publicações digitais e como e se poderão ser consultadas remotamente, já que as bibliotecas devem respeitar os direitos autorais vigentes nos países”.

A *Carta sobre a Preservação do Patrimônio Digital* da Unesco, no art. 8º – “Proteção do patrimônio digital” – menciona que “os Estados-Membros precisam de estruturas jurídicas e institucionais adequadas para garantir a proteção de seu patrimônio digital” e deve-se “assegurar que a legislação arquivística, bem como o depósito legal ou voluntário em bibliotecas, arquivos, museus ou outras instituições públicas para a preservação do patrimônio digital, deve ser elemento essencial da política nacional de preservação”, assim como o “acesso aos materiais depositados do patrimônio digital deve ser assegurado dentro de limites razoáveis, sem prejuízo da sua exploração”. Tudo isso para “evitar a manipulação ou

⁵⁷ GOODING, P.; TERRAS, M. *Electronic legal deposit: shaping the library collections of the future*. 2nd. ed. London: Facet Publishing, 2020. 256 p.

modificação ao patrimônio digital, por isso é muito importante um quadro jurídico e técnico para proteção da autenticidade”. É preciso também realizar a manutenção dos conteúdos e, igualmente, dos arquivos e a documentação. Essas são medidas necessárias para assegurar que um objeto digital autêntico seja preservado (UNESCO, c2004).

Observa-se que há uma necessidade de “soluções especiais e arranjos práticos para a preservação a longo prazo com apoio de *softwares* e *hardwares*, ou outras tecnologias de acesso necessárias para tornar as publicações digitais compreensíveis” (IFLA, 2011, 2013). Além disso, questões de direito autoral, de acesso, de espaço de armazenamento e de orçamento são desafios no cumprimento da missão institucional de guardar e preservar o patrimônio nacional, independentemente da instituição ou do suporte do material.

6 PRESERVAÇÃO DO PATRIMÔNIO

Antes de aprofundar a temática preservação digital, faz-se necessário entender o conceito de preservação, pelo fato de o termo ser parte integrante e fundamental das missões das instituições de memória.

O “*Glossary of World Heritage Terms*”⁵⁸, produzido pela Unesco em 1997, frisa que o documento “*The World Heritage Convention and the Operational Guidelines*” não define o conceito de preservação, que ele é usado como equivalente à conservação, salvaguarda e a diretrizes operacionais de proteção (UNESCO, 1997, tradução nossa).

Vaillant Callol (2013, p. 18) destaca “algumas controvérsias em torno do conceito e menciona que há abordagens distintas entre instituições”, entre elas, a *American Institute for Conservation* (AIC) que define preservação como “proteção dos bens culturais através de atividades que minimizem a deterioração e danos químicos e físicos e que evitem a perda de conteúdo informativo” (FAIC, 1996).

No documento “Memória do mundo: diretrizes para a salvaguarda do patrimônio documental”, publicado em 2002 pela Unesco, a organização apresenta preservação como “a soma das medidas necessárias para assegurar a acessibilidade permanente – para sempre – do patrimônio documental”, frisando que a definição se refere ao contexto do Programa Memória do Mundo (UNESCO, 2002). Pode-se inferir que a preservação é “uma consciência ou uma mentalidade política, seja individual, coletiva ou institucional para proteger e salvaguardar o patrimônio, proporcionando condições adequadas de uso e acesso” (FUNDAÇÃO PRÓ-MEMÓRIA DE SÃO CARLOS, c2021).

Para Cunha e Cavalcanti (2008), a preservação abrange as “medidas empreendidas com a finalidade de proteger, cuidar, manter e reparar ou restaurar os documentos”. É também considerado como um “conjunto de medidas jurídicas, administrativas, urbanísticas, arquitetônicas ou de outra natureza técnica que visa resguardar uma edificação, sítios urbanos, obras escultóricas em locais públicos [...]” (CUNHA; CAVALCANTI, 2008, p. 290). De acordo com a AIC, o principal objetivo da preservação é prolongar a existência dos bens culturais (FAIC, 1996).

Já o *Dicionário Brasileiro de Terminologia Arquivística* (ARQUIVO NACIONAL, 2005, p. 135) destaca que a preservação visa à “prevenção da deterioração e danos ao

⁵⁸ Preservation – The World Heritage Convention and the Operational Guidelines do not specifically define preservation. Preservation is used interchangeably with conservation, safeguarding and protection in the Operational Guidelines. (UNESCO, 1997).

documento, por meio de adequado controle ambiental”, definição profundamente ligada ao termo conservação – “um conjunto de ações estabilizadoras que visam desacelerar o processo de degradação de documentos ou objetos, por meio de controle ambiental e de tratamentos específicos (higienização, reparos e acondicionamento)” (CASSARES, 2000, p. 12). Verifica-se que ambas as definições estão voltadas para o controle ambiental, ou seja, preocupados com o ecossistema no qual o documento está salvaguardado.

Para Cassares (2000), a preservação também é um conjunto de medidas, no entanto, a autora acrescenta o termo “estratégias” à sua definição, podendo estas serem “de ordem administrativa, política e operacional que contribuem direta ou indiretamente para a preservação da integridade dos materiais” (CASSARES, 2000, p. 12).

No âmbito dos arquivos, no artigo “*Archival preservation practice in a nationwide context*” (Práticas de preservação de arquivos em um contexto nacional), “a preservação dos arquivos é a aquisição, organização, e distribuição de recursos (humano, físico, monetário) para assegurar uma proteção adequada da informação de valor duradouro para acesso pelas gerações presentes e futuras”. Além disso, a “preservação de arquivos engloba o planejamento e a implementação de políticas, procedimentos e processos, que em conjunto, previnem uma maior deterioração ou renovam a usabilidade de grupos seletos de materiais” (CONWAY, 1990, p. 206-207, tradução nossa).

Paul Conway (1996) também produziu o trabalho “*Preservation in the Digital World*”, que foi traduzido com o título “Preservação no universo digital” no ano de 2001, pelo “Projeto de Conservação Preventiva em Bibliotecas e Arquivos – CPBA⁵⁹”. Nele é ressaltado que “muitos dos princípios centrais da preservação, desenvolvidos no universo analógico, podem ser transportados para o universo digital, de forma a dar prosseguimento ao essencial desempenho da administração e dos serviços” (CONWAY, 2001, p. 12). Para Conway (2001), a preservação preocupa-se com os objetos e com as evidências. O autor consubstancia seu discurso quando cita a expressão utilizada por “Michel Buckland (1991) ‘informação como coisa’ – embutida no conteúdo intelectual dos objetos, para serem utilizadas com distintos propósitos” (CONWAY, 2001, p. 14). Ele apresenta nove conceitos, os quais divide em dois grupos: o primeiro de “contexto para ação de preservação”, composto por custódia, importância social, estrutura e cooperação; e o segundo, que chama de “prioridades para ação

⁵⁹O CPBA é uma experiência de cooperação entre instituições brasileiras e a organização norte-americana Commission on Preservation and Access, atualmente incorporada ao CLIR – Council on Library and Information Resources (Conselho de Recursos em Biblioteconomia e Informação).

de preservação” longevidade, escolha, qualidade, integridade e acesso (CONWAY, 2001, p. 23 e 24).

Grácio (2012) consolidou os conceitos de Conway (1996) expondo as semelhanças da preservação tradicional com a digital, conforme a figura 8.

Figura 8 – Contexto de ação da preservação

Conceito	Preservação tradicional	Preservação digital
Custódia	Faz referência à atividade da biblioteca de tratar o documento, desde sua seleção até o seu descarte.	Faz referência à necessidade da instituição de ter o compromisso de migrar os dados digitais para novas tecnologias.
Importância Social	As atividades de preservação têm a missão de atender à necessidade da sociedade de preservar a história e a memória, através de documentos, pessoas e instituições.	Está mais relacionada ao acesso aos serviços oferecidos às comunidades acadêmicas, eruditas e públicas.
Estrutura	Necessidade de uma estrutura organizacional que propicie recursos e estrutura para as atividades permanentes de preservação.	Está voltada para um processo de gerenciamento de riscos, com especialistas de outras áreas, como tecnologia.
Cooperação	Cooperação entre instituições com o objetivo de dividir os investimentos e, dessa forma, selecionar e preservar os documentos mais valiosos.	Também é necessária pela própria característica do ambiente digital.

Fonte: Grácio (2012, p. 63) e Conway (1996, 2001).

Os cinco elementos que compõem as “prioridades para ação de preservação” (CONWAY, 1996) estão ilustrados na figura 9:

Figura 9 – Prioridades para ação de preservação

Conceito	Preservação tradicional	Preservação digital
Integridade	Trata da integridade física e intelectual; a física se refere ao suporte e ao documento e é importante nos laboratórios de conservação; a intelectual se preocupa com a autenticidade do documento.	A integridade física tem menos relação com o suporte; na integridade intelectual também existe a preocupação com a autenticidade do documento digital.
Acesso	A preservação durante muito tempo se preocupou essencialmente em guardar o documento em um lugar seguro, tratando a preservação e o acesso de formas excludentes.	No universo digital, o acesso é uma das ideias centrais, e a capacidade de acesso à informação digital preservada torna-se o resultado final da ação de preservação.
Longevidade	A ideia é expandir a capacidade de utilização dos documentos, estabilizando as estruturas organizacionais e minimizando as possibilidades de deterioração física causadas por fatores internos e externos.	No universo digital, a preocupação é mais com o conteúdo do que com o suporte, e sua longevidade depende mais da expectativa de vida dos sistemas de acesso e das atividades de migração para os novos ambientes.
Escolha	Está relacionada à seleção do que deve ser preservado com base na definição de valores. É reconhecidamente uma das atividades mais difíceis de preservação.	É um avançado processo ligado ao uso efetivo das informações armazenadas em meio digital.
Qualidade	Necessidade de maximizar a qualidade nas atividades de preservação, com critérios de qualidade, normas, diretrizes, procedimentos e baixa tolerância a erros.	O objetivo é “assegurar, de modo mais amplo e tecnicamente possível, o conteúdo intelectual e visual, para então apresentá-lo aos usuários de maneira mais adequada às suas necessidades”.

Fonte: Grácio (2012, p. 63) e Conway (1996, 2001).

Para Conway (2001), no ambiente digital, o conceito ‘preservação’, relacionado com o gerenciamento e aliado à tecnologia, está assumindo outros significados, que o autor considera “possibilitar o uso” e “proteger o item original”, ligado ao crescimento da digitalização dos documentos, acarretando “manter os objetos digitais” acessíveis no futuro. Conway (2001) acredita que o grande desafio é criar ambientes organizacionais apropriados para ação de preservar.

É essencial compreender as diferentes perspectivas do conceito de preservação para que as instituições e os profissionais fomentem, em conjunto com a sociedade, uma

consciência da importância de salvaguardar o patrimônio, com medidas, estratégias e ações que garantam o uso e o acesso às futuras gerações, tanto no contexto analógico como no digital.

6.1 PRESERVAÇÃO DIGITAL – CONCEITOS E PERSPECTIVAS

Destacam-se alguns atores no cenário da preservação digital, que conceituam o termo, dependendo do contexto, como patrimônio, arquivístico, bibliotecário ou focado especificamente no digital.

Para profissionais da ciência da informação, [a preservação digital] refere-se à infraestrutura e ao comprometimento institucional necessário para proteger a informação representada digitalmente enquanto para os especialistas da ciência da computação seria uma maneira de atenuar a obsolescência tecnológica e aumentar a memória humana. (MÁRDERO ARELLANO, 2008, p. 42).

Para a Unesco, o conceito de preservação digital⁶⁰ “consiste nos processos que visam garantir a acessibilidade continuada dos materiais digitais, e envolve encontrar maneiras de representar o que foi originalmente apresentado aos usuários por uma combinação de ferramentas de *software* e *hardware* agindo sobre os dados” (UNESCO, c2021a, tradução nossa). O conceito apresentado pela Unesco está ligado ao patrimônio, de maneira ampla, e abarca diferentes instituições.

O tema preservação digital começou a ser discutido no Congresso Internacional de Arquivos (*International Congress on Archives*) de 1964, em Paris (CUNHA; LIMA, 2007; SILVA, 2019). De acordo com Silva (2019, p. 25), o “relatório do Professor Robert Henri Bautier para o Conselho Internacional de Arquivos – ICA, apresentado na *Conférence Internationale de la Table Ronde des Archives* – CITRA em 1971, em Bonn, Alemanha, [...] [é considerado] como ponto de partida para discussões sobre este tópico”.

O Projeto InterPARES “*The International Research on Permanent Authentic Records in Electronic Systems*”, iniciativa de pesquisa internacional com a participação de profissionais e instituições de diferentes áreas, foi fundado no ano de 1999, “com objetivo de desenvolver conhecimentos teóricos e metodológicos, para a preservação de longo prazo de registros autênticos criados e / ou mantidos em formato digital”. O projeto também fornece subsídios para produção de: normas, políticas, estratégias e planos de ação, a fim de garantir a longevidade do material, assim como a capacidade de seus usuários de confiar em sua

⁶⁰ Digital preservation consists of the processes aimed at ensuring the continued accessibility of digital materials. To do this involves finding ways to re-present what was originally presented to users by a combination of software and hardware tools acting on data. (UNESCO, [c2021a]).

autenticidade (INTERPARES, [c2021a], [c2021b]). Ainda que o projeto conte com profissionais de áreas diversas, o foco principal são os arquivistas e os documentos arquivísticos.

O projeto foi dividido em fases ao longo desses anos, e vários trabalhos foram publicados como fruto das pesquisas. No âmbito da terminologia, destacam-se o dicionário “*The InterPARES 2 Project Dictionary*” e o glossário “*The InterPARES 2 Project Glossary*”, ambos desenvolvidos na fase 2 (2002-2007), que apresentam, da mesma maneira, o conceito de preservação digital⁶¹ como “o processo específico de manutenção de materiais digitais através de diferentes gerações de tecnologia ao longo do tempo, independentemente do local onde residem” (INTERPARES 2, 2002, tradução nossa). Luciana Duranti, diretora do Projeto InterPARES, propõe a seguinte definição de preservação digital:

[...] conjunto de princípios, políticas, regras e estratégias destinadas a prolongar a existência de um objeto digital, mantendo-o em condições adequadas para utilização, quer no seu formato original, quer num formato mais persistente, protegendo ao mesmo tempo a identidade e integridade do objeto, ou seja, a sua autenticidade. (DURANTI, 2010, p. 157, tradução nossa).

Nota-se que a conceituação de Duranti (2010) é mais abrangente e difere das propostas no dicionário e no glossário do InterPARES em 2002.

O *Dicionário de Terminologia Arquivística* da Sociedade Americana de Arquivistas, “*Dictionary of Archives Terminology for Society of American Archivists*”, mostra a preservação digital⁶² como “a gestão e a proteção da informação digital para assegurar a autenticidade, integridade, confiabilidade e acessibilidade a longo prazo”. Além disso, em notas, dimensiona que, além da gestão, a preservação digital utiliza-se de estratégias, melhores práticas, padrões e políticas para garantir o acesso contínuo das informações digitais (SOCIETY OF AMERICAN ARCHIVISTS, 2021b).

A definição de preservação digital, apresentada pelo Projeto InterPARES destaca um cunho tecnológico, que difere, em certo ponto, da fornecida pela *Society of American Archivists* e daquela de Luciana Duranti com elementos arquivísticos, mas que, em conjunto, complementam-se.

Vale ressaltar a visão dos bibliotecários em torno do tema, com a *American Library Association* – ALA, que, em sua conferência anual em 2007, forneceu três aspectos da

⁶¹Digital preservation v., The specific process of maintaining digital materials during and across different generations of technology over time, irrespective where they reside. [Computer and Information Sciences] (INTERPARES 2, 2002).

⁶²Digital Preservation – the management and protection of digital information to ensure authenticity, integrity, reliability, and long-term accessibility. Disponível em: <https://dictionary.archivists.org/entry/digital-preservation.html>.

definição de preservação digital: primeiro, uma forma breve – “que seria a combinação de políticas, estratégias e ações que assegurem o acesso a conteúdos digitais ao longo do tempo”; segundo, uma definição média, somadas aos elementos da primeira – “[...] independentemente dos desafios do fracasso dos meios de comunicação e da mudança tecnológica, assegurando a autenticidade dos conteúdos migrados ao longo do tempo” (ALA, 2008, tradução nossa); e, por último, a definição longa:

A preservação digital combina políticas, estratégias e ações para assegurar a apresentação precisa dos conteúdos autênticos ao longo do tempo, independentemente dos desafios das falhas dos meios de comunicação e das mudanças tecnológicas. A preservação digital aplica-se tanto a conteúdos nascidos digitais como a conteúdos digitalizados. (ALA, 2008, p. 1, tradução nossa)⁶³.

A ALA afirma que, mediante políticas de preservação digital, as instituições confirmam, por meio de documento institucional, o compromisso de preservar conteúdos digitais para uso futuro; especificando os formatos de arquivos, assegurando o cumprimento de normas e utilizando de melhores práticas para a gestão responsável da informação digital. Ademais, as estratégias e as ações de preservação digital abordam a criação, a integridade e a manutenção desses conteúdos (ALA, 2008). A associação enseja com o documento contribuir com a comunidade de profissionais e bibliotecas no intuito de refletirem sobre a preservação digital.

A *Digital Preservation Coalition* – DPC, organização sem fins lucrativos com sede na Inglaterra, tem como missão capacitar os seus membros e conscientizar sobre os desafios estratégicos, culturais e tecnológicos da preservação digital, e apresenta a temática sobre três aspectos temporais (DPC, 2021, p. 1, tradução nossa):

- curto prazo – com acesso a materiais digitais quer por um período de tempo definido enquanto se prevê a sua utilização, mas que não se estende para além do futuro previsível e/ou até se tornar inacessível devido a mudanças na tecnologia.
- médio prazo – com acesso contínuo a materiais digitais para além das mudanças tecnológicas durante um período de tempo definido, mas não indefinidamente.
- longo prazo – com acesso contínuo a materiais digitais, ou pelo menos à informação neles contida, por tempo indeterminado.⁶⁴

⁶³ Long Definition Digital preservation combines policies, strategies and actions to ensure the accurate rendering of authenticated content over time, regardless of the challenges of media failure and technological change. Digital preservation applies to both born digital and reformatted content (ALA, 2008).

⁶⁴ Short-term preservation – Access to digital materials either for a defined period of time while use is predicted but which does not extend beyond the foreseeable future and/or until it becomes inaccessible because of changes in technology. Medium-term preservation – Continued access to digital materials beyond changes in technology for a defined period of time but not indefinitely. Long-term preservation – Continued access to digital materials, or at least to the information contained in them, indefinitely (DPC, 2021).

Segundo a DCP, a definição de preservação digital é muito ampla, no entanto, a organização considera como definição “todas as ações necessárias para manter o acesso a materiais digitais para além dos limites da falha dos meios de comunicação ou da mudança tecnológica e organizacional” (DPC, 2021, p. 1, tradução nossa).

Outra organização internacional ligada ao ambiente digital, a *National Digital Stewardship Alliance* – NDSA, configura-se como uma “organização que fornece proteção, experiência e apoio para a preservação do patrimônio digital”. Fundada em julho de 2010, pelo Programa Nacional de Infraestrutura e Preservação de Informação Digital da *Library of Congress*, em 2013, desenvolveu o glossário de termos, em que define preservação digital como uma “série de atividades, políticas, estratégias e ações gerenciadas para assegurar a reprodução exata dos conteúdos digitais durante o tempo necessário, independentemente dos desafios, do fracasso dos meios de comunicação e das mudanças tecnológica⁶⁵” (NATIONAL DIGITAL STEWARDSHIP ALLIANCE, 2021).

Thibodeau (2002), no artigo “*Overview of Technological Approaches to Digital Preservation and Challenges in Coming Years*”, em português, a “Visão geral das abordagens tecnológicas da preservação digital e dos desafios nos próximos anos”, ressalta que “a preservação de objetos digitais envolve uma variedade de desafios, incluindo questões políticas, papéis institucionais e relações, questões jurídicas, direito de propriedade intelectual e metadados” (THIBODEAU, 2002, p. 4, tradução nossa). O autor afirma que, para enfrentar qualquer desafio relacionado com a preservação digital, faz-se necessário entender o que são os objetos digitais, e aponta como dificuldade que “muitas formas de informação digital não podem ser expressas em mídia impressa ou analógica tradicional; por exemplo, páginas da *web* interativas”, bem como “declara que não há uma única definição ou modelo de documento digital que seja válido em todos os casos” (THIBODEAU, 2002, p. 5). Pode-se perceber a dificuldade de conceptualização no âmbito digital com as definições propostas por Rondinelli (2013) mencionadas na seção de patrimônio arquivístico digital.

Dessa forma, a variedade e a complexidade dos objetos digitais dão à preservação digital certo grau de dificuldade, pois todas as variantes deverão ser avaliadas. É importante destacar que Thibodeau (2002) inseriu os objetos digitais em três classes: primeiro o que o autor chamou de “**objetos físicos** – é simplesmente uma inscrição de sinais sobre um meio”; segundo os “**objetos lógicos** – unidades processáveis, ou seja, de acordo com a lógica de algum *software* de aplicação; em terceiro os “**objetos conceituais** – o que lidamos no mundo

⁶⁵ NATIONAL DIGITAL STEWARDSHIP ALLIANCE. *Glossary*. 2021. Disponível em: <https://ndsa.org/glossary/>. Acesso em: 08 jun. 2021.

real, isto é, uma entidade que reconhecemos como uma unidade de informação significativa” (THIBODEAU, 2002, p. 6-8, grifo nosso).

Além das três classes de objetos preconizados por Thibodeau (2002), somou-se o quarto elemento ou “grupo de elementos essenciais que contêm a mensagem, finalidade ou características para as quais foi decidido preservar o material”. Este foi utilizado para compreender os objetos digitais no “*Guidelines for the preservation of digital heritage*”, produzido pela *National Library of Australia* (2003, p. 38).

Moore (2008) menciona, no artigo “*Towards a Theory of Digital Preservation*”, que o conceito de preservação pode ser caracterizado como comunicação com o futuro, e fala da possibilidade de uma teoria da preservação digital, com base na definição do conjunto mínimo de processos de preservação (MOORE, 2008, p. 64). Ele afirma que é preciso enviar para o futuro não só a informação, mas também a descrição do ambiente utilizado para gerir e ler essas informações, e a teoria da preservação faz afirmações sobre a capacidade de manter o contexto, a disposição e a gestão de registros de informação, bem como o contexto de informação (políticas de gestão e procedimentos de preservação) do ambiente de preservação (MOORE, 2008, p. 71). Muito embora o foco no objeto físico tenha se deslocado para o objeto conceitual inerente à preservação digital, nunca se deve esquecer que os objetos digitais não podem existir sem a preservação adequada do suporte (NATIONAL LIBRARY OF AUSTRALIA, 2003, p. 39).

Infere-se que a preservação digital agrega novos desafios às instituições de patrimônio, “com planejamento, alocação de recursos e métodos, aliadas à tecnologia para que a informação digital de valor contínuo permaneça acessível” (HEDSTROM, 1998, p. 190).

As definições demonstradas até então são provenientes de atores do cenário internacional, que vêm, ao longo dos anos, estabelecendo um trabalho em torno do assunto, na salvaguarda do patrimônio, nas áreas de arquivologia, biblioteconomia e/ou tecnologia. No cenário nacional, ressaltam-se alguns atores relevantes, no entanto, para aprofundamento da pesquisa, eles serão combinados às referências internacionais.

A Rede Cariniana é um ator importante no campo da pesquisa brasileira. Traz como definição em seu glossário o termo “preservação digital de longo prazo”, como:

O ato de manter informação de forma correta e independente por longo prazo. Informação armazenada independentemente, que possui documentação suficiente para permitir que seja interpretada e usada por uma comunidade sem necessidade de recursos especiais obsoletos. Requer procedimentos específicos e técnicas apropriadas para cada tipo de formato e mídia. Com ela pretende-se garantir a inalterabilidade dos registros digitais.

É o conjunto de atividades ou processos responsáveis por garantir o acesso contínuo a longo prazo à informação existente em formatos digitais. A preservação digital consiste na capacidade de garantir que a informação digital de valor contínuo permaneça acessível e com qualidades de autenticidade suficientes para que possa ser interpretada no futuro ao se recorrer a uma plataforma tecnológica diferente da utilizada no momento da sua criação. Sinônimo: gerenciamento do ciclo de vida de dados digitais. (REDE CARINIANA, [2021], p. 1).

A Rede Cariniana considera o aspecto temporal de longo prazo, diferentemente da DCP, que também divide em curto e médio prazo. O Coordenador da Rede Cariniana, Miguel Márdero Arellano, sintetiza em seus trabalhos o conceito de preservação digital “como um conjunto de atividades necessárias para assegurar que os objetos digitais possam ser localizados, reproduzidos, utilizados e compreendidos no futuro”, ou seja, “devem manter a integridade e a acessibilidade dos objetos digitais ao longo do tempo, com objetivo de alcançar todas as características essenciais do objeto digital: físicas, lógicas e intelectuais” (MÁRDERO ARELLANO, 2004, 2017). O autor registra que as características físicas estão centradas no suporte, “como o armazenamento em fitas magnéticas (fitas cassetes, VHS, etc.) ou discos óticos (CD-ROM, WORM, etc.)”, enquanto as lógicas estão ligadas às “tecnologias utilizadas para inserir e acessar os dados (*software* e *hardware*, por exemplo)”. No caso das características intelectuais, estas estão ligadas “aos mecanismos que garantem a integridade e autenticidade da informação nos documentos eletrônicos” (MÁRDERO ARELLANO, 2004, p. 17-18).

No Glossário de Documentos Arquivísticos Digitais, de 2020, o termo preservação digital é apresentado como “conjunto de ações gerenciais e técnicas exigidas para superar as mudanças tecnológicas e a fragilidade dos suportes, garantindo o acesso e a interpretação de documentos digitais pelo tempo que for necessário” (CONARQ, 2020, p. 39). O Dicionário de Biblioteconomia e Arquivologia (2008) conceitua o termo como “processo de armazenamento, em condições adequadas para uso, de documentos ou objetos produzidos em formato digital” (CUNHA; CAVALCANTI, 2008, p. 290).

Pode-se também considerar o conceito de preservação digital “sob os aspectos dos processos de gestão envolvidos na administração das atividades necessárias para garantir que um objeto digital possa ser acessado e utilizado no futuro, a partir das TICs existentes na época e com garantias de sua autenticidade e integridade” (GRÁCIO, 2012, p. 61).

Refletindo-se sobre as definições apresentadas, percebe-se que as definições carregam as características salientadas por Márdero Arellano (2004), assim como por Thibodeau (2002). Estes expõem as características dos objetos digitais de maneira tridimensional, sendo a única diferença que a última categoria de objeto foi atribuída como objeto intelectual (MÁRDERO ARELLANO, 2004) e objeto conceitual (THIBODEAU, 2002). Há uma diversidade de

termos utilizados: preservação de registros, de informação, documento digital, artefato, recurso de conteúdos e/ou objetos digitais.

Tal complexidade de conceptualização dos objetos digitais foi abordada nos resultados do artigo “*Objetos Digitais: em busca da precisão conceitual*”:

Frequentemente o termo “objeto digital” é usado de forma intercambiada com termos como: artefato digital, documento digital, recurso digital, material digital e arquivo de computador. Porém o conceito de objeto digital pode ser simples ou complexo dependendo das diferentes abordagens e escopo. (YAMAOKA; GAUTHIER, 2013, p. 77).

O *Consultative Committee for Space Data Systems* desenvolveu o modelo de referência OAIS (*Reference Model for an Open Archival Information System – OAIS*), cujo documento, estabelece um sistema de arquivamento de informações digitais para fins de preservação e disponibilização para uma dada comunidade (CCSDS, 2012; SALES, 2014). O modelo OAIS foi amplamente divulgado na comunidade de pesquisa em preservação digital, e refere-se a objeto digital como “um objeto composto por um conjunto de sequências de *bits*” (CCSDS, 2012, p. 12).

Além disso, Sales (2014) menciona que a *National Information Standards Organization – NISO* (2004) considera que os objetos digitais “são conceitualmente equivalentes a itens que podem ser encontrados no acervo de bibliotecas, na coleção de museus e em fundos arquivísticos” (NISO, 2004 *apud* SALES, 2014, p. 97), e “que os objetos digitais precisam conter as informações contextuais e estruturais, para posterior identificação, recuperação, preservação e, precipuamente, que seu conteúdo e estrutura possam ser interpretados e acessados” (SALES, 2014, p. 97).

Segundo Sales (2014, p. 96), os objetos digitais podem ser “originados de duas formas: a partir da conversão de seus originais analógicos para um formato digital através de processos de digitalização ou obtidos diretamente de algum dispositivo digital”, e podem ser classificados como objetos digitais simples ou complexos (YAMAOKA; GAUTHIER, 2013, p. 77; SALES, 2014, p. 96).

Os **objetos digitais simples** – são aqueles expressos por um único arquivo ou formato, como por exemplo, um documento pdf ou uma imagem jpg. Os **objetos digitais complexos** – são documentos que incluem agregação de formatos de arquivos, assim como diversos tipos de metadados, por exemplo, um banco de dados ou uma *homepage* – que possuem metadados de preservação, de direitos, além dos metadados estruturais. (SALES, 2014, p. 97; DIGITAL CURATION CENTER, [2010], p. 2, grifo nosso)⁶⁶.

⁶⁶ DIGITAL CURATION CENTER – DCC. *Digital Curation 101*. [S. l.], [2010]. Disponível em: <https://www.dcc.ac.uk/sites/default/files/documents/DC%20101%20Curation%20and%20Preservation.pdf>. Acesso em: 20 abr. 2021.

Para a presente pesquisa, considera-se objeto digital como “unidade de informação em formato digital composta de uma ou mais cadeia de bits e de metadados que a identificam e descrevem suas propriedades”. Tal conceito está associado ao de patrimônio digital, “conjunto de objetos digitais que possuem valor suficiente para serem conservados, a fim de que possam ser consultados e utilizados no futuro” (CONARQ, 2020, p. 37-39). As páginas *web* estão classificadas como objetos digitais complexos (YAMAOKA; GAUTHIER, 2013; SALES, 2014; DIGITAL CURATION CENTER, [2010]).

Entende-se a preservação digital como um processo que envolve políticas, ações e estratégias contínuas, a fim de garantir autenticidade, integridade, confiabilidade, acesso e uso do patrimônio digital, para além das mudanças organizacionais e tecnológicas por período indeterminado (HEDSTROM, 1998; THIBODEAU, 2002; MÁRDERO ARELLANO, 2004, 2008; DURANTI, 2010; GRÁCIO, 2012; DIGITAL PRESERVATION COALITION, 2021; NATIONAL DIGITAL STEWARDSHIP ALLIANCE, 2021).

A preservação digital abarca óticas distintas, mas a missão institucional é um fator importante nesse contexto, para delimitar o escopo desta pesquisa. Grácio (2012), no livro *Preservação digital na gestão da informação: um modelo processual para as instituições de ensino superior*, define preservação digital, dividindo-a em três aspectos. (GRÁCIO, 2012, p. 200). O primeiro aspecto é organizacional, tendo como ponto importante “o objetivo institucional, que envolve uma equipe multidisciplinar, trata também das responsabilidades, dos recursos financeiros e dos atos administrativos da instituição”. O segundo são os aspectos legais, que se referem “à legislação, direito do autor e/ou mandato legal”. E o terceiro são os aspectos técnicos – “seleção, descarte; modelos, padrões e iniciativas (metadados, OAIS), suporte, infraestrutura tecnológica (*hardware* e *software*); estratégias de preservação (migração, emulação)” (GRÁCIO, 2012, p. 83-109).

Márdero Arellano (2008, p. 276), além de ressaltar as características (física, lógica e intelectual) essenciais ao objeto digital, afirma ser necessário que, no âmbito institucional, sejam seguidos alguns critérios para garantir a preservação digital. O autor os dividiu em seis categorias: “confiabilidade; responsabilidade política; sustentabilidade econômica; inclusão em repositório digital; transparência e acessibilidade a longo prazo”. Para Márdero Arellano, são os critérios institucionais

[...] que englobam as características necessárias para o funcionamento de um sistema de preservação digital, permitem determinar as práticas de preservação e as condições institucionais que afetam a forma como é enfrentada a preservação da informação em formatos digitais. (MÁRDERO ARELLANO, 2008, p. 276).

Enquanto Grácio (2012) mencionou três aspectos da preservação, Márdero Arellano (2008) citou seis categorias. A seguir, são elencados os dez mandamentos da preservação digital enunciados por Innarelli (2009, p. 39):

- 1- Manterás uma política de preservação;
- 2- Não dependerás de *hardware* específico;
- 3- Não dependerás de *software* específico;
- 4- Não confiarás em sistemas gerenciadores como única forma de acesso ao documento digital;
- 5- Migrarás seus documentos de suporte e formato periodicamente;
- 6- Replicarás os documentos em locais fisicamente separados
- 7- Não confiarás cegamente no suporte de armazenamento;
- 8- Não deixarás de fazer *backup* e cópias de segurança
- 9- Não preservarás lixo digital
- 10-Garantirás a autenticidade dos documentos digitais.

Segundo Innarelli (2009), a ideia de apresentar os dez mandamentos da preservação digital surgiu a partir de uma experiência prática e por acreditar que, dessa forma, torna-se mais fácil a apreensão dos princípios da preservação digital. Observa-se que a grande maioria dos mandamentos está ligada aos aspectos organizacionais e técnicos abordados no trabalho de Grácio (2012).

Percebe-se que a preservação em longo prazo envolve planejamento, tecnologia e orçamento, que abarcam as áreas administrativas, arquivísticas, informáticas, entre outras da instituição. Destacam-se como “três pilares fundamentais para a preservação digital – as estratégias, os sistemas informatizados e a interoperabilidade entre estratégias e sistemas” e estes pilares “devem estar em equilíbrio e em conformidade com os requisitos que são definidos primeiramente nas políticas institucionais de preservação digital” (SANTOS; FLORES, 2018, p. 34-35). Somente por meio da sincronia entre pilares e requisitos será possível “garantir o acesso, autenticidade e confiabilidade aos documentos em custódia” (SANTOS; FLORES, 2018, p. 35).

O livro *The Theory and Craft of Digital Preservation*, do bibliotecário e arquivista americano Trevor Owens, atualmente gestor da área de recursos digitais da *Library of Congress*, traz uma visão ampla sobre a preservação digital ao afirmar que “pretende ser um ponto de entrada na teoria e no ofício da preservação digital, conforme emergiu na prática⁶⁷” (OWENS, 2018, p. 2, tradução nossa).

São discutidos por Schisler (2020), no artigo “*The theory and craft of digital preservation: comentários sobre os 16 axiomas de preservação digital*”, apenas dez axiomas

⁶⁷ This book is intended as a point of entry into the theory and craft of digital preservation as it has emerged in practice (OWENS, 2018, p. 2).

(1 ao 6, 9, 11, 12 e 14) dos dezesseis propostos por Owens (2018). No entanto, abaixo serão destacados em sua totalidade⁶⁸:

1. Um repositório não é apenas um *software* (*A repository is not a piece of software*) – “um repositório é a soma de recursos financeiros, *hardware*, pessoal tempo, e a implementação contínua de políticas e planejamento para assegurar a longo prazo acesso ao conteúdo” (OWENS, 2018, p. 4). E Schisler (2020, p. 801) afirma que “a indústria no setor digital procura vender soluções para os nossos problemas, com sistemas, nuvens, *softwares* especializados” para salvaguardar dados digitais, no entanto, faz-se necessário o conjunto de recursos apropriados para realizar a preservação digital;

2. São as instituições que tornam possível a preservação (*Institutions make preservation possible*) – “[...] os principais facilitadores da preservação a longo prazo são as nossas instituições (bibliotecas, arquivos, museus, famílias, organizações religiosas, governos, etc.)” (OWENS, 2018, p. 4). É preciso que as instituições que atravessam gerações sejam fortalecidas e tenham seu papel reconhecido na missão de preservar (SCHISLER, 2020, p. 801);

3. Ferramentas de *software* podem tanto atrapalhar o processo quanto ajudar (*Tools can get in the way just as much as they can help*) – “[...] é melhor começar de forma simples e depois introduzir ferramentas que ajudem a melhorar o seu processo de trabalho, do que investir em um sistema complexo sem nenhum tipo de gestão” (OWENS, 2018, p. 5). Apesar de existir atualmente uma gama de soluções, a “preservação digital não é resolvida [apenas] com ferramentas” (SCHISLER, 2020, p. 801);

4. Nada tem sido preservado, temos somente coisas sendo preservadas (*Nothing has been preserved, there are only things being preserved*). Ou seja, “a preservação é o resultado de um trabalho contínuo das pessoas e compromissos de recursos”, pois “uma instituição que pensa em preservação digital terá que ter recursos fixos anuais alocados para a realização de um programa contínuo nesta área” (OWENS, 2018, p. 5). Para Schisler, o autor traz “um prisma interessante quando avalia a seriedade dos programas de preservação digital de uma instituição pela verba alocada anualmente para a tarefa” (SCHISLER, 2020, p. 800);

5. Acumulação ou armazenamento não é preservação (*Hoarding is not preservation*) – “[...] para realmente preservar, necessita ser capaz de torná-lo descoberto e acessível, exigindo uma abordagem clara e coerente para desenvolvimento de coleções, organização, descrição, assim como métodos para dar acesso” (OWENS, 2018, p. 5). Segundo

⁶⁸ Vale ressaltar que alguns axiomas/princípios não tiveram uma tradução literal, para que pudesse fazer sentido na língua portuguesa.

Schisler (2020, p. 800), “é importante ter pelo menos 3 cópias, 2 mídias diferentes, 1 destas cópias *offsite* e 1 cópia *off-line*”; no entanto, o armazenamento deve ser de maneira organizada e consciente, mitigando desperdício de recursos;

6. Fazer backup dos dados não é preservação digital (*Backing up data is not digital preservation*) – “[...] muitos dos aspectos envolvidos no *backup* de dados para uso atual são semelhantes aos aspectos do trabalho de preservação digital” (OWENS, 2018, p. 6), mas “a preservação digital pensa nos objetos digitais a longo prazo, para que possam continuar sendo acessíveis no futuro”, acarretando também “pensar nos formatos, migração, fixidez, e repositórios digitais seguros, entre outros” aspectos (SCHISLER, 2020, p. 800);

7. Os limites dos objetos digitais são imprecisos (*The boundaries of digital objects are fuzzy*) – Os “objetos individuais fazem referência, incorporam e usam aspectos de outros objetos como parte de sua função diária [...], ou seja, pode ser necessário dispor de um conjunto de fontes, um codec⁶⁹ de vídeo específico, por exemplo, para acessar algo no futuro” (OWENS, 2018, p. 6);

8. O objeto digital difere de acordo com o contexto de uso (*One person’s digital collection is another’s digital object [might be] another’s data set*) – “[...] A imprecisão dos limites de objetos digitais significa que o conceito de ‘item’ e ‘coleção’ é menos claro do que com item analógico”. Segundo o autor, “um *sítio web* pode ser um item num arquivo *web*, mas é também funcionalmente uma publicação em série que muda com o tempo, bem como uma coleção de *files*” (OWENS, 2018, p. 6);

9. A preservação digital consiste em fazer o melhor uso dos seus recursos para mitigar as ameaças e riscos mais prementes de preservação (*Digital preservation is about making the best use of your resources to mitigate the most pressing preservation threats and risks*) – “[...] não é algo que pode ser realizado ou concluído”, ou seja, “é um processo contínuo de compreender os riscos de perder conteúdo ou a capacidade de acessar/interagir, assim é preciso fazer uso de todos os recursos para mitigar esses riscos” (OWENS, 2018, p. 7). Schisler (2020, p. 801) considera este um axioma importante ante os desafios institucionais enfrentados: recursos tecnológicos, financeiros e humanos limitados para desenvolver um programa contínuo de preservação;

10. A resposta a quase todas as questões de preservação digital é “depende” (*The answer to nearly all digital preservation questions is “it depends”*) – “Decidir o que é importante sobre um objeto ou conjunto de objetos, depende de como será seu uso futuro”.

⁶⁹ Codec é o acrônimo de codificador/decodificador, dispositivo de *hardware* ou *software* que codifica/decodifica sinais. Disponível em: <https://pt.wikipedia.org/wiki/Codec>. Acesso em: 23 ago. 2021.

Segundo o autor, a “abordagem que uma organização irá adotar para garantir o acesso de longo prazo a uma coleção” depende também das características, do formato, da escalabilidade da coleção e, principalmente, será conforme sua missão institucional (OWENS, 2018, p. 7);

11. Já passou da hora de iniciar a tomada de ações (*It's long past time to start tacking actions*) – “[...] Existem coisas práticas e pragmáticas que todos podem e devem fazer agora para mitigar muitos dos riscos de perda mais urgentes” (OWENS, 2018, p. 7). Owens pretende com o axioma “encorajar a tomada de ação começando pelo que está à nossa frente e o que é mais simples” (SCHISLER, 2020, p. 802);

12. As definições altamente técnicas de preservação digital são cúmplices no silenciamento do passado (*Highly technical definitions of digital preservation are complicit in silencing the past*) – “Grande parte da linguagem e especificações da preservação digital evoluíram para complexos conjuntos de requisitos, que ofuscam muitas das coisas práticas” que qualquer pessoa ou organização poderiam fazer para aumentar a probabilidade de acesso ao conteúdo no futuro (OWENS, 2018, p. 7 e 8). Há um “entendimento que este campo da preservação é complexo, requer muitos recursos e acaba por afastar e intimidar as instituições com menos conhecimento, recursos humanos e materiais” a realizarem tal tarefa (SCHISLER, 2020, p. 802);

13. As possibilidades dos meios digitais suscitam a necessidade de preservação a ser enredada no desenvolvimento de coleções digitais (*The affordances of digital media prompt a need for digital preservation to be entangled in digital collection development*) – para o autor, “[...] “a política de preservação digital deve estar cada vez mais conectada e em sintonia com a política de desenvolvimento de coleções”. Isto é, as decisões sobre o que uma organização deseja coletar e preservar precisam estar claramente estipuladas para apoiar a tomada de decisão (OWENS, 2018, p. 8);

14. Tem-se que aceitar e abraçar a parcela arquivística (*Accept and embrace the archival sliver*) – “[...] o ‘digital’ faz parecer que se pode ou se deve tentar salvar tudo”. “[...] a preservação digital é antes de tudo um desafio técnico, e não social e ético, isto é, existem vários níveis de esforços que devem ser colocados à prova no futuro por diferentes tipos de materiais” (OWENS, 2018, p. 8). Será uma parte ou “a fração pequena do todo, pois através do trabalho das pessoas e instituições salvaguardando estes filetes de todos os tipos, possibilitará preservar pedaços selecionados deste universo da produção digital” (SCHISLER, 2020, p. 801);

15. A escala e as estruturas inerentes da informação digital sugerem trabalhar mais com uma pá do que com uma pinça (*The scale and inherent structures of digital information suggest working more with a shovel than with tweezers*) – deve-se aceitar o fato da impossibilidade de coletar e preservar tudo, é necessário perceber que, em muitos casos, o tempo e os recursos necessários para tomar decisões sobre coisas individuais poderiam ser usados em larga escala, ou em lotes (OWENS, 2018, p. 9);

16. Fazer a preservação digital requer pensar como um futurista (*Doing digital preservation requires thinking like a futurista*) – segundo Owens (2018), para “[...] garantir um acesso de longo prazo à informação digital, precisa-se pelo menos em algum nível, pensar e estar atento às tendências de desenvolvimento das tecnologias digitais”. Deve-se considerar a evolução tecnológica futura, com diversidade de formatos e padrões, sempre preservar com olhar no futuro, pois a preservação é contínua (OWENS, 2018, p. 9).

No artigo “Preservação da informação arquivística digital: repercussões para o patrimônio cultural”, os autores apresentam alguns conceitos importantes, como de preservação do patrimônio cultural, patrimônio digital e do documento arquivístico digital:

A própria concepção do termo “preservação” precisa ser compreendida como elemento que não visa desenvolver medidas estritamente emergenciais, quando determinado elemento patrimonial estiver na iminência de se perder ou ser destruído, mas sim, ser parte integrante de políticas consistentes quanto à salvaguarda digital. (SCHÄFER; FLORES, 2013, p. 182).

Nota-se a amplitude da preservação digital por meio dos trabalhos desses autores, abarcando diferentes níveis de conhecimento, construindo um grande ecossistema para preservar o patrimônio digital. No entanto, apenas um deles (GRÁCIO, 2012) destacou uma perspectiva jurídico-ética, mostrando o “aspecto legal” do cenário digital; enquanto a maioria dos trabalhos (CONWAY, 1996, 2001; HEDSTROM, 1998, THIBODEAU, 2002; MÁRDERO ARELLANO, 2004, 2008; MOORE, 2008; INNARELLI, 2009; GRÁCIO, 2012; SANTOS; FLORES, 2018; OWENS, 2018) concentra-se no prisma organizacional e técnico da preservação digital.

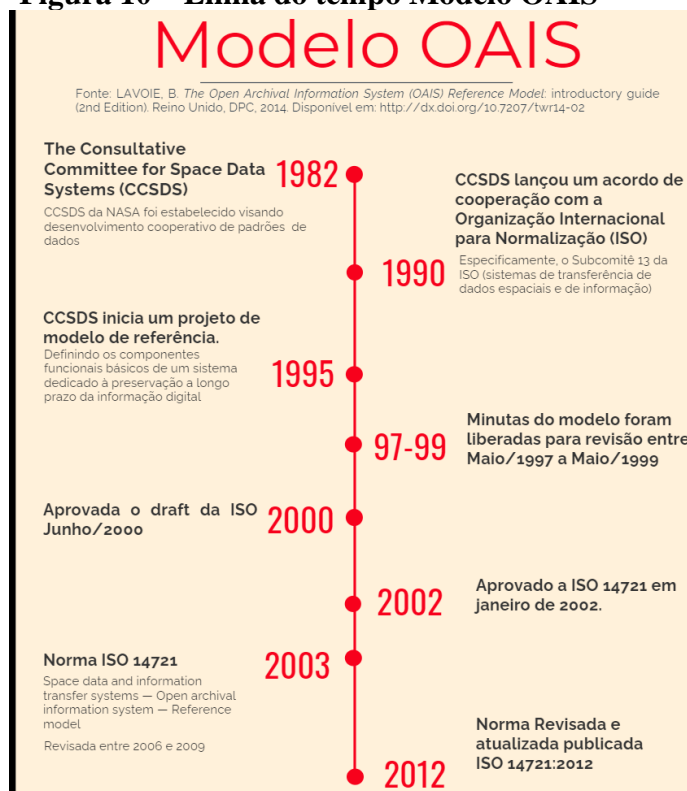
7 ETAPAS PARA ARQUIVAMENTO E PRESERVAÇÃO DA WEB

A presente seção trata de aspectos técnicos e de governança importantes para realizar o arquivamento e a preservação da *web* e servirá de apoio para a questão do acesso aos conteúdos em longo prazo, como o modelo OAIS, padrão de metadados, Norma ISO 28500:2017, e introduz o IIPC.

7.1 OPEN ARCHIVAL INFORMATION SYSTEM

O modelo de referência⁷⁰ OAIS foi desenvolvido pela *The Consultative Committee for Space Data Systems*, conhecido como Magenta Book – CCSDS 650.0-M-2. É apresentado como uma prática recomendada com intuito de prover a preservação da informação digital. A recomendação teve sua primeira versão em 2003, atualmente em vigor, e equivale à norma ISO 14721:2012.

Figura 10 – Linha do tempo Modelo OAIS



Fonte: Elaborado pela autora, baseado em Lavoie (2014).

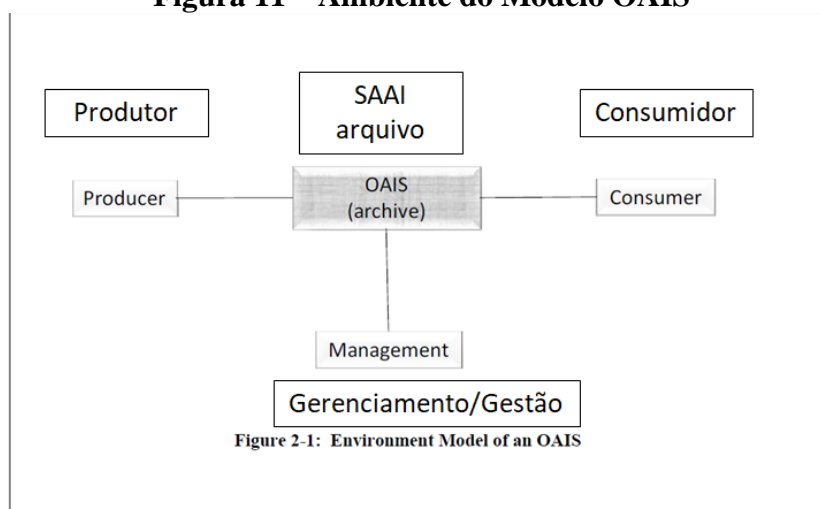
⁷⁰ Modelo de Referência é uma estrutura para o entendimento de relacionamentos significativos entre as entidades de algum ambiente, e para o desenvolvimento de normas consistentes ou de especificações que deem suporte àquele ambiente. Um modelo de referência é baseado em um pequeno número de conceitos unificantes e pode ser usado como base para educação e explicação de normas para um não especialista ABNT ISO 14721 (2021, p. 22).

No Brasil, a norma ABNT NBR 15472 (2007) foi recentemente cancelada e substituída pela norma ABNT NBR ISO 14721 (2021) – Sistemas espaciais de transferência de dados e de informação – Sistema Aberto de Arquivamento de Informação (SAAI) – Modelo de referência, publicada em 28 de julho de 2021. Nesse modelo de referência, o enfoque é a informação digital, com objetivo de preservar em longo prazo, considerando as questões relativas às mudanças tecnológicas e o acesso por tempo indefinido.

O documento tem a intenção de prover um conjunto de fundamentos que visa à preservação e ao acesso às informações em longo prazo. Desse modo, o modelo OAIS ou SAAI pode ser aplicado a qualquer arquivo, embora seja especificamente aplicável às organizações e indivíduos produtores de informação com a responsabilidade de torná-las disponíveis por longo prazo. Entretanto, o modelo de referência não exemplifica ou especifica em seu escopo um desenho de implementação (CCSDS, 2012; ABNT, 2021).

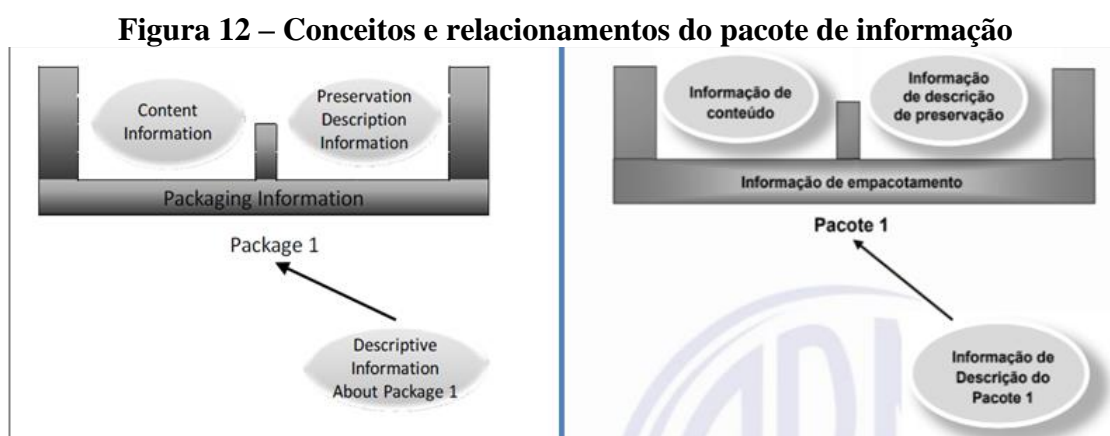
O modelo de referência OAIS divide-se em três ambientes principais: produtor – é o papel exercido por pessoas ou sistemas clientes, que fornecem informações a serem preservadas; gerenciamento/gestão – é efetuado por aqueles que estabelecem a política global para OAIS e não se envolve com as questões cotidianas do arquivo; consumidor – é desempenhado por pessoas ou sistemas clientes, que interagem com os serviços OAIS, para encontrar e adquirir informações que foram preservadas (CCSDS, 2012; ABNT, 2021). O ambiente OAIS/SAAI é o de preservação. O modelo OAIS é apresentado na figura 11:

Figura 11 – Ambiente do Modelo OAIS



Fonte: Adaptado da CCSDS (2012) e ABNT ISO 14721 (2021).

O modelo OAIS ⁷¹ ou SAAI considera informação como “qualquer tipo de conhecimento que pode ser intercambiado” (ABNT, 2021, p. 20). As informações são intercambiadas nesses ambientes e interagem mediante pacotes de informação. Os pacotes são formados por meio de um conteúdo conceitual que provém de dois tipos de informação: *Content Information* (Informação de Conteúdo) e *Preservation Description Information* – PDI (Informação de Descrição de Preservação), ambos encapsulados e identificados por uma *Packaging Information* (Informação de Empacotamento). O pacote resultante é visto como sendo ‘encontrável’ (passivo de descoberta), por meio da *Descriptive Information* (Informação de Descrição) (CCSDS, 2012; ABNT, 2021). A figura 12 abaixo apresenta o referido pacote de informação:



Fonte: CCSDS (2012) e ABNT ISO 14721 (2021).

A informação de conteúdo é aquela que é objetivo original da preservação. A referida informação:

Consiste no objeto de dados de conteúdo (objeto físico ou objeto digital, isto é, bits) e está associada à informação de representação que é necessária para tornar compreensível o objeto de dados de conteúdo para a comunidade-alvo. Por exemplo, o objeto de dados de conteúdo pode ser uma imagem que é fornecida sob a forma de um conteúdo de bits de um arquivo de CD-ROM, juntamente com outros arquivos do mesmo CD-ROM, que também contém a Informação de Representação. (ABNT, 2021, p. 29).

Ademais, a informação de descrição de preservação ou PDI divide-se em cinco tipos. Ela é necessária para preservar a informação de conteúdo, e, também, para assegurar que esta

⁷¹ **Information:** any type of knowledge that can be exchanged. In an exchange, it is represented by data. An example is a string of bits (the data) accompanied by a description of how to interpret the string of bits as numbers representing temperature observations measured in degrees Celsius (the Representation Information) (CCSDS, 2012, p. 12).

seja identificada e que o ambiente de sua criação possa ser compreendido (CCSDS, 2012; ABNT, 2021). O quadro 4 relaciona os cinco tipos de PDI:

Quadro 4 – Os cinco tipos da *Preservation Description Information*

Os cinco tipos da <i>Preservation Description Information</i> – PDI (Informação de Descrição de Preservação) – PCRFA	
<i>Provenance</i> (Proveniência)	descreve a fonte da informação de conteúdo, que teve a custódia desde a sua origem e a sua história (incluindo a história de processamento)
<i>Context</i> (Contexto)	descreve como a informação de conteúdo se relaciona com outras informações fora do Pacote Informação
<i>Reference</i> (Referência)	fornece um ou mais identificadores, ou sistemas de identificadores, por meio dos quais as informações de conteúdo podem ser identificadas de forma única
<i>Fixity</i> (Fixidez)	fornece um invólucro, ou escudo protetor, que protege a informação de conteúdo de alteração não documentada
<i>Access Rights</i> (Direitos de Acesso)	fornece os termos de acesso, incluindo preservação, distribuição, e utilização da informação de conteúdo

Fonte: Adaptado da CCSDS (2012) e ABNT ISO 14721 (2021).

Aliadas às informações apresentadas acima, estão a *Packaging Information* (informação de empacotamento), que vincula de forma lógica a informação de conteúdo com a informação de descrição de preservação, e a *Descriptive Information* (informação de descrição), que é utilizada para descobrir o pacote que contém a informação de conteúdo (CCSDS, 2012; ABNT, 2021).

Além do pacote de informação apresentado, o modelo OAIS demonstra a necessidade de que haja uma distinção entre um pacote informativo que é preservado por um OAIS e os pacotes informativos que são submetidos ou difundidos a partir de um OAIS, que são conhecidos como pacotes variantes. Essas variantes são divididas em três tipos de pacotes:

- *Submission Information Package* (SIP) / Pacote de Informação de Submissão (PIS) – pacote que é enviado a um OAIS/SAAI por um produtor, e a sua forma e conteúdo detalhado são tipicamente negociados entre o produtor e OAIS/SAAI. A maioria dos SIP/PIS vai ter alguma informação de conteúdo e algumas informações de descrição de preservação (CCSDS, 2012; ABNT, 2021);
- *Archival Information Package* (AIP) / Pacote de Informação de Arquivamento (PIA) – consiste no conteúdo informação e a descrição de preservação associada informação que é preservado dentro de um OAIS. Um ou mais SIP são transformados em um ou mais AIP para preservação (CCSDS, 2012; ABNT, 2021);

- *Dissemination Information Package (DIP)* / Pacote de Informação de Disseminação (PID) – configura-se em um pacote de informação, derivado de um ou mais AIP enviados ao consumidor em resposta a um pedido. O OAIS pode fornecer a totalidade ou uma parte de uma AIP a um consumidor na forma de DIP. O DIP pode também incluir coletâneas de AIP, e pode ou não ter informação de descrição de preservação completa (CCSDS, 2012; ABNT, 2021).

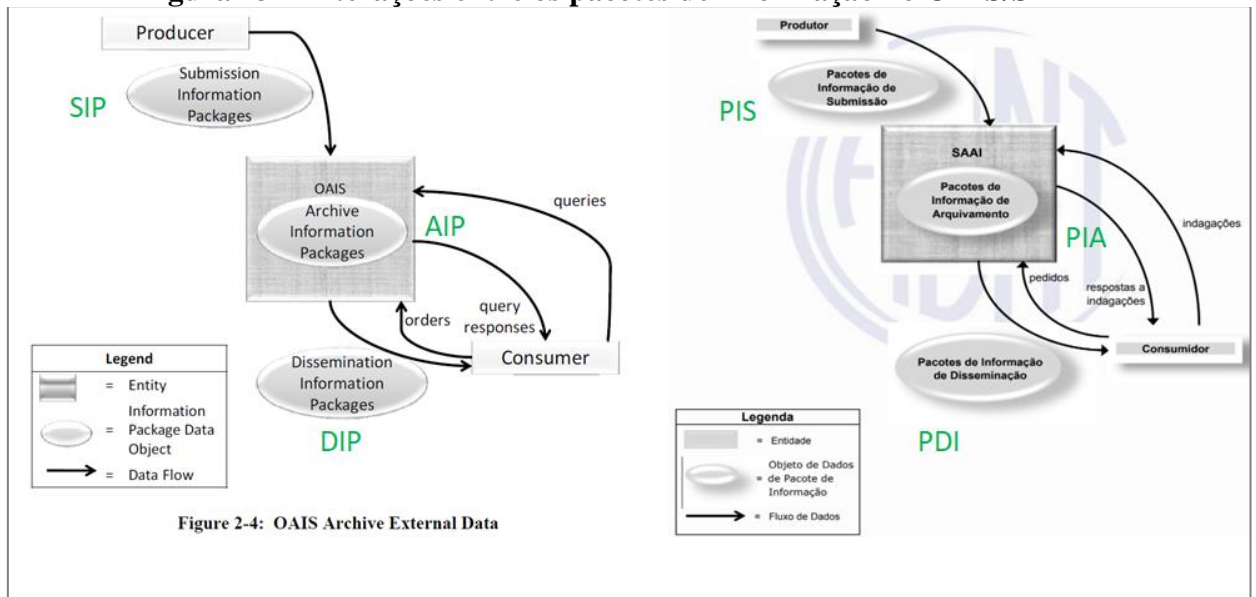
O modelo OAIS é amplamente utilizado para a preservação digital pelas instituições. No Brasil, o Conarq publicou a Resolução nº 43, de 4 de setembro de 2015, que estabelece diretrizes para a implementação de repositórios arquivísticos digitais confiáveis, conhecido como RDC-Arq. De acordo com o Conarq, “um repositório digital confiável é um repositório digital que é capaz de manter autênticos os materiais digitais, de preservá-los e prover acesso pelo tempo necessário” (CONARQ, 2015, p. 9). Além disso, para realizar o gerenciamento dentro de RDC-Arq, é preciso estar em conformidade com o modelo OAIS. O referido modelo conta com seis entidades funcionais, que são:

1. Admissão (*Ingest*) – fornece os serviços e funções para aceitar Pacotes de Informação de Submissão (SIPs) de Produtores, ou de elementos internos sob controle da Administração, e preparar o conteúdo para armazenamento e gestão dentro do Arquivo (CCSDS, 2012; ABNT, 2021);
2. Armazenamento (*Archival Storage*) – fornece os serviços e funções para o armazenamento, manutenção e recuperação de AIPs (CCSDS, 2012; ABNT, 2021);
3. Gestão de Dados (*Data Management*) – fornece os serviços e funções para povoar, manter e acessar tanto a Informação Descritiva, que identifica e documenta o acervo do arquivo, como os dados administrativos utilizados para gerir o OAIS (CCSDS, 2012; ABNT, 2021);
4. Administração (*Administration*) – fornece os serviços e funções para o funcionamento global do sistema, que incluem a solicitação e negociação de acordos de submissão com Produtores, onde são auditadas as submissões para assegurar que o cumprimento das normas, além de manter a gestão da configuração do *hardware* e *software* do sistema (CCSDS, 2012; ABNT, 2021);
5. Planejamento de Preservação (*Preservation Plan*) – fornece os serviços e funções para o controle do ambiente OAIS, apresentando recomendações e planos de preservação para assegurar que as informações armazenadas permaneçam acessíveis e compreensíveis para a comunidade a longo prazo, mesmo que o ambiente informático original se torne obsoleto (CCSDS, 2012; ABNT, 2021);

6. Acesso (*Access*) – fornece os serviços e funções que apoiam os consumidores na determinação da existência, descrição, localização e disponibilidade da informação armazenada no OAIS, e permite aos consumidores solicitar e receber produtos de informação (CCSDS, 2012; ABNT, 2021).

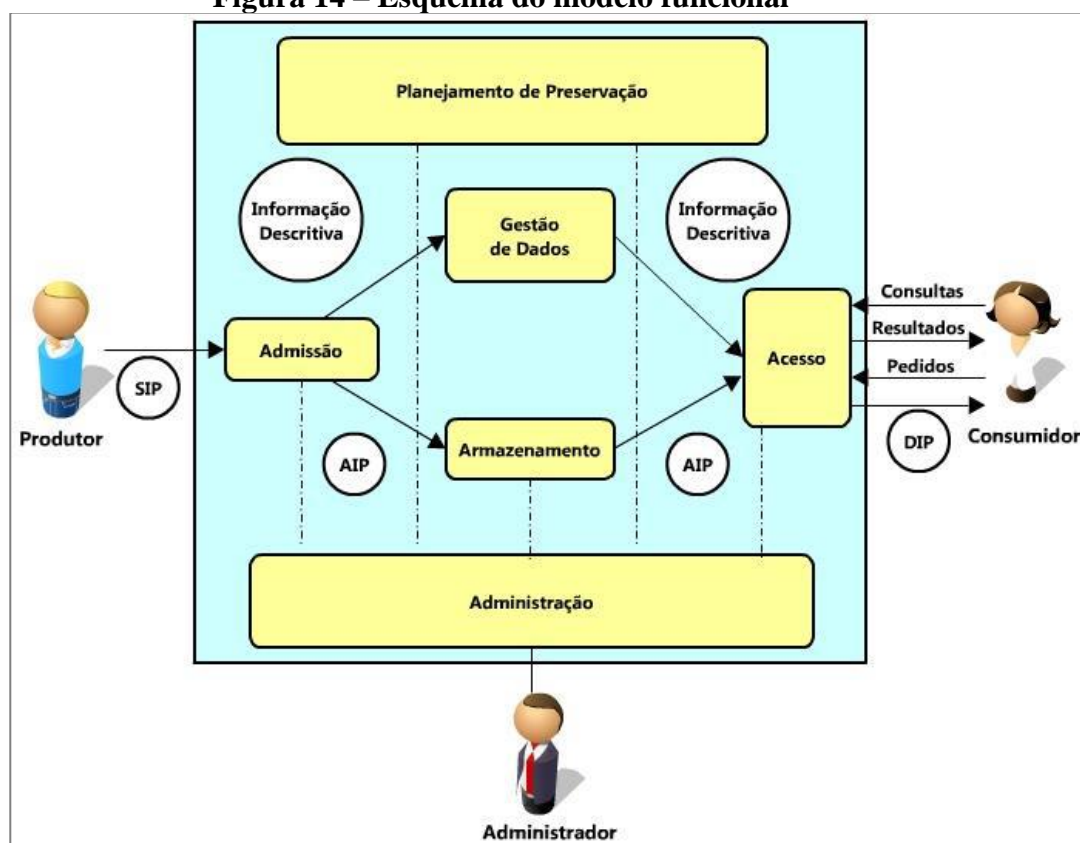
A seguir, as figuras 13 e 14 apresentam, respectivamente, as relações entre os pacotes de informação e os principais fluxos de informação do modelo OAIS.

Figura 13 – Interações entre os pacotes de informação no OAIS/SAAI



Fonte: Adaptado da CCSDS (2012) e ABNT ISO 14721 (2021).

Figura 14 – Esquema do modelo funcional



Fonte: CONARQ (2015, p. 20).

A Resolução Conarq nº 43/2015 aponta como relevante o documento *Trusted Digital Repositories: Attributes and Responsibilities*, produzido pela a *Research Library Group* (RLG) em conjunto com a *Online Computer Library Center* (OCLC), que trata sobre atributos e responsabilidades de um repositório digital confiável a partir do modelo OAIS. O relatório da RLG-OCLC, publicado em 2002, estabeleceu características e responsabilidades fundamentais para a criação e manutenção de repositórios digitais confiáveis que atendessem aos acervos de instituições culturais e científicas, visando garantir a integridade, a confiabilidade e o acesso às informações em longo prazo (CONARQ, 2015).

De acordo com o relatório da RLG-OCLC (2002), além da conformidade com o modelo OAIS, devem ser atributos de um repositório digital confiável: a responsabilidade administrativa (*Administrative Responsibility*), promovendo com transparência todas as ações; a viabilidade organizacional (*Organizational Viability*) – a missão deve refletir o compromisso de guarda, gestão e acesso a longo prazo aos objetos digitais em nome dos depositantes e usuários; a sustentabilidade financeira (*Financial Sustainability*) – é necessário ter um plano de negócios sustentável em vigor e contínuo; a adequação tecnológica (*Technological and Procedural Suitability*) – diz respeito à série de estratégias de preservação,

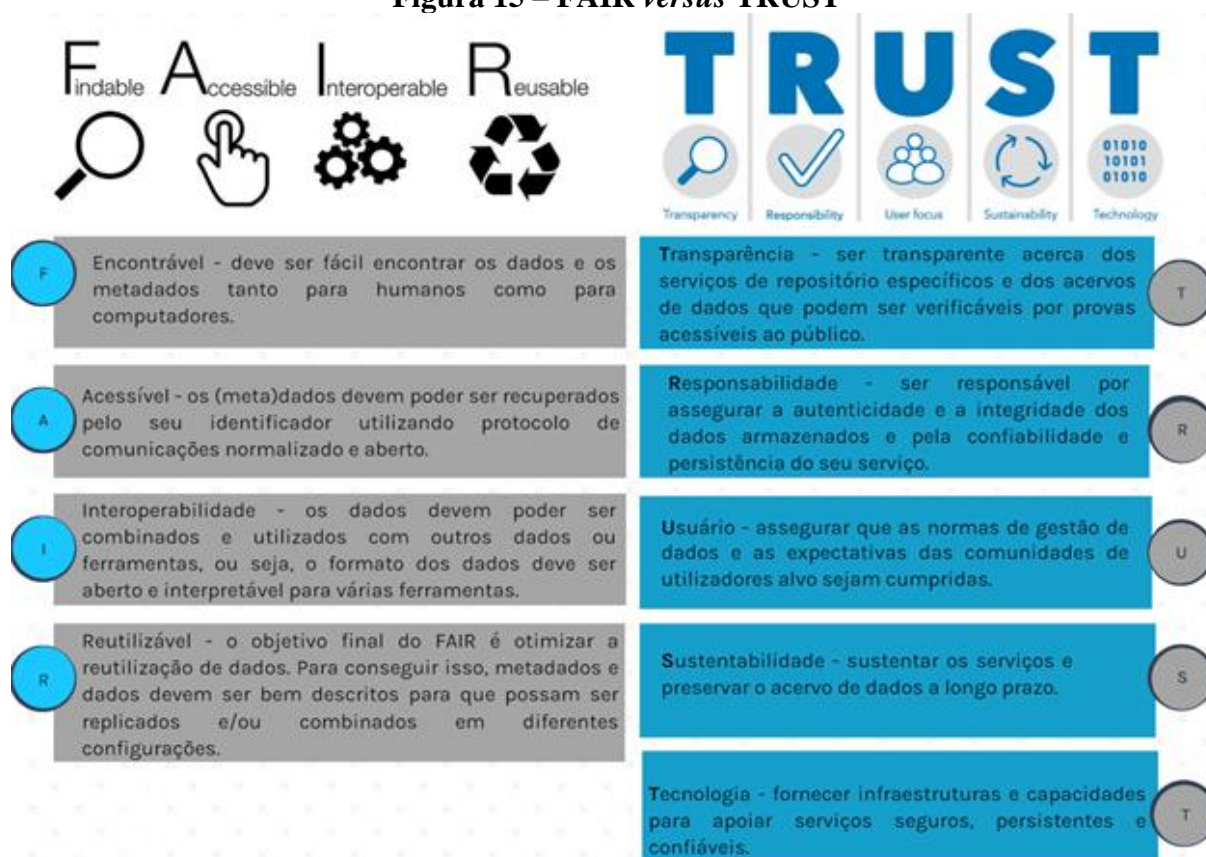
políticas e planos, infraestrutura e auditoria dos processos; a segurança do sistema (*System Security*), com as políticas e práticas que cumprirão os requisitos da comunidade, particularmente os relativos aos processos de cópia, sistemas de autenticação, *firewalls* e sistemas de *backup*; e, como último atributo, aponta a responsabilidade processual (*Procedural Accountability*), ligada às ações executadas pelos agentes e aos mecanismos de monitoramento que asseguram o funcionamento contínuo de todos os sistemas e os procedimentos em vigor (RESEARCH LIBRARIES GROUP, [c2002]).

Em consonância com os atributos elencados no relatório da RLG-OCLC (2002) existem os princípios FAIR (*Findable, Accessible, Interoperable e Reusable*), publicados em 2014, e TRUST (*Transparency, Responsibility, User Focus e Technology*), em 2019, ambos ligados ao movimento de ciência aberta. Tal movimento consiste em promover a inovação por meio de um “conjunto de políticas e ações de disseminação do conhecimento, em geral por meios digitais, para que todos os resultados de uma pesquisa sejam acessíveis a todos, passíveis de reutilização e de reprodução” (FAPESP, [c2021]).

Conforme Lin (2019), os princípios FAIR são aplicados aos objetos de dados, enquanto os princípios de TRUST aplicam-se aos repositórios de dados. Segundo Dillo (2020), eles formam um par perfeito para obter uma definição de alta qualidade e uma gestão transparente dos dados (FAIR), proporcionando uma infraestrutura (TRUST) necessária para salvaguardar a acessibilidade dos dados para o futuro.

A seguir, a figura 15 relaciona os princípios FAIR e TRUST.

Figura 15 – FAIR versus TRUST



Fonte: Baseado e adaptado de GO FAIR ([2021]); Lin *et al.*, 2020; FAIR guiding principles for data resources (2016); Research Data Alliance ([2022]).

Observa-se que FAIR e TRUST são complementares e convergem com os cinco tipos (proveniência, contexto, referência, fixidez e direito de acesso) da PDI (Informação de Descrição de Preservação) do Modelo OAIS citados anteriormente.

Percebe-se que todo o aparato apresentado está ligado ao aspecto técnico abordado por Grácio (2012), e o modelo OAIS é o ponto de partida para viabilizar a preservação dos objetos digitais complexos, como são as páginas *web*.

7.2 METADADOS

As bibliotecas e os arquivos têm longa experiência em enfrentar os desafios da preservação. Tais instituições utilizam padrões de metadados para realizar a descrição dos seus acervos. As páginas *web* impõem ao ofício da descrição novos obstáculos, pois tendem a ser atualizadas de maneira dinâmica, podendo desaparecer sem qualquer aviso.

Cientes desse desafio, a OCLC lançou, em 2018, a publicação *Descriptive Metadata for Web Archiving*, por meio de um grupo de trabalho de arquivo de metadados da *web*

(WAM), para desenvolver recomendações para metadados descritivos. A abordagem é adaptada às características únicas de *websites* arquivados, com intuito de ajudar as instituições a melhorar a consistência e eficiência de suas práticas nesta área emergente (DOOLEY; BOWERS, 2018).

Compreende-se como metadados “a informação estruturada que descreve, explica, localiza, ou torna mais fácil recuperar, utilizar, ou gerir um recurso de informação” (NISO, c2004, p. 1). Eles são normalmente utilizados para qualquer esquema formal de descrição de recursos, aplicando-se a qualquer tipo de objeto digital ou não digital. Segundo a NISO (c2004), existem três tipos principais de metadados: descritivos, estruturais e administrativos, e esses últimos subdividem-se em metadados de gerenciamento de direitos, técnicos e metadados para preservação (NISO, c2004, p. 1). Esses metadados são apontados no quadro abaixo.

Quadro 5 – Metadados

Tipos de Metadados	
Descritivos (<i>Descriptive metadata</i>)	descrevem um recurso para fins, tais como descoberta e identificação. Podem incluir elementos como título, resumo, autor e palavras-chave.
Estruturais (<i>Structural metadata</i>)	indicam como são colocados juntos os objetos compostos, por exemplo, como as páginas são ordenadas para formar capítulos ou páginas <i>web</i> para formar um <i>website</i> .
Administrativos (<i>Administrative metadata</i>) Subdivide-se em técnicos, direitos e preservação.	fornecem informações para ajudar a gerir um recurso, tais como quando e como ele foi criado, tipo de arquivo e outras informações técnicas.
• Técnicos (<i>Technical metadata</i>)	informação sobre arquivos digitais necessária para os decodificar e renderizar
• Gerenciamento de direitos (<i>Rights metadata</i>)	direitos de propriedade intelectual ligados ao conteúdo, por exemplo, uma licença Creative Commons.
• Preservação (<i>Preservation metadata</i>)	apoiam a gestão em longo prazo e a futura migração ou emulação de arquivos digitais.

Fonte: Elaborado pela autora, adaptação de Niso (c2004, p. 1) e Riley (2017, p. 6).

Os metadados são peças fundamentais para assegurar que os recursos continuem acessíveis no futuro. Os metadados descritivos são utilizados para facilitar a descoberta de informação relevante e podem ajudar a organizar os recursos digitais, por exemplo, o número de recursos baseados na *web* cresce exponencialmente. Dessa forma, são utilizados *softwares* para extrair automaticamente e reformatar a informação de conteúdos *web* baseados nos metadados descritivos atribuídos. Os metadados também têm como função promover a

interoperabilidade e a integração, fornecer uma identificação digital e dar suporte ao arquivamento e a preservação.

Na publicação *Descriptive Metadata for Web Archiving*, os autores fizeram uma revisão de literatura com foco no usuário e algumas reflexões sobre esse aspecto. Uma delas é a necessidade de serviços de apoio ao usuário devido à complexidade de acesso e utilização de arquivos da *web*. Segundo os autores, faz parte do papel executado pelas instituições de memória a disseminação da *web* atual, como do seu potencial uso no futuro pelos usuários. Dooley e Bowers (2018) afirmam que é preciso fornecer primeiro mecanismos para a compreensão dos arquivos da *web* (como encontrar e como utilizá-los) e depois entender quais os desafios dos usuários durante suas pesquisas, a fim de encontrar formas para melhorar o acesso e a recuperação (DOOLEY; BOWERS, 2018, p. 12).

Ressaltam-se alguns questionamentos apontados no documento, com base nos documentos institucionais e/ou diretrizes utilizadas para a construção das recomendações. São eles:

- Criador/proprietário do *website*: É um editor, um criador, um assunto ou os três? clarificar a natureza/âmbito do site? Devem ser soletrados os acrônimos? O título deve incluir uma frase do tipo como ‘Website do ...’?
- Datas: que datas são simultaneamente importantes e viáveis de registrar? Início/fim do site existência, data(s) de captura, datas de conteúdo ou data de direitos de autor?
- Extensão: como é que isto é expresso de forma mais útil? 1 website arquivado, 1 recurso on-line, 6,25 GB ou aproximadamente 300 *websites*?
- Proveniência: a proveniência refere-se ao criador do website, ou do repositório que recolhe o site e hospeda o arquivo *web*, formas como o site evoluiu, frequência e datas de captura, ou todos estes?
- Avaliação: significa a razão pela qual o site merece ser arquivado, uma coleção de sites nomeados pelo repositório ou as partes, páginas *web*, que foram ou não recolhidas?
- Formato: é importante que a descrição indique claramente que o recurso é um arquivo web? Se sim, qual a melhor forma de fazer – no título, extensão ou descrição?
- URL: quais os URLs que devem ser incluídos? Links, página de acesso ou página de destino?
- Quando codificado no formato MARC 21, um *website* deve ser considerado um recurso contínuo, recurso integrado, Arquivo de computador, publicação textual, material misto, manuscrito, [Arquivístico] ou qualquer um destes, dependendo do
- Instituição Host: caso a instituição que seleciona, recolhe e hospeda o *website* seja considerado um repositório (para nomear única instituição que recolheu, tem a custódia e dá acesso ao material), colecionador, editor, selecionador ou criador?
- Título: deve ser transcrito textualmente diretamente do site? Editado para
- contexto? [Se o contexto for um arquivo, como seria a descrição?]. (DOOLEY; BOWERS, 2018, p. 15, tradução nossa).

De acordo com Dooley e Bowers (2018), a revisão de literatura, no contexto dos profissionais de informação, revelou necessidades generalizadas que estavam para além do âmbito do grupo *web archiving metadata* (WAM), tais como a importância de melhores ferramentas de descoberta, o desafio quanto ao conteúdo da *web* arquivado em diferentes lugares, a ausência de ferramentas de *software* para facilitar a análise dos dados e a falta de conhecimento pelos usuários de que uma grande quantidade de conteúdos da *web* estão salvaguardados e preservados pelas bibliotecas e arquivos (DOOLEY; BOWERS, 2018, p. 12).

Além de revisões com foco no usuário e nos profissionais, a publicação conta com uma análise das ferramentas de *crawler*, baseada no relatório de metadados descritivos para arquivo da *web*, em que estudaram onze ferramentas com vista à sua funcionalidade para a extração de metadados descritivos. Ou seja, o documento estudou se seria possível uma automatização por intermédio das ferramentas, em vez de criar e/ou reescrever os metadados.

No referido relatório de análise de ferramentas, foi concluído que a maioria das ferramentas de *crawler* de arquivos da *web* concentra-se na captura e armazenamento de metadados técnicos para transmissão e recriação, e capturam metadados descritivos mínimos. Portanto, este último, os metadados descritivos, necessitam que sua inserção ou criação seja de forma manual, quer dentro da ferramenta, quer externamente. Além disso, o título do *website* (tal como registrado nos seus metadados) e a data da captura são rotineiramente capturados, mas pode não ser possível extraí-los automaticamente. Quando ocorre, a extração aparece como ‘página inicial’ ou ‘título’. Também não há um padrão de definição de metadados descritivos. Entre as ferramentas, a esperança seria que os criadores de páginas *web* textuais incorporem rotineiramente mais metadados acessíveis para captura (DOOLEY; BOWERS, 2018, p. 13; SAMOUELIAN; DOOLEY, 2018, p. 9). Tendo em conta a importância dos metadados de descrição, o documento apresenta alguns critérios para sua inclusão. A figura 16 aponta tais critérios:

Figura 16 – Critérios para inclusão de Metadados no contexto *web*



Fonte: Elaborado pela autora, baseado e adaptado de Dooley e Bowers (2018, p. 16).

Segundo os autores, em relação aos critérios elencados acima, poderia simplesmente ter sido indicado o uso do Dublin Core⁷², ao invés de delimitar um conjunto de elementos de metadados da WAM. Todavia, cada elemento foi cuidadosamente considerado e adaptado para o contexto do arquivamento da *web*. O resultado foi um conjunto de 14 elementos de metadados da WAM: contribuinte, criador, data, descrição, língua, relação, assunto e título, que são similares ao Dublin Core. Os outros seis elementos da WAM são diferentes no nome

⁷² Dublin Core – é um conjunto de quinze elementos (propriedades) "nucleares" para descrever recursos. Estes quinze elementos Dublin Core™ foi formalmente padronizado como ISO 15836, ANSI/NISO Z39.85, e IETF RFC 5013. As propriedades do núcleo fazem parte de um conjunto maior de Termos de Metadados do DCMI. O Dublin Core é amplamente utilizado para descrever objetos digitais, inclusive por utilizadores do Archive-It (DCMI, [c2021]; DOOLEY; BOWERS, 2018).

e/ou significado: coletor, extensão, gênero/forma, direitos, fonte de descrição e URL (DOOLEY; BOWERS, 2018, p. 16).

No relatório, não é designado nenhum conjunto de elementos necessários ou essenciais, ficando a critério das instituições que inserem os metadados estarem em consonância com as normas e requisitos mínimos, para interoperabilidade dos dados. Entretanto, o título e a URL de acesso são frisados como os elementos mínimos necessários para identificação e descoberta de qualquer *site* ou coleção arquivada. Estes são frequentemente incluídos como “sementes” em uma descrição multinível, por exemplo, auxiliando na descoberta de arquivos da *web*.

Quadro 6 – Os elementos de metadados WAM

COLETOR (<i>Collector</i>) É a organização responsável pela curadoria e gestão de um <i>website</i> arquivado ou coleção.	Utiliza-se para a organização que seleciona o conteúdo da <i>web</i> para arquivamento, cria metadados e executa outras atividades associadas à ‘propriedade’ de um recurso.	DESCRIÇÃO (<i>Description</i>) Uma ou mais notas que explicam o conteúdo, contexto e outros aspectos de um <i>website</i> ou coleção arquivada.	Utiliza-se para informação textual sobre múltiplos aspectos do conteúdo descrito. Uma vez que consiste em texto não estruturado, é aqui que o conteúdo e o contexto do <i>site</i> podem ser articulados de forma mais clara.
CONTRIBUIDOR (<i>Contributor</i>) Uma organização ou pessoa secundariamente responsável pelo conteúdo de um arquivo <i>website</i> ou coleção.	Utiliza-se para entidades ou pessoas que tenham feito contribuições significativas, mas de forma secundária, para o conteúdo de um <i>website</i> ou coleção e que não estejam especificados no elemento Criador.	EXTENSÃO (<i>Extent</i>) É uma indicação do tamanho de um <i>website</i> ou de uma coleção arquivada.	Pode ser expressa como o número de <i>websites</i> , a quantidade de dados armazenados (em megabytes, gigabytes ou outra medida) ou o número e/ou tipo de arquivos recolhidos.
CRIADOR (<i>Creator</i>) Uma organização ou pessoa responsável principalmente pela criação do intelectual conteúdo de um <i>website</i> ou coleção arquivada.	Incluir este elemento para uma organização apenas quando esta tiver claramente a responsabilidade principal de ter criado o conteúdo intelectual. Em caso de dúvida, utilizar o Contribuinte.	GÊNERO/FORMA (<i>Genre/Form</i>) Termo que especifica o tipo de conteúdo de um <i>website</i> ou de uma coleção arquivada.	Pode fornecer acesso útil a vários tipos de conteúdo, incluindo <i>websites</i> e coleções arquivadas. O uso de um vocabulário controlado é recomendado para encorajar a consistência da descrição.
DATA (<i>Date</i>) Uma única data ou intervalo de datas associado a um evento no ciclo de vida de um <i>website</i> ou coleção arquivada.	Utiliza-se para registrar qualquer data conhecida ou intervalo de datas associado a um <i>website</i> arquivado ou recolhidos que ajudará os utilizadores a compreender o conteúdo.	FONTE DE DESCRIÇÃO (<i>Source of description</i>) Informação sobre a recolha ou criação dos próprios metadados, tais como fontes de dados ou a data em que os conteúdos foram obtidos.	Utiliza-se para identificar a fonte de todos ou alguns dos metadados, particularmente para descrições de <i>websites</i> . Incluir a data em que o <i>site</i> foi examinado e o local a partir do qual a informação foi obtida.

DIREITOS (<i>Rights</i>) São as declarações e permissões legais concedidas pelo direito de propriedade intelectual ou outros acordos jurídicos	Utiliza-se para as condições que restringem o acesso dos usuários ao conteúdo arquivado, e permissão dos detentores dos direitos de autor para a reutilização após o acesso.	IDIOMA (<i>Language</i>) A(s) língua(s) do conteúdo arquivado, incluindo recursos visuais e áudio com componentes linguísticos.	Incluir o elemento Língua para qualquer <i>website</i> ou coleção em que a língua é essencial para compreender o conteúdo. Se estiver em mais de uma língua, indicar tudo o que parece significativo.
ASSUNTO (<i>Subject</i>) Tópico(s) primário(s) que descreve(m) o conteúdo de um <i>site</i> ou coleção arquivada.	Utilizar tantos termos quantos forem necessários para permitir o acesso ao conteúdo do assunto principal. O uso de um vocabulário controlado é fortemente recomendado para encorajar a consistência da descrição	RELAÇÃO (<i>Relation</i>) Utilizado para exprimir relações entre partes de um único <i>website</i> arquivado e qualquer coleção a que pertença.	Ao descrever um <i>website</i> que faça parte de uma coleção de <i>sites</i> arquivados, inclua o título da coleção no elemento Relação para fornecer o contexto dentro do qual o <i>site</i> foi recolhido.
TÍTULO (<i>Title</i>) O nome pelo qual um <i>website</i> ou coleção arquivada é conhecida.	O título de um único <i>site</i> é geralmente transcrito do cabeçalho da página inicial. Se a coleção consiste unicamente em <i>site</i> recolhidos de um único domínio da instituição, incluir o nome da instituição.	URL Endereço Internet de um <i>website</i> ou coleção arquivada.	Utilizar este elemento para registar URLs, URNs ou URIs que sejam úteis aos utilizadores, particularmente sementes e URLs de acesso. Incluir texto para explicar a sua função e repetir o elemento tantas vezes quantas forem necessárias.

Fonte: Elaborado pela autora, baseado e adaptado de Dooley e Bowers (2018, p. 17).

O quadro 6 sintetiza os quatorze elementos propostos pelo WAM e suas respectivas funções no processo de descrição dos objetos digitais complexos. Além disso, são apresentados nas recomendações exemplos práticos para cada elemento destacado, servindo de base para a implementação dos processos de trabalho das instituições de guarda da memória e preservação da *web*.

Apesar do Dublin Core ser considerado simples no âmbito da preservação digital, seus quinze elementos podem ser expandidos. “E para o contexto *web* é sem dúvida um expoente de interoperabilidade, sua eficácia é maior no âmbito descritivo” (MIRANDA; BORBA, 2021, p. 223).

Dooley e Bowers (2018) evidenciaram a necessidade de pesquisas futuras nesse âmbito. O foco das recomendações ficou em torno dos metadados descritivos, por serem itens essenciais para fins de descoberta e identificação. No entanto, os autores frisaram que, ao longo da pesquisa, surgiram muitas outras questões para as quais são necessárias

investigações e soluções, como a relação com outros tipos de metadados, por exemplo, os de preservação, e o desconhecimento por parte dos usuários sobre os arquivos da *web* custodiados pelas instituições de memória. Outro ponto seriam os múltiplos níveis de descrição. As recomendações da WAM tocam nessas questões, mas sem aprofundar em descrição bibliográfica ou arquivística.

Considera-se que as estratégias de preservação digital são diversas e a adoção de metadados para esta finalidade auxiliam as instituições nos “processos relativos ao arquivamento e à manutenção do acesso à informação digital a longo prazo, com garantias de autenticidade, de integridade e de confiabilidade” e viabilizam a interoperabilidade (FORMENTON *et al*, 2017, p. 92). Formenton (2015), por meio de sua dissertação “Identificação de padrões de metadados para preservação digital”, reuniu os principais padrões ligados à preservação digital, conforme figura 17 abaixo:

Figura 17 – Lista de Padrões de Metadados

Nome	Especificação
ANSI/NISO Z39.87, <i>Data Dictionary – Technical Metadata for Digital Still Images</i> ³²	Padrão de metadados técnicos para desenvolver, intercambiar e interpretar arquivos de imagem fixa digital, facilitar a interoperabilidade entre sistemas, serviços e <i>software</i> , assim como apoiar a gestão a longo prazo e o contínuo acesso a coleções de imagens digitais (BARBEDO et al. 2007; NATIONAL INFORMATION STANDARDS ORGANIZATION, 2004, c2011).
<i>Dublin Core</i> (DC) ³³	Padrão para catalogação e suporte a descoberta de recursos eletrônicos na <i>Web</i> (CAMPOS; SARAMAGO, 2007; GRÁCIO, 2012; SAYÃO, 2010).
<i>Encoded Archival Description</i> (EAD) ³⁴	Padrão para codificação de instrumentos arquivísticos de acesso, tais como inventários, índices, dentre outros (ANDRADE, 2007; BARBEDO et al. 2007; BARBEDO; CORUJO; SANT'ANA, 2011; CHAN; ZENG, 2006).
<i>Learning Object Metadata</i> (LOM) ³⁵	Padrão que define um conjunto mínimo de atributos para gerir, localizar e avaliar recursos ou objetos de aprendizagem (CHAN; ZENG, 2006; NATIONAL INFORMATION STANDARDS ORGANIZATION, 2004).
<i>Machine Readable Cataloging</i> (MARC) ³⁶	Padrão para a representação e a comunicação de registros ou informações bibliográficas entre bibliotecas, através da utilização de computadores (CHAN; ZENG, 2006; GRÁCIO, 2012; THOMAZ, 2004).
<i>Metadata Encoding and Transmission Standard</i> (METS) ³⁷	Padrão para codificar metadados descritivos, administrativos e estruturais sobre objetos digitais num repositório (ALMEIDA; CENDÓN; SOUZA, 2012; BARBEDO et al. 2007; CAMPOS; SARAMAGO, 2007).
<i>Metadata Object Description Schema</i> (MODS) ³⁸	Padrão de metadados descritivos derivado do esquema bibliográfico MARC 21, cujo o enfoque são os recursos eletrônicos e os serviços de biblioteca (NATIONAL INFORMATION STANDARDS ORGANIZATION, 2004).
ISO/IEC Moving Picture Experts Group (MPEG) <i>Multimedia Metadata</i> ³⁹	Padrões para a descrição e a representação de recursos ou objetos audiovisuais e multimídia digitais, tais como MPEG-7 e MPEG-21 (NATIONAL INFORMATION STANDARDS ORGANIZATION, 2004; SAYÃO, 2010).
<i>PREservation Metadata: Implementation Strategies</i> (PREMIS) <i>Data Dictionary</i> ⁴⁰	Padrão que define um conjunto de elementos básicos para codificar, armazenar, gerir e intercambiar metadados de preservação no contexto de um sistema de repositório para preservação digital (ARELLANO, 2008; CAMPOS; SARAMAGO, 2007; SAYÃO, 2010).
<i>Record Export for Art and Cultural Heritage</i> (REACH) ⁴¹	Padrão para a descrição de objetos no âmbito de coleções de museus (GRÁCIO, 2012).

Fonte: Formenton (2015, p. 62).

Posteriormente, no artigo “Os padrões de metadados como recursos tecnológicos para a garantia da preservação digital”, foram salientados os elementos úteis para a preservação digital, a partir de seis exemplos de metadados provenientes da figura 17. Entre eles, estão o Dublin Core, MODS, EAD, ANSI/NISO Z39.87, METS e PREMIS (FORMENTON *et al.*, 2017, p. 89 e 90):

- Dublin Core (qualificado) – título, criador, descrição, data, formato, identificador, fonte, relação, cobertura, direitos, detentor de direitos e proveniência;
- MODS – informação de título, nome, informação de origem, descrição física, índice; item relacionado, identificador, localização, parte, condição de acesso, informação de registro;
- EAD – cabeçalho EAD e descrição arquivística;
- ANSI/NISO Z39.87 – informação básica do objeto digital, informação básica da imagem, metadados de captura da imagem, metadados de avaliação da imagem e histórico de mudança;
- METS – cabeçalho METS, metadados descritivos, mapa estrutural, ligações estruturais e comportamento;
- PREMIS – entidade intelectual, objeto, eventos, agentes e direitos⁷³.

Vale ressaltar que todos os padrões apresentados acima podem ser representados em XML, facilitando o intercâmbio de dados e a combinação entre eles, sendo possível utilizar o METS para padronização de pacotes de informação do modelo OAIS em combinação com o PREMIS. Segundo os autores, deve-se “ponderar também a ocorrência de problemas de duplicação, sobreposição e equivalência entre os conjuntos de metadados no uso concomitante destes diferentes esquemas” (FORMENTON *et al.*, 2017, p. 91).

Formenton e Gracioso (2022), no artigo “Padrões de metadados no arquivamento da web: recursos tecnológicos para a garantia da preservação digital de *websites* arquivados”, afirmam que

A preservação digital no arquivamento da web só será possível com o uso efetivo de padrões de metadados, pois são eles que determinaram a persistência, a coerência, a compreensibilidade, o acesso e a representação de sites selecionados, coletados e armazenados em arquivos da Web, além de definirem a arquivabilidade de sites e a interoperabilidade entre sistemas. (FORMENTO; GRACIOSO, 2022, p. 1).

Os autores concluíram que, além do METS e do PREMIS, salientados como opções para a preservação digital, o Dublin Core “mostrou ser um expoente para arquivamento da

⁷³ São direitos e permissões vinculadas ao objeto relevantes para a preservação, por exemplo, permissão para se fazer uma cópia em PDF (SAYÃO, 2010, p. 25).

web por seu uso em iniciativas notáveis da área” (FORMENTO; GRACIOSO, 2022, p. 23). Uma vez que a preservação e o arquivamento da *web* são tarefas complexas, que requerem conhecimentos e competências diversas, vislumbra-se que os avanços nessas questões de metadados serão crescentes e fundamentais para salvaguarda os conteúdos *web*.

7.3 PADRÃO DE ARQUIVAMENTO: ISO 28500:2017

Os arquivos da *web* consistem em grandes coleções de conteúdos; e, devido à sua natureza similar à da própria internet, mostra-se desafiador o dever das instituições em preservar e dar acesso aos conteúdos em longo prazo. As instituições de memória tentaram encontrar formas mais adequadas para recolher essa vasta quantidade de material, utilizando ferramentas de *crawler*. No entanto, para guardar e dar acesso, é fundamental o uso de normas e padrões para o cumprimento de suas missões institucionais.

Em 2009, foi projetada a primeira versão da Norma ISO 28500 — *Information and documentation — WARC file format*, que especifica o formato de arquivo WARC e que foi produzida no âmbito do Comitê de Interoperabilidade técnica da *International Organization for Standardization*. Atualizada em 2017, a norma tem como escopo tratar do armazenamento do conteúdo e informações de controle dos principais protocolos da Internet, como HTTP, DNS e FTP; também versa sobre o armazenar metadados, todas as informações de controle de protocolo e resultados referentes aos dados vinculados ou eventos duplicados. Além disso, serve para apoiar na compressão de dados e manter a integridade do registro de dados, evitando a interrupção da funcionalidade existente (IIPC, 2021d).

O formato WARC, em seu contexto original, é mantido pelo IIPC. O WARC oferece uma convenção para a concatenação de múltiplos registros de recursos (objetos de dados), cada registro consistindo em um cabeçalho seguido por um bloco de conteúdo. O cabeçalho tem campos obrigatórios que documentam, por exemplo, o URI, a data, o tipo e a duração do registro. O bloco de conteúdo pode conter recursos em qualquer formato, como uma página HTML, uma imagem binária ou um arquivo de vídeo. O formato WARC é uma extensão do formato do arquivo ARC, que tem sido tradicionalmente utilizado para armazenar ‘*crawls*’ como sequências de blocos de conteúdo recolhidos na WWW. Cada captura em um arquivo ARC é precedido por um cabeçalho de uma linha, que descreve muito resumidamente o conteúdo recolhido e o seu tamanho (ISO, 2017; AUBRY, 2019; BLUMENTHAL, 2021a).

O arquivo ARC foi originalmente utilizado pelo *Internet Archive* para recursos da *web*. Refletindo sob as necessidades de preservar o contexto sobre os recursos recolhidos, o padrão

WARC foi formalizado em 2009 para incluir os tipos detalhados de metadados técnicos. Como acontece com todos os padrões ISO, o padrão WARC é revisado periodicamente para garantir que continue atendendo às necessidades que emergem da prática. A primeira revisão foi publicada em agosto de 2017, apoiada por uma força-tarefa IIPC, sendo conhecida como versão WARC 1.1, e um processo de revisão está previsto para 2022. A revisão introduziu principalmente novos campos nomeados para deduplicação e a possibilidade de ter carimbos mais precisos de data e hora (AUBRY, 2019).

O arquivo WARC inclui metadados em várias etapas: criação e conteúdo, registros de solicitações e respostas do servidor. Em outras palavras, esse formato registra tudo o que foi feito na transferência de informações de um servidor *web* para seu leitor (BLUMENTHAL, 2021a).

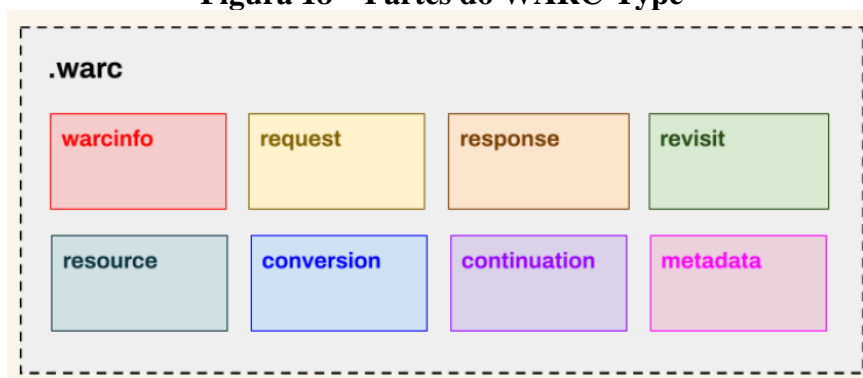
De acordo com a norma ISO 28500:2017, os campos nomeados dentro de um registro WARC fornecem informações sobre o registro atual. O WARC reutiliza cabeçalhos apropriados de outros padrões e define novos cabeçalhos, todos começando por “WARC-”. Para fins específicos, alguns são obrigatórios, conforme quadro 7 abaixo:

Quadro 7 – Campos obrigatórios WARC

Campos nomeados obrigatórios	Funções do Campo
<i>WARC-Record-ID</i>	é um identificador único atribuído ao registro, ou seja, deve-se ter um URI legal e indicar claramente um esquema documentado, por exemplo, através de um prefixo de esquema URI como ‘http: ou urn:
<i>Content-Length</i>	Comprimento ou tamanho do conteúdo
<i>WARC-Date</i>	é um carimbo temporal UTC conforme descrito na ISO 8601[1], por exemplo YYYY – MM – DDThh:mm:ssZ, o carimbo da hora representa o instante em que se iniciou a captura de dados para a criação de registro
<i>WARC-Type</i>	é o tipo de registro da WARC e divide-se em 8 tipos: ‘warcinfo’, ‘response’, ‘resource’, ‘request’, ‘metadata’, ‘revisit’, ‘conversion’ e ‘continuation’.

Fonte: Elaborado pela autora, adaptado da ISO 28500 (2017).

Destaca-se o campo obrigatório *WARC-Type*, pois se desdobra em oito peças distintas, cada uma com o seu próprio significado e atributos de metadados, conforme figura 18 abaixo.

Figura 18 – Partes do WARC-Type

Fonte: Blumenthal (2021a).

Apenas com os registros ‘warcinfo’, ‘request’ e ‘response’ é possível utilizar um *software* de renderização (como, por exemplo, o *Wayback*) para solicitar um documento WARC. Dessa maneira, pode-se obter a mesma resposta que foi gerada no momento da recolha, ler o mesmo HTML ou carregar a mesma imagem em um navegador *web* (BLUMENTHAL, 2021a).

Figura 19 – Definições das partes do WARC-Type

<p>warcinfoc</p> <p>Este registro identifica o arquivo como uma WARC. Conta-nos um pouco sobre como e quando foi criado, quem o criou, e - no caso do Arquivo - a coleção a que pertence. Diz-nos precisamente quando esta aquisição ocorreu, o software que foi utilizado para o fazer, e até a localização e a máquina anfitriã que fez o trabalho, tudo isto é informação de boa proveniência para o futuro.</p>	<p>request</p> <p>Ferramentas de recolha devem solicitar cada página web o documento descarregável a partir do seu servidor web original, ao vivo. Este pedido começa com um cabeçalho de metadados, que inclui informação sobre o pedido, o requerente, e como entregar-lhes os conteúdos relevantes. Sob o cabeçalho vemos o pedido preciso tal como o servidor o recebeu, para que seja documentado e preservado.</p>
<p>response</p> <p>É a resposta do servidor ao vivo a este pedido é também escrita na WARC. Inicia com os metadados do cabeçalho para contextualizá-lo individualmente. E mais uma vez, o cabeçalho é seguido pelo conteúdo original dessa entrega - o arquivo ou código original da web que podemos querer reproduzir num navegador da web.</p>	<p>revisit</p> <p>Descreve a resposta a um pedido de material que já foi arquivado, que não foi alterado. pode incluir apenas um corpo de conteúdo abreviado que tem de ser interpretado em relação a um registro anterior. A utilização de um registro "revisitar" em vez de outro tipo é opcional, para quando os benefícios são a redução do tamanho de armazenamento ou são desejadas melhores referências cruzadas de material.</p>
<p>resource</p> <p>É criado pelo processo de arquivamento da web, para capturar e descrever material relacionado com um recurso arquivado, mas que pode não ter um URL próprio. Por exemplo, um arquivo diretamente recuperado de um repositório acessível localmente ou o resultado de uma recuperação em rede onde a informação do protocolo foi descartada.</p>	<p>conversion</p> <p>Este registro reserva espaço para a eventual migração de materiais da web arquivados para formatos sucessores, se e quando essa necessidade surgir. Por exemplo, um registro HTML5 poderia aparecer aqui para aumentar ou melhorar o acesso ao conteúdo que foi recolhido no formato obsoleto Adobe Flash. Cada transformação deve resultar num registro completo, sem dependência da sobrevivência do original.</p>
<p>continuation</p> <p>Permite a um software de renderização ler e representar um documento arquivado através de dois registros separados, se necessário. Baseia-se na premissa de que o processo de escrever o conteúdo de um registro num arquivo WARC poderia ser interrompido, e que o processo poderia, portanto, ser 'continuado' num registro subsequente, apenas retomando onde parou na linha seguinte.</p>	<p>metadata</p> <p>Contém conteúdos criados para descrever, explicar, ou acompanhar um recurso recolhido, que não foram abrangidas por outros tipos de registros. Por exemplo, um registro de metadados para um recurso incorporado como uma imagem ou vídeo pode descrever como a ferramenta de recolha o identificou e subsequentemente o arquivou.</p>

Fonte: Adaptado de ISO 28500 (2017); Blumenthal (2021a); IIPC (2021d).

De acordo com Blumenthal (2021a), os registros 'conversion', 'continuation' e 'metadata' estão ligados aos processos de gestão e preservação digital.

Desse modo, a normalização abriu o caminho para utilização e implementação de uma gama de ferramentas para coleta, acesso, mineração, intercâmbio e preservação de recursos digitais. Embora o WARC represente o formato padrão único para arquivos da *web*, foi adotado, para além da comunidade, para armazenar materiais nascidos digitais ou digitalizados.

Além dos aspectos técnicos abordados, associar-se a uma comunidade de arquivamento e preservação da *web* é um passo importante na governança institucional, a fim de transpor os desafios nesse âmbito.

7.4 CONSÓRCIO INTERNACIONAL DE PRESERVAÇÃO DA INTERNET

O IIPC, formado em grande parte por instituições europeias, foi criado em 2003, mesmo ano da publicação da carta da Unesco, sendo “formalmente constituído na Biblioteca Nacional da França com mais de 11 instituições participantes”: Biblioteca Britânica, Biblioteca Real Dinamarquesa, Biblioteca Nacional da Finlândia, Biblioteca Nacional da Suécia, Biblioteca Nacional e Universitária da Islândia, Biblioteca e Arquivos do Canadá, Biblioteca do Congresso Americano, Biblioteca Nacional da Noruega, Biblioteca Nacional da Austrália, Biblioteca Nacional da Itália de Florença e o *Internet Archive*, com foco no trabalho colaborativo para arquivar e preservar internet (IIPC, 2004, [c2021a], [c2021b]). “Atualmente o IIPC possui membros de mais de 45 países, incluindo bibliotecas e arquivos nacionais, instituições universitárias e regionais” (IIPC, 2016, 2020a).

O IIPC tem “a missão de adquirir, preservar e tornar acessível o conhecimento e a informação da *internet* para as gerações futuras em todos os lugares, promovendo o intercâmbio global e as relações internacionais⁷⁴” (IIPC, 2016, 2020a, tradução nossa).

Os objetivos⁷⁵ do IIPC que constavam no Acordo⁷⁶ de 3 de fevereiro de 2016, com vigência de 2016 a 2020, eram: permitir a preservação de uma rica coleção de conteúdos da internet em todo o mundo, de forma que os mesmos sejam guardados, protegidos e acessados ao longo do tempo; promover o desenvolvimento e o uso de ferramentas, técnicas e padrões comuns e incentivar e apoiar bibliotecas e arquivos nacionais, e organizações de pesquisa em todos os lugares para tratar do arquivamento e preservação da internet (IIPC, 2016, 2020a, tradução nossa). A recuperação da informação sobre o Acordo de 2016 foi possível por constar no arquivo *web* do IIPC preservados pelo *Internet Archive*.

⁷⁴ The mission of the IIPC is to acquire, preserve and make accessible knowledge and information from the Internet for future generations everywhere, promoting global exchange and international relations (INTERNATIONAL INTERNET PRESERVATION CONSORTIUM, 2016, 2020a).

⁷⁵ To enable the collection of a rich body of Internet content from around the world to be preserved in a way that it can be archived, secured and accessed over time. To foster the development and use of common tools, techniques and standards that enable the creation of international archives. To encourage and support national libraries, archives and research organizations everywhere to address Internet archiving and preservation. (INTERNATIONAL INTERNET PRESERVATION CONSORTIUM, 2016, 2020a).

⁷⁶ INTERNATIONAL INTERNET PRESERVATION CONSORTIUM. *IIPC strategic plan*. 2016. Disponível em: <https://netpreserve.org/resources/IIPC-Strategic-Plan.pdf>. Acesso em: 1 jun. 2021.

No ano de 2021, foi publicado o acordo, válido de 2021 a 2025, entre as instituições membros e atualizado no *website* do IIPC, com os objetivos de: identificar e desenvolver as melhores práticas para selecionar, recolher, coletar, preservar e fornecer acesso ao conteúdo da *internet*. O acordo visa promover uma ampla cobertura internacional em conteúdos de arquivos da *web* por meio de divulgação e curadoria na construção de coleções colaborativas, com intuito de desenvolver a defesa internacional de iniciativas e legislação que incentivem a coleta e preservação de conteúdos da *internet*, a fim de incentivar e facilitar a utilização de arquivos da *internet* para fins de investigação (IIPC, [c2021c]).

Para alcançar os objetivos (2021-2025), o IIPC continuará promovendo um “fórum de compartilhamento de conhecimentos sobre o arquivamento da *web*; desenvolvendo e recomendando normas para a coleta, preservação e acesso a longo prazo ao conteúdo da *internet*”. Pretende também “facilitar o desenvolvimento e a sustentabilidade de *software* e das ferramentas”, com desejo de “aumentar a sensibilização para as questões e iniciativas de preservação da *internet* através de atividades de colaboração, coleta, conferências, *workshops*, eventos de formação e publicações”. Para isso, deve-se “trabalhar diretamente com investigadores e redes de investigação para promover a utilização e análise de conteúdos arquivados da *internet*” (IIPC, [c2021c]).

O IIPC, por intermédio dos seus objetivos, demonstra a importância de iniciativas para guarda, preservação e acesso da *web* em longo prazo, ressaltando que as instituições devam ser incentivadas e apoiadas para alcançar resultados nesse âmbito. Ainda, desenvolve uma série de ações para apoiar a preservação da *web*, como grupos de trabalhos, eventos, financiamento de projetos colaborativos, estudo de casos e treinamentos.

Uma iniciativa relevante do IIPC foi a liderança para a expansão do original formato ARC, utilizado pelo *Internet Archive* desde 1996, para o formato standard WARC (ISO, 2017), padrão utilizado atualmente e ponto-chave dos aspectos técnicos do arquivamento da *web*.

Desde 2007, é realizada a “*The Annual General Assembly and the Web Archiving Conference*”. Grande parte da documentação sobre arquivamento e preservação da *web* é fruto de pesquisadores e instituições membros do consórcio.

Após a publicação da carta da Unesco e a criação do IIPC, muitas instituições criaram ações para “evitar a perda deste importante material representativo do acervo cultural atual que encontramos na *web*, através de programas de coleta/captura e arquivamento de páginas *web*, registrada em seus domínios nacionais” (ALENCAR BRAYNER, 2016, p. 318). Ademais, associar-se a um consórcio internacional é um passo institucional importante para

desenvolver ações no âmbito do arquivamento e preservação digital, na tentativa de garantir o acesso a longo prazo do patrimônio digital.

Compreende-se como membros do IIPC: instituições de patrimônio, como arquivos e bibliotecas, assim como instituições de pesquisas, universidades ou prestadores de serviço que trabalham com preservação da internet.

A participação no referido consórcio proporciona o acesso a uma comunidade de especialistas, informações e serviços ligados ao arquivamento da *web*, incluindo projetos de financiamento. Não há um limite de associados, e as candidaturas são analisadas pelo Comitê Diretor, que procede a uma votação para decidir se aceita ou não a candidatura, tendo autoridade para definir outras categorias de membros além da institucional. O custo pode variar, de acordo com o orçamento operacional anual de cada instituição, e divide-se em três segmentos de cotas em dólar (IIPC, 2021c).

No *website* do IIPC, consta a lista de membros, representados por meio de um mapa. Além dos membros fundadores – 12 instituições –, há também a relação de todos os membros, totalizando 54 instituições. Entre estas, 40 possuem arquivos da *web*, área institucional dedicada ao arquivamento da *web*.

A pesquisa destacará, no primeiro momento, os membros fundadores e, depois, as organizações ibero-americanas participantes do IIPC.

Percebe-se que instituições internacionais já estão preocupadas com arquivamento e preservação da *web* há bastante tempo. Marcial e Vieira (2021) citam a importância da preservação da memória institucional, afirmando que “riscos a obsolescência tecnológica, a falta de políticas públicas para gestão e preservação da informação, no âmbito brasileiro associadas às restrições orçamentárias, à visão a curto prazo e descontinuidade existentes no Estado brasileiro”, põem em risco a memória nacional (MARCIAL; VIEIRA, 2021, p. 165-166). Portanto, “sem uma iniciativa de arquivamento da *web* no Brasil, a maior parte dos conteúdos serão perdidos definitivamente dentro de poucos anos, o que levará à perda a longo prazo e irreversível de conteúdos culturais valiosos⁷⁷” (ALENCAR BRAYNER, 2016, p. 330, tradução nossa).

⁷⁷ Without A Web archiving initiatives, most of this content will be permanently lost within a few years. A delay in moving forwards with a comprehensive national Web archiving initiative in Brazil will lead to the long term and irretrievable loss of valuable cultural content” (ALENCAR BRAYNER, 2016, p. 330).

8 ANÁLISE E DISCUSSÃO DOS RESULTADOS

As instituições que fundaram o IIPC⁷⁸ são: *Bibliothèque nationale de France*, *Internet Archive*, *The British Library*, *Royal Danish Library*, *National Library of Finland*, *National Library of Sweden*, *National and University Library of Iceland*, *National Library of Norway*, *Library and Archives Canada*, *Library of Congress*, *National Library of Australia* e *National Library of Italy*⁷⁹, sendo que esta última integrou o consórcio até 2012⁸⁰, portanto, não será contemplada na pesquisa, pois não permanece como membro.

Quadro 8 – Instituições Membro Fundadoras do IIPC

As Instituições Membros Fundadoras do IIPC	Links
1- Bibliothèque nationale de France	https://netpreserve.org/about-us/members/biblioth%C3%A8que-nationale-de-france-national-library-france/
2- Internet Archive	https://netpreserve.org/about-us/members/internet-archive/
3- The British Library	https://netpreserve.org/about-us/members/british-library/
4- Royal Danish Library	https://netpreserve.org/about-us/members/netarchivedk-royal-danish-library/
5- National Library of Finland	https://netpreserve.org/about-us/members/kansalliskirjasto-national-library-finland/
6- National Library of Sweden	https://netpreserve.org/about-us/members/kungliga-biblioteket-national-library-sweden/
7- National and University Library of Iceland	https://netpreserve.org/about-us/members/landsb%C3%B3kasafn-%C3%ADslands-h%C3%A1sk%C3%B3lab%C3%B3kasafn-national-and-university-library-iceland/
8- The National Library of Norway	https://netpreserve.org/about-us/members/nasjonalbiblioteket-national-library-norway/
9- Library and Archives Canada	https://netpreserve.org/about-us/members/library-and-archives-canada-biblioth%C3%A8que-et-archives-canada/
10- Library of Congress	https://netpreserve.org/about-us/members/library-congress/
11- National Library of Australia	https://netpreserve.org/about-us/members/national-library-australia/

Fonte: Adaptado do IIPC.

Aos países-membros fundadores, foram acrescentadas as instituições que cooperam ou dividem a tarefa de preservação de páginas *web* naquele país: na França, o *Institut National de l'Audiovisuel* – INA; no Reino Unido, o *The National Archives* e o *UK Parliament Web Archive*; e no Canadá, a *Bibliothèque et Archives Nationales du Québec* (BANQ). Dentre as

⁷⁸ <https://web.archive.org/web/20040603043437/http://netpreserve.org/about/members.php>

⁷⁹ <https://web.archive.org/web/20120722084143/http://www.netpreserve.org/about/memberList.php>

⁸⁰ <https://web.archive.org/web/20131113223031/http://netpreserve.org/about-us/members>

entidades citadas, apenas o *UK Parliament Web Archive* não consta como membro do IIPC e não entrará no quadro de caracterização das instituições.

As instituições ibero-americanas são provenientes de Portugal, Espanha e Chile, respectivamente: Arquivo.pt (*Portuguese Web Archive*), *National Library of Spain*, *Library of Catalonia* (Patrimonio Digital de Cataluña – PADICAT) e *Chilean National Library*. A seleção desses países foi com base nos integrantes da Organização dos Estados Ibero-Americanos para a Educação, a Ciência e a Cultura – OEI⁸¹, criada em 1949 em Madri. A OEI configura-se como um organismo governamental de cooperação de países na área da educação, da ciência, da tecnologia e da cultura, que trabalha para fortalecer as políticas públicas e que tem como idiomas oficiais o português e o espanhol.

Quadro 9 – Instituições Ibero-Americanas do IIPC

Instituições Ibero-Americanas Participantes do IIPC	Links
Arquivo.pt (Portuguese Web Archive)	https://netpreserve.org/about-us/members/portuguese-web-archive-fccn-fct/
National Library of Spain	https://netpreserve.org/about-us/members/biblioteca-nacional-de-esp%C3%B1a-national-library-spain/
Library of Catalonia (PADICAT)	https://netpreserve.org/about-us/members/biblioteca-de-catalunya-library-catalonia/
Chilean National Library	https://netpreserve.org/about-us/members/biblioteca-nacional-de-chile-national-library-chile/

Fonte: Adaptado do IIPC.

8.1 CARACTERIZAÇÃO DAS AÇÕES DE GUARDA E PRESERVAÇÃO NAS INSTITUIÇÕES MEMBROS FUNDADORAS DO IIPC

A presente subseção é para cumprir o objetivo B, identificar e caracterizar as instituições de guarda da memória fundadoras do IIPC, tendo sido consultados o *website* do IIPC (<https://netpreserve.org/about-us/members>) e os *websites* institucionais.

8.1.1 National Library of France e Institut National de l’Audiovisuel

A instituição começou o trabalho de arquivamento no ano de 2002, segundo consta no *website* do IIPC, sob a coordenação de Masanès em 2003. No entanto, as primeiras coleções

⁸¹ <https://oei.int/quienes-somos/oei>

foram criadas em 1996, em caráter experimental, com colaboração do *Internet Archive* (BIBLIOTHÈQUE NATIONALE DE FRANCE, 2021a; IIPC, 2021).

O *Bibliothèque nationale de France Web Archives* (<https://www.bnf.fr/fr/archives-de-linternet>) é garantido por legislação, para depósito legal, que é regido pelo Código do Patrimônio⁸², contemplando livros, periódicos, documentos cartográficos, música, documentos gráficos e fotográficos, documentos sonoros, videogames, documentos multimídia e, desde 2006, sítios *Web*, *softwares* e bases de dados, os três últimos conforme a “subseção 4: depósito legal de serviços de comunicação ao público por meio eletrônico (artigos R132-23 a R132-23-2)” (BIBLIOTHÈQUE NATIONALE DE FRANCE, 2021b; IIPC, 2021).

Observa-se que o arquivamento da *web* faz parte da missão da BnF, e a responsabilidade pelo depósito legal das publicações *on-line* francesas e da *web* é compartilhada com o INA. O arquivo *web* francês refere-se ao domínio daquele país, ou seja, aos *sites* registrados nesse país (.fr), sob uma extensão vinculada ao território nacional (.re, ou .bzh por exemplo), ou sob uma extensão genérica (.com ou .org por exemplo), produzidos na França ou com autor residente no território francês. Além da coleta por domínio nacional, também é realizada de forma seletiva, temática e por eventos, produzindo um número expressivo de conteúdo *web* (BIBLIOTHÈQUE NATIONALE DE FRANCE, 2021c; IIPC, 2021).

No caso dos sítios *web*, o recolhimento é realizado por meio de “um *software-robô* no qual os *sites* são explorados e copiados à medida que todos os elementos constituintes das páginas são recolhidos, isto é, textos, imagens, arquivos de áudio e vídeo, animações, folhas de estilo e *links*” (BIBLIOTHÈQUE NATIONALE DE FRANCE, 2021c).

O depósito legal da França no setor audiovisual foi confiado ao INA em 1992, e hoje cobre 168 canais de televisão e estações de rádio. O foco da recolha são todos os programas de origem francesa transmitidos pela primeira vez. Em 2006, foi adicionado em sua missão parte do depósito legal na internet da França, tornando-se parceiro da BnF. O sistema do INA captura, de forma digital por satélite e fibra ótica, os programas de 168 canais de rádio e televisão, gravados 24 horas por dia e 7 dias por semana (INA, 2020, [c2022]).

De acordo com o INA, a coleção de *sites* foi lançada experimentalmente em fevereiro de 2009 como parte do depósito legal e, atualmente, cobre mais de 16 mil *sites* franceses. O

⁸² BIBLIOTHÈQUE NATIONALE DE FRANCE. *Code du patrimoine: Articles L131-1 à L133-1 et R131-1 à R133-1*. Disponível em: <https://www.bnf.fr/fr/code-du-patrimoine-articles-l131-1-l133-1-et-r131-1-r133-1>. Acesso em 16 mar. 2021.

decreto de 19 de dezembro de 2011 sobre o depósito legal da *web* implementou a Lei⁸³ 2006-961 de 1 agosto de 2006 (Lei DADVSI sobre os direitos de autor e conexos na Sociedade da Informação), em que estabelece as condições de seleção e consulta da informação recolhida na internet para depósito legal pelo INA e pela BnF. O escopo do INA abrange: *websites* de departamentos de mídia audiovisual; *web TV* e *web rádios*; *websites* que cobrem principalmente programas de rádio e televisão e *websites* de organizações de comunicação audiovisual profissional e institucional. Além disso, o depósito legal também inclui 15 mil contas de Twitter (desde 2016) e 7.800 contas de plataforma de vídeo (desde 2008), 5.300 *podcasts* e 2.500 *hashtags* (INA, [c2022], [2021a]). O INA⁸⁴ disponibiliza conjunto de dados provenientes das coleções de *sites* recolhidos por meio de um *website* (data.gouv.fr) para reuso e um guia de fundos pesquisáveis (DATA GOUV FR, [c2021]). O processo inovador de recolha do INA permite levar em consideração os aspectos temporais e quantitativos dos *sites* vinculados aos meios audiovisuais. Outrossim, os programas recolhidos para o depósito legal para os quais a instituição não detém quaisquer direitos de exploração são arquivados e melhorados através do processamento técnico. Logo depois, são disponibilizados a qualquer pessoa cujo projeto de investigação exija o recurso a fontes de rádio e televisão (INA, 2020, [c2022]).

Em conformidade com a BnF, a coleção salvaguardada não pretende ser exaustiva, mas busca o princípio da representatividade nacional, e a modalidade de acesso é restrito – em Paris, aos usuários cadastrados, por respeito aos direitos de propriedade intelectual, as coleções não estão acessíveis *on-line* (IIPC, 2021).

Apesar de o arquivo *web* francês não oferecer o acesso remoto, dispõe de uma seção de visita guiada⁸⁵ ao um conjunto de arquivos da *web* coletados pela BnF, divididas por temas, muitos relacionados com a história da *web* e sua evolução (BIBLIOTHÈQUE NATIONALE DE FRANCE, 2021d).

A preservação digital dos arquivos recolhidos é mediante um sistema próprio da biblioteca, no *Système de Préservation et d'Archivage Réparti*⁸⁶ (Sistema Distribuído de Preservação e Arquivamento – SPAR), que visa “garantir a preservação dos documentos digitais, para que a informação permaneça legível, compreensível e reutilizável a longo prazo, mesmo que mude o ambiente técnico no qual os documentos foram produzidos”

⁸³ <https://www.legifrance.gouv.fr/loda/id/JORFTEXT000000266350/>

⁸⁴ <https://www.data.gouv.fr/fr/datasets/collecte-des-sites-web-francais-lies-a-laudiovisuel-francais/>

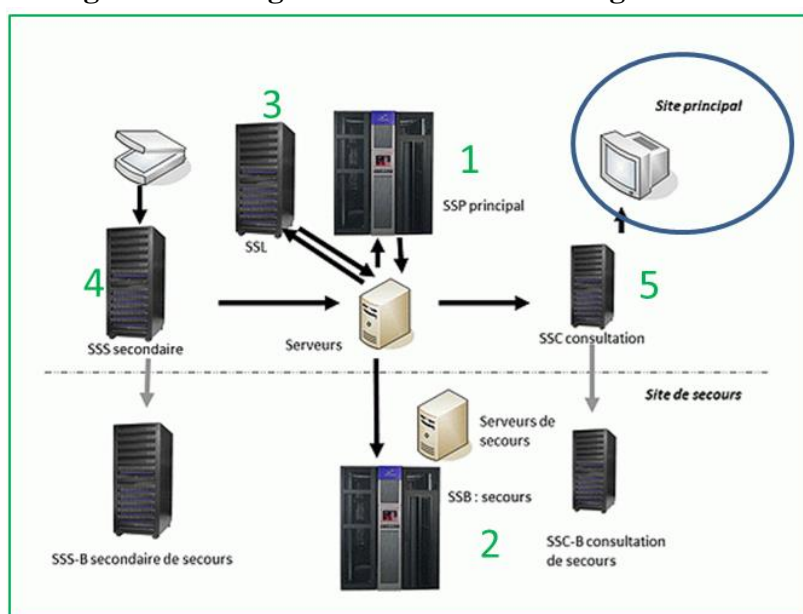
⁸⁵ <https://www.bnf.fr/fr/centre-d-aide/parcours-guides-dans-les-archives-de-linternet#step01>

⁸⁶ <https://www.bnf.fr/fr/spar-systeme-de-preservation-et-darchivage-reparti>

(BIBLIOTHÈQUE NATIONALE DE FRANCE, 2021e). A instituição também concentra esforços no monitoramento preventivo e nas operações de migração de *hardware* e *software*.

O sistema permite a realização de operações de arquivamento digital, assim como a gestão documental, monitoramento da autenticidade dos documentos para fins de comprovação e garantia da segurança do sistema. Por último, o SPAR configura-se como um sistema distribuído, pois permite a gestão de várias cópias de documentos, em diferentes locais, de forma a prevenir a sua perda e destruição (BIBLIOTHÈQUE NATIONALE DE FRANCE, 2021e).

Figura 20 – Diagrama da infraestrutura geral SPAR



Fonte: Adaptado Bibliothèque Nationale de France (2021f)⁸⁷.

A infraestrutura SPAR é composta por um conjunto de elementos de *hardware* e *software* espalhados por dois lugares separados geograficamente. Essa distribuição garante alta disponibilidade de dados e sistema e torna-se um elemento essencial para a proteção das coleções. O acesso ao *site* principal é exclusivo aos produtores e usuários dos dados.

A figura 20 mostra os elementos da SPAR, em que o número (1) configura-se como a unidade de armazenamento principal (SSP), sendo baseada em tecnologia de fita magnética com altas taxas de leitura/escrita – é o lugar que contém a cópia matriz dos dados arquivados; (2) é a unidade de armazenamento de cópia de segurança (SSB), baseada em tecnologia de fita magnética de alta capacidade, sendo o local de salvaguarda onde contém uma segunda cópia dos dados arquivados; (3) é a unidade de armazenamento *on-line* (SSL) que,

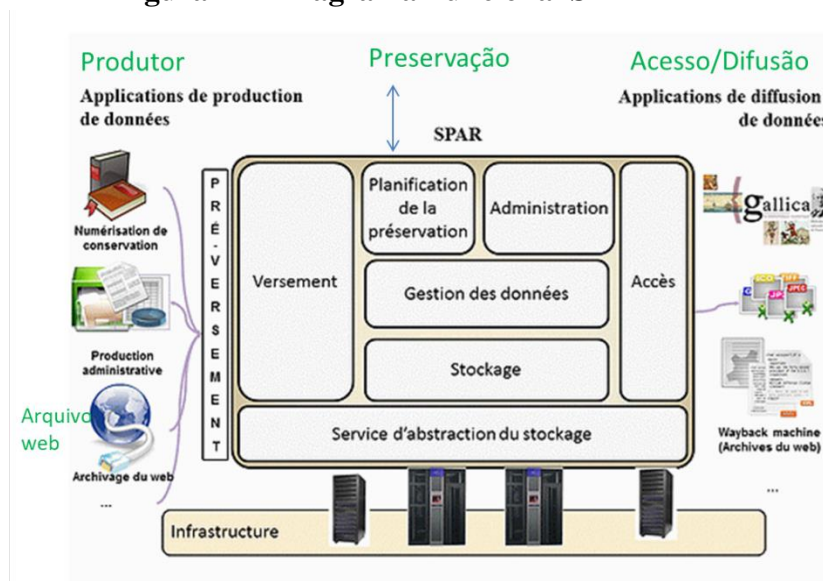
⁸⁷ <https://www.bnf.fr/fr/infrastructure-materielle-de-spar-systeme-de-preservation-et-darchivage-reparti>

diferentemente das duas anteriores, é baseada na tecnologia do disco rígido – consiste em compartimentos que contêm uma terceira cópia de alguns dos dados arquivados; (4) é a unidade de armazenamento secundário (SSS), são compartimentos de disco, para dados temporários que podem não ser permanentes, mas que precisam ser armazenados – contém, em particular, os dados que aguardam o arquivamento; e, por último, (5) unidade de armazenamento e recuperação (SSC), que são compartimentos de disco para dados destinados a consulta e que requerem acesso rápido. Além dessa infraestrutura de armazenamento distribuída, as tecnologias são atualizadas regularmente para evitar o risco de obsolescência (BIBLIOTHÈQUE NATIONALE DE FRANCE, 2021f).

A BnF tomou decisões para criar o SPAR de acordo com o contexto de suas coleções e, também, seguindo normas internacionais, como o modelo OAIS, para promover a interoperabilidade e certificação do sistema. Desse modo, o SPAR foi desenvolvido usando *software* aberto tanto quanto possível para se adaptar a novas necessidades e se encaixar em um cenário complexo e evolução. Assim, os documentos são analisados, validados e enriquecidos com todas as informações e metadados necessários à sua preservação.

O SPAR insere-se, de fato, em um ambiente denso, fruto de uma longa experiência na constituição de coleções digitalizadas e nativas digitais, conforme a figura 21 do diagrama funcional:

Figura 21 – Diagrama Funcional SPAR



Fonte: Adaptado Bibliothèque Nationale de France (2021e)⁸⁸.

⁸⁸ <https://www.bnf.fr/fr/spar-systeme-de-preservation-et-darchivage-reparti>

O ambiente SPAR recebe e preserva documentos recebidos via depósito legal, aquisição, reprodução, produção administrativa, doação etc., além de muitos aplicativos/*softwares* para garantir sua produção, gerenciamento e distribuição. Portanto, as evoluções do sistema levam em consideração todas as novidades referentes às coleções ou na forma de tratá-las.

Logo, as atividades de preservação do arquivo da *web* da França incluem deter, pelo menos, duas cópias em fitas para longo prazo, em dois locais distantes, e uma cópia de acesso ao disco rígido; utiliza-se o formato de arquivo WARC, tendo utilizado o formato de arquivo ARC até 2014. Ambos os formatos estão a cargo do sistema de preservação SPAR. Além disso, as diferentes coleções são preservadas separadamente umas das outras, dependendo de sua ferramenta de produção e/ou de seu produtor, ou seja, dependendo da sua proveniência e contexto os documentos digitais são tratados conforme suas especificidades, para garantir a integridade e a preservação a longo prazo (IIPC, [c2021b]).

8.1.2 Internet Archive

Além da BnF, a organização *Internet Archive* é também membro fundador do IIPC. Iniciado em 1996 nos EUA, configura-se como uma organização sem fins lucrativos, “dedicada a manter e preservar um banco de dados histórico de *sites* e outros conteúdos digitais”, uma entidade de vanguarda na preservação digital nos primórdios da *web* (INTERNET ARCHIVE, [2021a]; IIPC, 2020).

Em 1996, a *Internet Archive*, em conjunto com *Smithsonian Institution*, realizou um projeto para recolher e armazenar na internet os materiais das Eleições Presidenciais dos EUA; a partir daí, a organização foi evoluindo o trabalho de salvaguardar os conteúdos digitais. Fundado por Brewster Kahle, com a colaboração dos conselheiros Peter Lyman e Bruce Gilliat, a organização tinha como intuito oferecer a historiadores, investigadores, acadêmicos e outros o acesso a uma vasta coleção de dados, com “a missão apresentada em seu *website* em 1997, de recolher e armazenar materiais públicos da *Internet* tais como a *World Wide Web* e *Netnews*, e forma de visualização seria por meio do *software* de navegação *Alexa Internet*” (INTERNET ARCHIVE, 1997⁸⁹).

Considera-se uma captura válida de uma página da *web* na *Internet Archive* quando o documento é HTML, texto simples ou em um PDF (GOEL, 2016). Atualmente, a instituição

⁸⁹ <https://web.archive.org/web/19970126045828/http://www.archive.org/>

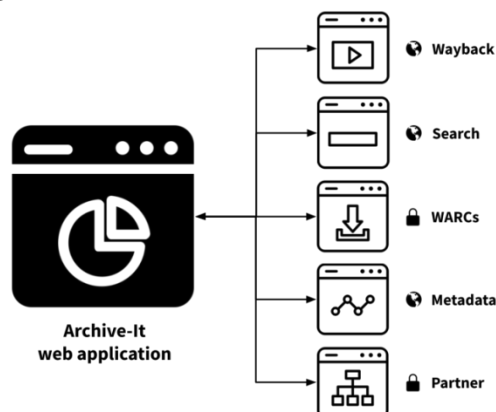
possui 24 anos de arquivamento da *web*, tornando o conteúdo acessível através da ferramenta *Wayback Machine*, com uma média de utilização de 400-500 pedidos por segundo. A interface é em inglês, e a abrangência da coleta pode ser por domínio nacional e/ou regional, seletiva, temática e por evento; além disso, cerca de 549 bilhões de capturas foram recolhidas desde 1996 até o presente, provenientes de cada domínio, mais de 200 milhões de sítios *web* e em mais de 40 línguas. Essa base de dados de arquivos expande-se com a internet, com crescimento aproximado de quase 100 TB (comprimidos) todos os meses (IIPC, 2020).

Em 2002, a *Internet Archive* lançou o *Heritrix*, o rastreador da *web* de código aberto, que é a ferramenta de *software* que captura conteúdo da WWW. Em 2009, ao invés de usar o arquivo gerado pelo rastreador *Heritrix*, foi adotado como um padrão ISO para arquivamento da *web* o arquivo WARC, demonstrando a prevalência de programas ativos de arquivamento da *web* e a importância do próprio rastreador da *web* (BLUMENTHAL, 2021a).

A organização trabalha com mais de 750 bibliotecas, entre outros parceiros, por meio da ferramenta *Archive-It* para identificação das páginas importantes da *web* em todo o mundo. É a “principal solução da *Internet Archive*, para uma vasta gama de organizações, incluindo arquivos, e outras instituições de patrimônio cultural” (INTERNET ARCHIVE, 2020, [2021a]). Desde o seu lançamento em 2006, possibilitou a mais de 800 organizações parceiras utilizarem a solução para criar, armazenar e fornecer acesso às coleções de conteúdos da *web* (INTERNET ARCHIVE, 2020; IIPC, 2020).

A aplicação *Archive-It*, além da funcionalidade central de capturar e preservar conteúdos baseados da *web*, facilita aos usuários adicionar, importar e exportar metadados descritivos, e ainda permite a navegação pública e a pesquisa de texto completo por meio do *website archive-it.org*. Inclusive, fornece uma interface de programação de aplicação (*Application Programming Interface* – API) e outras ferramentas para facilitar integrações externas com *websites*, repositórios locais ou serviços de armazenamento e preservação. Os contratantes dos serviços do *Archive-It* beneficiam-se do acesso à formação, à documentação e apoio técnico especializado (ARCHIVE IT, 2014). E como mencionado, em 2013 o WALCM (Modelo de ciclo de vida do arquivamento da *web*) foi publicado para apoiar o uso de melhores práticas e metodologias para o arquivamento da *web*, sendo desenvolvido com a colaboração dos parceiros do *Archive-It* (BRAGG; HANNA, 2013).

O sistema do *Archive-It* divide os usuários por níveis de conhecimentos: básicos e avançados. Dessa maneira, a ferramenta oferece as opções necessárias para cada tipo de utilizador (ARCHIVE IT, 2014). A figura 22 abaixo ilustra algumas funcionalidades e integrações da ferramenta *Archive-It*:

Figura 22 – Funcionalidades da Ferramenta

Fonte: Archive-It (2021)⁹⁰

O processo de arquivamento da *web* consiste em etapas, que funcionam de maneira síncrona, “para o objetivo final de interagir com um *site* da mesma forma de quando foi arquivado”. Esse processo começa com uma pessoa iniciando um rastreador (*crawler*) *web*. “O *crawler*, ou robô / bot como também são conhecidos, busca e faz a seleção da *web* e reúne todo o material de origem que compõe essa página da *web*. Isso inclui todas as imagens, textos, qualquer *javascript*”, ou seja, todos os elementos que tornem a página dinâmica, qualquer CSS (*Cascading Style Sheets*), é um mecanismo para adicionar estilo a um documento *web*, isto é, a sua ‘aparência’. Após capturar o conteúdo selecionado, inserem-se alguns metadados sobre o *site*, então o rastreador/*crawler* armazena tudo em um WARC. Os WARCs têm um limite de capacidade de 1GB, então geralmente são necessários vários WARCs para armazenar um *site*. Pelo fato de os arquivos WARC não serem automaticamente reproduzidos em uma tela quando abertos, é necessário utilizar tecnologia para reproduzi-los, o mecanismo de reprodução do *Archive-It* é denominado *Wayback Machine* (BLUMENTHAL, 2021b; LOHNDOR, [2021a], [2021b]).

A figura 23 abaixo ilustra o processo realizado pelo *Internet Archive*, com a ferramenta *Archive-It* e *Wayback Machine*:

⁹⁰ <https://support.archive-it.org/hc/en-us/articles/360001231286>

Figura 23 – Arquivamento da *web* pelo Archive-It



Fonte: Adaptado de Lohndor⁹¹ [2021a].

O resultado de todas essas etapas é a aparência da página no dia em que foi arquivada, ou seja, após o processo de renderização, a página pode ser visualizada.

A organização ainda possui um conjunto de serviços⁹² para a comunidade interessada no arquivamento e preservação, conhecida como “*Internet Archive web service & data*”. Conta com o serviço de preservação digital (*Digital preservation & storage service*) – “conjunto de serviços de preservação digital de baixo custo, construído para oferecer acesso a longo prazo ao conhecimento arquivado”, por meio de “software de código aberto, os parceiros são capazes de selecionar um ou vários serviços em um menu de opções, desde replicação de dados até redundância geográfica e relatórios de fixidez, que atendam a seus objetivos de preservação” (INTERNET ARCHIVE, [2021c]).

Ainda, oferece às instituições: um serviço de recolha (*Harvesting service*), recolhendo os conteúdos *web*; serviço de acesso (*Access Service*), com duas maneiras de disponibilizar um rastreamento de domínio para os usuários; serviço de pesquisa (*Research service*) oferecendo acesso a petabytes de dados da *web* arquivada; (*Scholarship Preservation service*) fornecimento de preservação a longo prazo direcionado aos conteúdos acadêmicos; e *Archive-it* o serviço voltado para arquivamento da *web* (INTERNET ARCHIVE, [2021c]).

A *Internet Archive* lançou uma página de comemoração virtual (<https://anniversary.archive.org/#welcome>) para celebrar os seus 25 anos. A organização disponibiliza um grande acervo digital, incluindo 588 bilhões de páginas da *web*, 28 milhões de livros e textos, 14 milhões de gravações de áudio (incluindo 220 mil *shows* ao vivo), 6 milhões de vídeos (2 milhões de programas de notícias de televisão), 3,5 milhões de imagens e 580 mil programas de *software*. É considerado o maior arquivo *web* no mundo, com mais de

⁹¹ <https://support.archive-it.org/hc/en-us/articles/360041674111-What-is-web-archiving>

⁹² <https://webservices.archive.org/>

70 petabytes de dados para acesso ao público (INTERNET ARCHIVE, [2021a]; KAHLE, 2021).

8.1.3 The British Library e The National Archives, U.K.

O arquivamento da *web* no Reino Unido é dividido entre diferentes instituições e com focos distintos, por depósito legal, recolha seletiva, colaboração, visando cumprir o papel de preservar os conteúdos produzidos na *web*.

A biblioteca britânica, membro fundadora do IIPC, começou a recolher *websites* em 2004 de forma seletiva, com intuito de tornar acessível e preservar os recursos *web* de importância acadêmica e cultural para o Reino Unido. Além disso, a partir de 2013, a coleta passou a ser em nível de domínio completo, adquiridos por meio de depósito legal.

A *British Library*, no âmbito da coleta seletiva, prioriza *sites* que refletem a diversidade de interesses e atividades em todo o país; que contêm valor ou são de interesse de pesquisa; os eventos políticos, culturais, sociais e econômicos de interesse nacional ou que demonstrem o uso inovador da *web*. Opta-se por conteúdos exclusivamente publicados na *web* ou *sites* que correm risco de desaparecerem (WEBBER, [2021]).

O depósito legal exige que os editores forneçam uma cópia do que publicaram no Reino Unido para as bibliotecas. Teve início em 1662, e desde 2013⁹³, os regulamentos desse marco legal foram expandidos para incluir as publicações digitais do contexto *web*. “Isto permite que a biblioteca obtenha uma cópia de qualquer recurso publicado digitalmente no Reino Unido (incluindo *websites* e meios de comunicação social) e que disponibilize essa cópia arquivada nas seis Bibliotecas de Depósito Legal do Reino Unido” (BRITISH LIBRARY, [c2021a], [2021b], [2021c]; IIPC, [2019]). Além da *British Library*, a lei do depósito legal inclui *National Library of Scotland*, *National Library of Wales*, *Bodleian Libraries*, *Oxford University*, *Cambridge University Libraries* e *Trinity College*, em Dublin. As referidas instituições têm como objetivo assegurar que a produção publicada no Reino Unido seja recolhida sistematicamente de forma mais abrangente possível em um trabalho conjunto para garantir a preservação a longo prazo das publicações digitais e impressas (IIPC, 2019, BRITISH LIBRARY, [2015]).

As instituições têm direito a copiar o material disponível e publicado a partir da internet para o arquivo *web*, conhecido como *UK Web Archive*

⁹³ <https://www.legislation.gov.uk/ukxi/2013/777/contents/made>

(<https://www.webarchive.org.uk/en/ukwa/index>). Essa área dedicada aos conteúdos *web* tem a missão de coletar todos os *sites* de domínios (.uk, .scot, .london, .wales ou .cymru) e qualquer *site* alojado em um servidor sediado naquele país, pelo menos uma vez por ano. A estratégia consiste em um modelo de coleção mista, permitindo o rastreio periódico da *web* do Reino Unido na sua totalidade, juntamente com a priorização das partes que são consideradas curatorialmente importantes pelas seis bibliotecas, podendo ser nos idiomas inglês, galês e escocês. Estes receberão então maior atenção na curadoria e na verificação da qualidade (BRITISH LIBRARY, [c2021b]).

Em relação ao acesso às publicações depositadas, elas são disponibilizadas em até sete dias após sua recolha na *web*, no entanto, o acesso é restrito ao espaço físico das bibliotecas. Os *sites* são recolhidos por meio de *softwares crawlers* com diferentes propósitos e frequências variadas, podendo ser diária, semanal, mensal, trimestral ou semestral. O processo de recolha consiste em programar uma lista de endereços de domínio (URLs) iniciais para que o *software* solicite uma cópia da raiz ou da página inicial ao servidor de origem. Depois, segue automaticamente os *links* para os níveis seguintes dentro do mesmo domínio, emitindo um pedido separado para todos os URLs identificados. Os *sítios web* emitem automaticamente uma resposta, entregando uma cópia da página ou arquivo referente daquela URL. Além disso, os procedimentos de recolhimento contam com limite de rastreamento, isto é, quando vários pedidos de páginas ou arquivos diferentes são emitidos para o mesmo *website*, o *software* é programado para deixar um intervalo entre cada pedido, para não sobrecarregar o *site* anfitrião, utilizando como *software* de recolha *Web Curator Tool* (WCT) e o *Heritrix* (UK WEB ARCHIVE, [c2021a]; BRITISH LIBRARY, [c2021b]).

Os componentes do modelo de coleção são o recolhimento de domínio anual / bianual, destinado a captar o domínio britânico da forma mais abrangente possível, fornecendo uma visão geral; os *websites-chave*, aqueles que representam organizações e indivíduos britânicos que são de interesse geral em um setor particular da vida do Reino Unido e/ou das nações que o constituem; os *websites* de notícias, contendo notícias publicadas por organizações jornalísticas; e as coleções baseadas em eventos, eventos políticos, culturais, sociais e econômicos de interesse nacional (HOCKX-YU, 2013).

Conforme o *Joint Collecting Framework for UK Legal Deposit, 2015-2020*, documento desenvolvido por especialistas do consórcio de depósito legal, o objetivo dos últimos anos dessas bibliotecas foi recolher os *websites* e melhorar a capacidade de arquivamento da *web*. Assim, ampliar os conhecimentos especializados para que fosse possível compreender, explicar e conceder o uso do conteúdo do arquivo da *web* do Reino

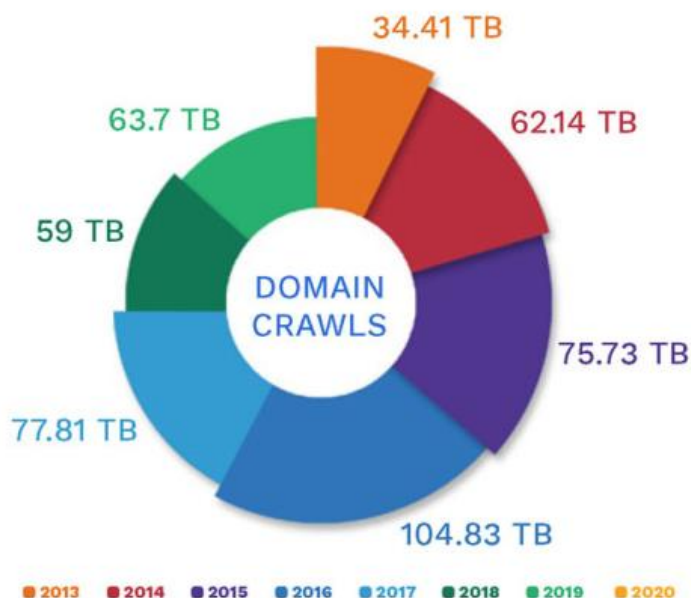
Unido, apoiando a pesquisa, a análise e a visualização desses dados, melhorando o acesso e a visualização das coleções. De forma similar, desenvolver coleções centradas em UKWA e que não se limitem as produções correntes (BRITISH LIBRARY, [2015]).

Há uma parte do *website* chamado de Topics and Themes. Essa área contém mais de 100 coleções de *websites* relacionados com pesquisa, cotidiano, eventos do Reino Unido que podem ser acessadas no UKWA, sendo que algumas delas possuem subdivisões. O UKWA tem autorização dos proprietários para disponibilizar mais de 19 mil *websites* dando acesso via internet. No entanto, muitos dos *websites* arquivados só podem ser acedidos em salas de leitura das Bibliotecas de Depósito Legal e são identificados com a mensagem ‘visível apenas nas instalações da Biblioteca’ (BRITISH LIBRARY, [c2021c]). As temáticas encontradas no UKWA são variadas e com conteúdo produzido por diversas áreas da sociedade. Destaca-se, para a presente pesquisa, a coleção *IT Collection*, que foi iniciada em 2015, cobrindo o universo tecnológico de UK recolhidos pelas bibliotecas. Uma das subcoleções é *Web Archives and Digital Preservation* (Arquivos da web e preservação digital)⁹⁴ composta por 35 *websites*, sendo que 17 estão com acesso restrito ao espaço físico das bibliotecas, e 18, com acesso aberto via internet (UK WEB ARCHIVE, [c2021a]).

É possível conhecer as coleções seletivas também por meio dos *posts* publicados em Blog da *British Library* e no Twitter. O UKWA completou 15 anos em 2020, realizando seu sétimo rastreamento anual. Tal tarefa foi iniciada após a promulgação da legislação, em 2013, de depósito legal para publicações não impressas; e conforme Helena Byrne, curadora dos arquivos da web da *British Library*, cerca de 5-10 milhões de *websites* são arquivados todos os anos e são comprimidos e armazenados à medida que a tecnologia se desenvolve. Em 2019, foram recolhidos 63,7 TB de dados comprimidos, elevando o total recolhido durante *crawls* de domínio, de 2013 a 2019, para 477,62 TB (BYRNE, 2020). A figura 24 abaixo ilustra a quantidade de conteúdo recolhidos nos últimos anos pelas bibliotecas:

⁹⁴ <https://www.webarchive.org.uk/en/ukwa/collection/527>

Figura 24 – Volume de recolhimento anual do UKWA



Fonte: Byrne (2020).

Segundo Byrne, devido ao grande volume de dados que são recolhidos todos os anos a partir do *crawl* anual do domínio e dos *crawls* frequentes, há um atraso significativo entre o conteúdo arquivado até sua disponibilização no *website* do UKWA. Os dados recolhidos entre 2013 e 2019 foram de 250,34 TB, elevando o total combinado para 727,96 TB de dados comprimidos. Para facilitar a pesquisa de conteúdos, o *website* permite pesquisar em todo o conteúdo *Selectively Crawled* (recolha seletiva) de 2005 a 2013, conteúdo *Frequent Crawl* (recolha frequente) de 2013 a 2017, e conteúdo *Domain Crawl* (recolha de domínio) de 2013 a 2015 (BYRNE, 2020).

Recentemente, foi anunciada uma parceria envolvendo a Biblioteca da Universidade de Cambridge e liderada pela Biblioteca Nacional da Escócia, integrantes do depósito legal. Conseguiram o financiamento de £ 230.958 do *Wellcome Trust*, para arquivar e explorar recursos *on-line* sobre informações de saúde e a pandemia de covid-19. O projeto-piloto foi intitulado *The Archive of Tomorrow: Health Information and Misinformation in the UK Web Archive* (O arquivo do amanhã: informações sobre saúde e desinformação no arquivo da *web* do Reino Unido), o qual pretende ter a duração de 14 meses e preservará 10 mil *sites* (oficiais ou não) relacionados com a saúde, e a coleção será disponibilizada no UKWA (UNIVERSITY OF CAMBRIDGE, [c2021]).

Tal iniciativa reafirma o compromisso das bibliotecas do depósito legal publicado no documento *Joint Collecting Framework for UK Legal Deposit, 2015-2020*, em melhorar a capacidade de arquivamento da *web* e acesso às coleções para os usuários e pesquisadores.

Em relação à preservação digital, a *British Library* conta com uma estratégia de preservação digital publicada em 2017-2020⁹⁵, tendo como foco substituir e melhorar o repositório digital. Será implementado um conjunto de ferramentas para preservar e gerir as coleções digitais, inserir metadados para preservação em longo prazo, gerir os relatórios e os processos de forma que sejam documentados e coerentes com a política de preservação digital. A *British Library* possui uma área dedicada ao tema e conta com uma equipe de especialistas para desenvolver uma série de programas para lidar com os desafios da obsolescência tecnológica gerenciamento o ciclo de vida dos *softwares* e *hardwares*, fragilidade das mídias e formatos, somados ao grande volume de dados preservados, com cerca de 3 milhões de itens adicionados por ano no acervo (BRITISH LIBRARY, [c2021d]). Além disso, em 2013, foi revisado o documento publicado pelo *The Preservation Advisory Centre da British Library*, orientado como construir uma política de preservação (BRITISH LIBRARY, 2013).

Outra instituição que realiza arquivamento da *web* é o *The National Archives, U.K* (<https://www.nationalarchives.gov.uk/webarchive/>), também membro do IIPC. Desde 1996, seu foco é o arquivamento dos *sites* do governo, no idioma inglês. Conta com um acesso via internet com textos completos, busca por assunto ou coleções (IIPC, c2021b). Conhecido como UKGWA, contém mais de 3 bilhões de URLs e frequentemente mais de 10 milhões de páginas acessadas todos os meses. O Arquivo Nacional tem a missão de preservar o conteúdo *web* governamental em todas as suas formas (incluindo páginas *web*, publicações oficiais, conjuntos de dados, *tweets* e multimídia) e procura preservar esta parte do registro no seu contexto original. No início, o Arquivo Nacional trabalhava com *Internet Archive* para capturar e preservar o registro governamental. O UKGWA é agora um serviço que é fornecido, sob contrato, ao Arquivo Nacional pela *Internet Memory Foundation*, uma organização sem fins lucrativos (THE NATIONAL ARCHIVES, 2014).

O trabalho de arquivamento da *web* governamental é regido pela política de recolha de registros nacionais de 2012 (*Records collection policy*), que estabelece quais os registros que o Arquivo Nacional deverá recolher para preservação permanente, assim como quando e onde

⁹⁵ THE BRITISH LIBRARY. *Sustaining the value: The British Library Digital Preservation Strategy 2017-2020*. 2017. Disponível em: https://www.bl.uk/digital-preservation/britishlibrary/~/_/media/c0068e4bb25c4ac39d1f4c5a020a21c8.ashx. Acesso em: 7 jan. 2022.

eles serão mantidos. Ademais, é realizada com base na política de seleção de 2014 (*Operational Selection Policy* OSP27), um instrumento de trabalho para os envolvidos na seleção de registros públicos. São recolhidos os registros de todos os departamentos e organismos governamentais de caráter público, com base na lei dos registros públicos (*Public Records Act*, 1958), que podem existir em qualquer formato ou qualquer meio; dessa forma, recolhe-se para que seja preservado e acessível em longo prazo. Com a expansão da utilização de *websites* para publicar documentos oficiais e comunicar-se com o cidadão, foi reconhecido que esses *websites* tinham valor de arquivo (THE NATIONAL ARCHIVES, 2012, 2014; UNITED KINGDOM, 1958).

O *site* do UKGWA conta com uma área para orientar os gestores dos *sites* a cumprirem os requisitos técnicos para que possam ser arquivados posteriormente. De acordo com o documento, são, pelo menos, cinco etapas no processo: 1) são feitas a identificação e a seleção dos *websites*-alvo; 2) depois a recolha, com uma média de 100 *sites* por mês; 3) controle de qualidade para assegurar que o *site* recolhido represente as características do *site* ao vivo; 4) publicação, quando o *site* passa ter uma URL permanente com o prefixo http://webarchive.nationalarchives.gov.uk/*/; e 5) o acesso, quando o *site* entra para o catálogo do *The National Archives* e com índice UKGWA, na função de texto completo (THE NATIONAL ARCHIVES, 2017).

Vale ressaltar que o *The National Archives* publicou alguns documentos relevantes da temática de arquivamento da *web* e preservação digital. Em 2011, foi publicado o guia de arquivamento da *web* (*Web Archiving Guidance*) para orientar os gestores públicos, arquivistas e sociedade sobre a importância dos *websites* com registro, questões técnicas, de acesso e preservação em longo prazo. Depois, um documento intitulado de *Digital Strategy* (2017), com intuito de ampliar o acervo, os serviços e o acesso aos arquivos; tudo apoiado por algumas políticas, entre elas, a política de preservação publicada em 2018. Inclusive, conta com uma área exclusiva para preservação digital, assim como a *British Library*, no entanto, não fornece detalhes de como é processo de preservação digital na instituição.

Além da *British Library* e das bibliotecas do depósito legal, que têm como prioridade recolher material *web* nacional e o *National Archives* com foco nos conteúdos governamentais, o Reino Unido possui um arquivo *web* dedicado ao parlamento britânico. Conhecido como *UK Parliament Web Archive* (UKPWA) (<http://webarchive.parliament.uk/>), inclui *websites* e meios de comunicação social que datam de 2009 até o presente e tem como objetivo preservar os conteúdos do Parlamento do Reino Unido, e qualquer pessoa pode acessar *on-line*. O arquivo parlamentar possui dois documentos ligados à preservação digital,

uma política publicada em 2009 e a estratégia de preservação digital em 2008 (UK PARLIAMENT, [c2021]).

Observa-se que as instituições britânicas empregam esforços para realizar o arquivamento da *web* em diferentes esferas. O UKPWA não faz parte do IIPC, mas foi ressaltado para mostrar a diversidade e organizações em um mesmo país tratando da temática.

8.1.4 Royal Danish Library

A Royal Danish Library, também foi fundadora do IIPC, iniciou o trabalho de recolha de eventos na internet de conteúdo dinamarquês em 2005, com uma área dedicada a essa tarefa. O arquivo *web* dinamarquês é conhecido com *Netarchive.dk* (<https://www.kb.dk/en/find-materials/collections/netarkivet>) (IIPC, 2021b).

De acordo com Webster (2017), foi a partir de 1999, após um *site* utilizado como fonte de pesquisa acadêmica ser retirado do ar sem qualquer aviso, que surgiu a inquietação pela temática por pesquisadores daquele país. Foi então que Brügger e seu colega Finnemann anunciaram a intenção de trabalhar para o estabelecimento de um arquivo *web* dinamarquês. Constituiu-se então uma parceria entre os representantes da Royal Danish Library e a Biblioteca Estadual e Universitária em Aarhus para o estabelecimento do arquivo *web* dinamarquês; possibilitando, posteriormente, a construção da base legal, com a revisão da lei de depósito legal (WEBSTER, 2017).

Dessa maneira, o arquivo *web* da Dinamarca tem como base legal a Lei nº 1.439, de 22 de dezembro de 2004, de depósito legal, do Ministério da Cultura, que abarca o “material publicado em redes de comunicação eletrônicas”. A lei dinamarquesa sobre o depósito legal constitui a base para a recolha e preservação de parte do patrimônio cultural (IIPC, [c2021b]; ROYAL DANISH LIBRARY, [2021a]).

Conforme o marco legal, “o material dinamarquês publicado em redes de comunicação eletrônica está sujeito a depósito legal”, sendo uma obrigação da biblioteca a solicitação de acesso ou reprodução de cópias do material. São considerados material sujeito ao depósito legal dinamarquês conteúdos “publicado a partir de domínios da Internet, etc., que são especificamente atribuídos à Dinamarca”, ou dirigidos a um público daquele país (DINAMARCA, 2004). Caso haja despesas referentes à solicitação de cópias de material, será pago pela biblioteca. A Dinamarca realiza depósito legal desde 1697; e, desde 1781, a lei tem sido utilizada como instrumento de recolha do patrimônio cultural impresso. Com a lei atual e a sua predecessora imediata, o depósito legal foi alargado para incluir outras publicações para

além da impressão – o mais recente material da internet, bem como rádio e televisão (ROYAL DANISH LIBRARY, [2021a]).

A coleta dos conteúdos ocorre de forma automática utilizando o *software NetarchiveSuite*, que foi desenvolvido e mantido pela própria Royal Danish Library. Conta com a colaboração da Biblioteca Nacional da França, Biblioteca Nacional da Áustria, Biblioteca Nacional da Espanha e a Biblioteca Nacional da Suécia (SØRENSEN; HAVE, 2020).

A ferramenta *NetarchiveSuite* começou a ser desenvolvida em 2004, mas foi inaugurada em 2005 para a colheita da *web* dinamarquesa. Tem como principais funções: planejar, programar e executar a recolha da *web*, desde pequenas porções temáticas (por exemplo, relacionadas com eventos especiais, ou domínios especiais) até a recolha e o arquivamento do conteúdo de todo um domínio nacional. O *software* tem a funcionalidade de preservação de bits⁹⁶ e foi construído com base do rastreador da *web Heritrix* (do *Internet Archive*) (SØRENSEN; HAVE, 2020).

Compreende-se como materiais aptos a serem recolhidos aqueles que estão públicos na *web*, ou seja, *websites* privados como intranet são excluídos do recolhimento. Além disso, são utilizadas diferentes estratégias de coleta: transversal, seletiva, eventos, especiais e colaborativas. As recolhas de forma transversal referem-se aos domínios dinamarqueses (em uma média, de quatro vezes por ano); de forma seletiva, são recolhidos conteúdos de todos os meios de comunicação dinamarqueses (de 12 vezes ao dia ou até semanalmente). Estão compreendidos na recolha seletiva os *websites* de partidos políticos, organizações e associações, ministérios e agências, perfis selecionados das redes sociais, vídeos do YouTube (semanalmente); eventos, como eleições parlamentares ou a pandemia de covid-19, uma média de 2 ou 3 eventos anuais; especiais, são aquelas coleções baseada em desejos de pesquisa, além de projetos específicos em colaboração com outras instituições internacionais (ROYAL DANISH LIBRARY, [2021b]).

Para acessar o material recolhido pelo *Netarchive.dk*, fica restrito ao ambiente da biblioteca e ao cumprimento da legislação de direitos autorais e de proteção de dados pessoais. Desse modo, a maioria do material do arquivo *web* só é acessível para fins de investigação mediante pedido (HIELMCRONE, 2005; ROYAL DANISH LIBRARY, [2021c]). É possível realizar uma busca por palavras em páginas *web* (html) no *Netarchive*, e o número encontrado é comparado com o número total de páginas daquele ano. Atualmente, o arquivo *web*

⁹⁶ A preservação de bits é a preservação de sequências de bits que assegura a localização e legibilidade da sequência correta (original) de 0s e 1s para a posteridade (ROYAL DANISH LIBRARY, 2019, p. 13).

dinamarquês possui 730 TB de conteúdo e mais de 32 bilhões de objetos arquivados (ROYAL DANISH LIBRARY, [2021b]).

Para realizar a preservação digital desse grande volume de informação, a Royal Danish Library publicou a versão atualizada em 2019 da sua política de preservação digital. O documento estabelece, de forma geral, as ações e os princípios que a instituição utiliza para a preservação digital, desenvolvido para assegurar que a preservação digital seja efetuada e garanta o acesso ao material digital, tanto em curto como em longo prazo. Levam em conta as questões legais assegurando o cumprimento do depósito legal e dos direitos autorais; financeiro, pois a preservação digital requer investimento constante, desenvolvimento tecnológico, metodológicos e organizacionais das quais dependem a preservação digital, como a tecnologia utilizada para as atividades de recolha da *web*, migração de formatos etc.; método de auditoria, adoção de normas e padrões como OAIS; ações colaborativas com outras instituições em nível nacional ou internacional. Além disso, o aprimoramento contínuo de pesquisas voltadas para temática é contemplado no documento, pois o tamanho e o formato dos materiais digitais impactam na sua preservação (ROYAL DANISH LIBRARY, 2019).

8.1.5 National Library of Finland

A *National Library of Finland*, membro fundadora do IIPC, iniciou o trabalho de arquivamento da *web* em 2006, apoiada em duas legislações: a de depósito legal e de direitos autorais. Para fins de depósito legal, compreendem-se os materiais com domínios ‘.fi ou .ax’, aqueles que estão alocados no país ou direcionados ao público finlandês (IIPC, 2021b).

A Lei nº 1.433, de 28 de dezembro de 2007, de depósito legal, tem como propósito “preservar o material cultural nacional” e disponibilizá-lo ao público finlandês e às gerações futuras. Além dos materiais impressos, estão contemplados na referida lei: os materiais *on-line*, gravações, filmes, programas de rádio e TV. Dessa forma, a biblioteca tem a tarefa de recuperar e armazenar o material *on-line* de conteúdos diversos para que possa representar diferentes momentos da sociedade finlandesa (MINISTÉRIO DA EDUCAÇÃO DA FINLÂNDIA, 2007). A biblioteca nacional divide a tarefa do depósito legal com outras cinco bibliotecas: *Turku University Library*, *Jyväskylä University Library*, *Åbo Akademi University Library*, *Oulu University Library* e *University of Eastern Finland Librar* (NATIONAL LIBRARY OF FINLAND, [2021a]).

Kansalliskirjaston verkkoarkisto ou *Finnish Web Archive* (<https://verkkoarkisto.kansalliskirjasto.fi/>) coleta o material, pelo menos, uma vez por ano com um

rastreador da *web*. Além dos domínios nacional ‘.fi e .ax’, são também recolhidos os ‘.com ou .net’. São contemplados os conteúdos nos idiomas finlandês, inglês e sueco, e a pesquisa dos materiais pode ser feita por meio da URL ou por texto completo (NATIONAL LIBRARY OF FINLAND, [2021a]). A instituição também recolhe os conteúdos por temas, como eventos e assuntos governamentais significativos (por exemplo, eleições); outros eventos importantes, como competições esportivas, eventos culturais, política global, desastres naturais etc. Realizam-se também recolhas em cooperação com outras organizações de memória ou com instituições de pesquisa. As colheitas anuais e temáticas são descritas na bibliografia nacional ao nível de coleção, sendo que as temáticas são identificadas por intermédio de uma lista no arquivo da *web*. Atualmente, o arquivo *web* conta com uma lista de 120 coleções de temas arquivados entre 2006 e 2021, e o tamanho do arquivo *web* fica em torno de 240 terabytes. O acesso ao material é restrito no ambiente institucional das bibliotecas integrantes do depósito legal (NATIONAL LIBRARY OF FINLAND, [2021a], [2021b]; IIPC, 2021b).

The National Library of Finland utiliza o rastreador da *web Heritrix*, tendo como principais alvos *websites*, mas outros dados também são coletados (por exemplo, de servidores FTP). O rastreador da *web* não pode coletar todas as publicações *on-line* finlandesas, ou seja, ficam excluídos intranets, grupos de notícias ou fóruns, publicações *on-line* pagas, e bancos de dados não estão disponíveis para colheita automática. A preservação desse material requer cooperação com os editores (NATIONAL LIBRARY OF FINLAND, [2021a]).

Em relação à preservação, a instituição possui uma política de preservação⁹⁷ (2009) que abrange também os materiais digitais e uma política de digitalização. Na referida política, a instituição reforça o compromisso que “assegurar que o material e o seu histórico de processamento possam ser transferidos do sistema de preservação a longo prazo para outro sistema na sua totalidade, utilizando interfaces e padrões abertos” (NATIONAL LIBRARY OF FINLAND, 2009, p. 11). Também possui o planejamento estratégico 2021-2030⁹⁸, que aborda a visão institucional sobre a preservação digital nos próximos anos.

⁹⁷ NATIONAL LIBRARY OF FINLAND. *Preservation policy*. 21 set. 2009. Disponível em: https://www.doria.fi/bitstream/handle/10024/94315/NLF_Preservation_Policy.pdf?sequence=2. Acesso em: 10 jan. 2022.

⁹⁸ NATIONAL LIBRARY OF FINLAND. *Duties and strategy*. 3 mar. 2020. Disponível em: <https://www.kansalliskirjasto.fi/en/about-us/duties-and-strategy>. Acesso em: 25 jan. 2022.

8.1.6 National Library of Sweden

A *Kungliga Biblioteket*, membro fundadora do IIPC, realiza o depósito legal de materiais impressos desde 1661; posteriormente, foram incluídos materiais eletrônicos que passaram a fazer parte do marco legal *The Swedish Act* (2012:492), como gravações de música, filmes, revistas específicas da *web* e artigos de jornais, *e-books* e publicações de agências governamentais (NATIONAL LIBRARY OF SWEDEN, [2021a]). No entanto, o marco legal não abrange páginas *web*, bases de dados, mídias sociais ou outros meios de comunicação interativos. Segundo Gooding e Terras (2020, p. 105, tradução nossa), tais conteúdos “não foram incluídos na lei de depósito legal, pois foi considerado como grande o desafio de preservar esses objetos digitais complexos a longo prazo”.

Embora a recolha da *web* não seja obrigação por depósito legal, em 1997, a *National Library of Sweden* deu início ao recolhimento da *web* sueca, pelo projeto conhecido como *Kulturarw3 – Kungliga bibliotekets webbkarkiv*, com objetivo de coletar, preservar e fornecer acesso aos documentos eletrônicos daquele país. Com a ajuda de um robô *web*, a instituição coleta da parte sueca da internet com uma frequência de uma ou duas vezes ao ano os domínios ‘.se, .nu, .com e .org’. O conteúdo recolhido pelo *Kulturarw3* é armazenado em fita; de acordo com a instituição, significa que, dessa maneira, levam-se vários minutos para carregar cada página após uma pesquisa, mas a *Kungliga Biblioteket* está trabalhando nesse sentido a fim de melhorar a recuperação da informação (NATIONAL LIBRARY OF SWEDEN, [2021b], [2021c]; ASCHENBRENNER, [200?]).

Estima-se que a *Kungliga Biblioteket* detém 350 terabytes de dados, o que corresponde 5 bilhões de arquivos coletados desde 1997. Todavia, todo esse material possui o acesso restrito a local físico, isto é, somente por meio dos computadores da biblioteca (NATIONAL LIBRARY OF SWEDEN, [2021b]).

Durante a busca por documentos institucionais, não foi localizado no *website* da *Kungliga Biblioteket* uma política de preservação digital ou estratégias nesse sentido; dessa forma, foi enviado *e-mail* para *kulturarw3@kb.se* solicitando mais informações. A resposta foi que a Biblioteca Nacional da Suécia tem, atualmente, uma política de preservação digital publicada em 2012, enviada por anexo em sueco, e que, em breve, será substituída por uma política de preservação (analógica e digital) em geral. O ofício é partilhado entre diferentes áreas, contudo, a biblioteca possui uma equipe de preservação digital com perfis profissionais de diferentes departamentos e divisões – como técnicos de colheita da *web*, técnicos em digitalização, TI (criadores, arquitetos e técnicos), especialistas em metadados

(predominantemente bibliotecários), administradores de depósitos legais e arquivistas. A política futura será menos detalhada e centrar-se-á nos princípios orientadores gerais da preservação. Além disso, a Equipe de Preservação Digital delineará orientações detalhadas relativas às estratégias de preservação digital, em um nível mais granular do que a política de preservação digital de 2012. Todavia, a *Kungliga Biblioteket* está no processo de copiar todo o arquivo *web* para o sistema baseado em disco e convertendo a parte mais antiga do arquivo para o formato WARC; ainda não foi decidido se o arquivo *web* será armazenado tanto em fita como em disco. Nenhuma atividade especial de preservação tem sido usada para o arquivo *web*, além do sistema que tem gerenciado esse material.

8.1.7 National and University Library of Iceland

A *National and University Library of Iceland*, membro fundadora do IIPC, deu início ao recolhimento do domínio ‘.is’ em 2004, embora possua coleções datadas de 1996, arquivadas pelo *Internet Archive*. A instituição dispõe de legislação de depósito legal nº 20/2002, que entrou em vigor em 2003, na qual abarca publicações eletrônicas provenientes de redes de computadores, incluindo *sites* e outros tipos de dados “publicados ou acessíveis ao público na parte islandesa da *World Wide Web*, ou seja, no domínio islandês .is ou material publicado em outros domínios na língua islandesa” (IIPC, [c2021b]; NATIONAL AND UNIVERSITY LIBRARY OF ICELAND, [c2021a], [c2021b]).

Tal tarefa é desempenhada pela *National and University Library of Iceland*, por meio dos *softwares Heritrix* e *OpenWayback*, com coletas por domínio, seletiva e por eventos, que são posteriormente armazenados em duas cópias em discos magnético e outras duas cópias em fitas. O acesso é aberto por intermédio de pesquisa por URL, exceto para os *websites* cujos conteúdos são pagos (NATIONAL AND UNIVERSITY LIBRARY OF ICELAND, [c2021b]).

Durante a busca por documentos institucionais, não foi localizado no *website* uma política de preservação digital ou estratégias nesse sentido, então foi enviado *e-mail* para a instituição solicitando mais informações. No entanto, não houve retorno.

8.1.8 The National Library of Norway

A *National Library of Norway*, membro fundadora do IIPC, iniciou o trabalho de recolhimento do domínio nacional ‘.no’ em 2001, com base em uma legislação de depósito

legal datada de 1990. O marco legal do Ministério da Cultura, de 1º de julho de 2018, nos termos da Lei nº 32, de 9 de junho de 1989, relativa à obrigação de entrega dos documentos acessíveis ao público, § 4, § 4b e § 5, foi revisado em janeiro de 2016 e regulamentado em 2018, com o objetivo de garantir a preservação e disponibilização dos materiais noruegueses para pesquisa e documentação, incluindo os documentos digitais que foram disponibilizados publicamente por meio de redes de comunicações eletrônicas (IIPC, [c2021b]; NORUEGA, 2016, 2018).

A instituição faz o recolhimento a fim de preservar para posteridade os *websites* de forma que sejam garantidas suas características originais no momento da colheita. O nome oficial da ferramenta de coleta da Biblioteca Nacional é “Veidemann”, que significa caçador, que utiliza uma versão especial do navegador Chrome para renderizar os *sites*. Caso o *website* não suporte tal navegador, a biblioteca não garante que o material recolhido terá a mesma aparência da versão original (NATIONAL LIBRARY OF NORWAY, [c2021a]).

No documento de planejamento estratégico 2018-2022, a biblioteca pretende realizar compartilhamento da memória norueguesa, com coleções digitais armazenadas em repositórios que garantam a preservação em longo prazo; além de destacar como importante a *expertise* dos profissionais em conservar e preservar os materiais fornecendo *know-how* e serviços nessas áreas para outras instituições, de arquivos, bibliotecas e museus, assim como a responsabilidade de fornecer às instituições aconselhamento jurídicos sobre a lei de direitos autorais (NATIONAL LIBRARY OF NORWAY, [c2021b]).

De acordo com a *National Library of Norway*, em relação ao arquivamento digital, procura-se utilizar padrões internacionais e repositório digital de confiável para preservar o acervo para o próximo milênio. A instituição tem vários desafios e custos ligados ao estabelecimento e operação de uma solução de arquivamento digital em larga escala, e, dessa forma, esforça-se constantemente para estabelecer soluções para que seja oferecido um alto grau de segurança, integridade, migração segura e conversão de conteúdo digital, quando necessário. Tal tarefa requer pessoas altamente qualificadas em preservação e arquivamento digital, para desenvolvimento e operação de infraestrutura (NATIONAL LIBRARY OF NORWAY, [c2021b], [c2021c]).

8.1.9 Library and Archives Canada e National Library and Archives of Quebec

A *Library and Archives Canada*, membro fundadora do IIPC, desde 2005, é a autoridade para recolher e preservar uma amostra representativa do conteúdo da *web*

canadense, recolhendo recursos federais no contexto do seu programa de arquivo da *web* (LIBRARY AND ARCHIVES CANADA, 2018a).

A instituição possui como marco legal a *Library and Archives of Canada Act* (S.C. 2004, c. 11), aprovada em 22 de abril de 2004, e o Regulamento de Depósito Legal de Publicações (SOR / 2006-337), aprovado em dezembro de 2006, compreendendo como aptos ao depósito legal os “*websites*, incluindo portais, sites pessoais, sites de serviços, intranets e sites que consistem principalmente de links para outros sites; e bancos de dados dinâmicos e dados brutos” (GOVERNMENT OF CANADA, 2006), tais materiais possuem a obrigatoriedade de duas cópias para depósito.

O arquivamento da *web* é conduzido de acordo com a *Library and Archives of Canada Act*, que define que uma amostragem da internet deve ser guardada para fins de preservação digital. A *Library and Archives Canada*, por meio de suas políticas, afirma que os recursos baseados na *web* são patrimônio documental digital e deve-se coletar e preservar tais recursos para acesso e uso. A instituição lançou o Programa de Arquivamento da *Web* em dezembro de 2005, e tem sido uma atividade operacional contínua desde 2013 (LIBRARY AND ARCHIVES CANADA, 2019).

A *Library and Archives Canada* emprega uma metodologia robusta para coletar recursos da *web* e mídia social, que inclui recolha por domínio, isto é, a presença do Governo do Canadá na *web*; temáticas, por eventos e, também, o arquivamento de recursos sob risco conhecido de desaparecer; além de suplementar outras coleções da biblioteca e do arquivo com acervos da *web*, em colaboração com outros especialistas internos e externos (por exemplo, *Truth and Reconciliation Web Archive*). A instituição não fornece acesso às coleções *web* não governamentais, que compreendem cerca 30 terabytes. O acesso aos conteúdos governamentais é por meio do *Canadian Government Information PLN Web Archive*⁹⁹, arquivado pelo *Archive-It*. A coleção da *Library and Archives Canada* possui em torno de 1,5 bilhão de objetos digitais e 60 terabytes de dados e a partir de 2016; os acervos de arquivamento da *web* acumulam-se a uma taxa mínima de 13 terabytes ao ano (LIBRARY AND ARCHIVES CANADA, 2019).

A instituição possui algumas políticas institucionais, como a “*Policy on Holdings Management*” aprovada em 28 de julho de 2014, em que se aplica às atividades da *Library and Archives Canada* em todas as áreas de negócio relacionadas com a gestão do patrimônio documental sob o controle e/ou propriedade da instituição. Com apoio do conjunto de

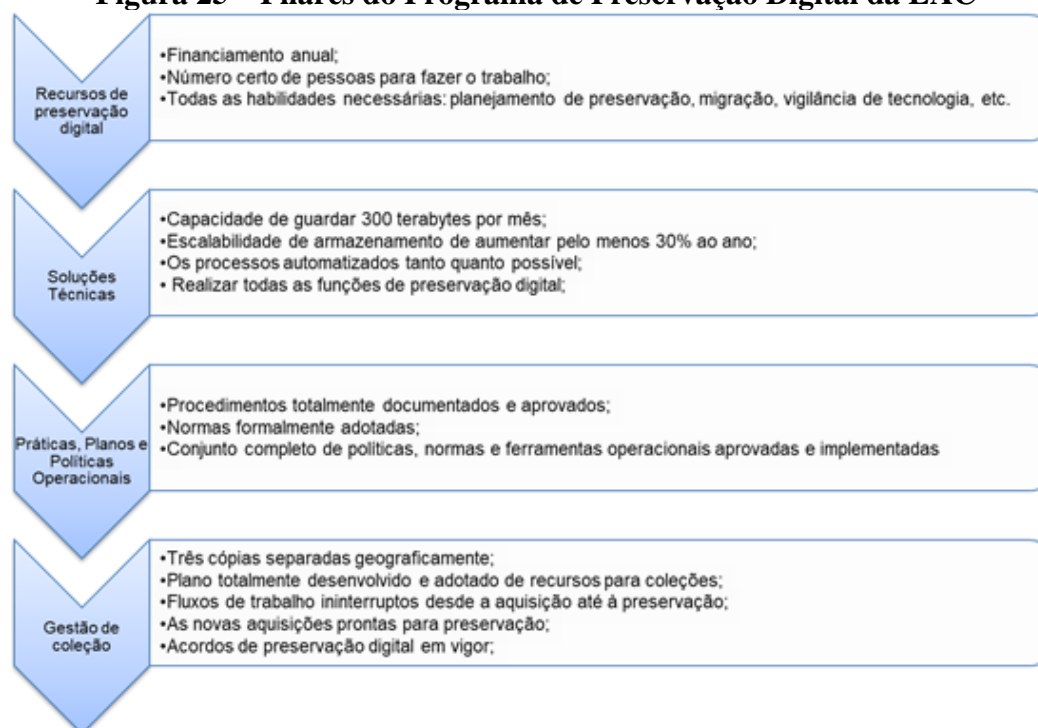
⁹⁹ <https://archive-it.org/organizations/700>

políticas “*Stewardship Policy Framework*” aprovado em 12 de julho de 2013, descreve a abordagem institucional sobre a preservação em longo prazo do patrimônio documental do Canadá, abordando os princípios da confiabilidade, sustentabilidade e colaboração com outras instituições do patrimônio e deve ser aplicado em conjunto com as diretivas, diretrizes e ferramentas da *Library and Archives Canada*. A “*Policy on Holdings Management*” substituiu quatro documentos institucionais: *Library and Archives Canada Preservation Orientations* (2010); *Library and Archives Canada Digital Preservation Policy* (2008); *National Archives of Canada Preservation Policy* (2001) e *National Library of Canada Preservation Policy* (1989) (LIBRARY AND ARCHIVES CANADA, 2018b).

A *Library and Archives Canada* possui uma área física dedicada à sua missão de preservar, que está sendo atualizada para atender às novas necessidades desse campo; a nova instalação tem previsão de inauguração em 2022. No campo da preservação digital, a instituição possui um programa com objetivo de garantir a viabilidade em longo prazo dos materiais digitais e conta com o documento *Strategy for a Digital Preservation Program* lançado em 2017, no qual apresenta as estratégias baseadas em normas ISO como a 14721:2012 do modelo OAIS, a ISO 16363:2012 de Auditoria e certificação de repositórios digitais confiáveis e o modelo de ciclo de vida do *Digital Curation Center* (DCC). Atualmente, o meio de armazenamento utilizado é a fita *linear tape open* (LTO), guardando em torno de 5 petabytes de ativos digitais dividido em 2 cópias (LIBRARY AND ARCHIVES CANADA, 2017).

A instituição pretende, até 2024, implementar a norma ISO 16363:2012 para que o programa se torne sustentável e confiável em longo prazo. A *Library and Archives Canada* almeja que seu programa tenha quatro pilares principais: recursos de preservação digital; soluções técnicas; práticas, planos e políticas operacionais; e gestão de coleções (LIBRARY AND ARCHIVES CANADA, 2017). A figura 25 ilustra o que a instituição define como ideal para o seu programa de preservação digital.

Figura 25 – Pilares do Programa de Preservação Digital da LAC



Fonte: Elaborado pela autora, traduzido e adaptado da *Library and Archives Canada* (2017).

A instituição afirma que as abordagens utilizadas para realizar a preservação digital devem evoluir constantemente, assim, os documentos necessitarão de revisões e atualizações, visando garantir, em longo prazo, um ecossistema digital confiável para salvaguardar a produção documental do país.

O Canadá conta com outra instituição membro do IIPC, a biblioteca regional situada em Québec, a BAnQ (National Library and Archives of Quebec). Tem como marco legal uma legislação de depósito legal sancionada em 2001, que respeita a BAnQ, a qual foi atualizada em 1º de outubro de 2021. O depósito legal permite que a BAnQ colete, preserve e distribua todo o patrimônio documental publicado em Quebec. Além disso, o editor que publica um documento deve depositar a cada edição, no prazo de sete dias após a sua publicação, duas cópias gratuitas na Biblioteca e Arquivo Nacional de Québec, exceto as publicações digitais que têm caráter voluntário (BIBLIOTHÈQUE ET ARCHIVES NATIONALES DU QUÉBEC, 2021a; GOUVERNEMENT DU QUÉBEC, 2021).

De acordo com o IIPC, a instituição, em 2012, iniciou coletas seletivas de *sites* representativos de Quebec, com acesso *on-line*¹⁰⁰ ou local, autorizados pelos seus detentores. (IIPC, c2021b). Os critérios de seleção para coleta da BAnQ são: *sites* que estão em francês ou inglês ou mesmo em uma versão multilíngue; de departamentos e agências governamentais

¹⁰⁰ https://www.banq.qc.ca/collections/collections_patrimoniales/archives_web/index.html?language_id=1

sujeitos à Lei de Arquivos; *sites* temáticos ou *sites* de eventos, criados durante eventos específicos e cuja vida útil é efêmera, por exemplo, eleições e eventos comemorativos. Ao ano, são, no mínimo, coletados cerca de 500 *sites*, devido ao orçamento limitado, restrições legais e falta de um domínio de Quebec. O *software* de coleta é o *Heritrix* versão 3, e os arquivos são mantidos em formato WARC, exceto as versões anteriores que são mantidas em formato ARC. Ademais, a visualização é através do *OpenWayBack* (navegação no *site*) (BIBLIOTHÈQUE ET ARCHIVES NATIONALES DU QUÉBEC, 2021b).

A BAnQ conta com uma série de publicações institucionais, que abarca armazenamento, preservação, digitalização; no entanto, nada específico para preservação digital. No planejamento estratégico 2019-2023, é citada a Carta sobre preservação digital da Unesco e ressaltado que, atualmente, a BAnQ detém cerca de 2,5 petabytes e que poderá ter de gerir da ordem dos 20 petabytes em 10 anos de objetos digitais. No documento, salienta que, embora a instituição tenha feito progressos nos últimos anos, os requisitos de armazenamento, segurança e curadoria de dados ao longo do seu ciclo de vida são atendidos de forma parcial na atual solução para salvaguardar os acervos digitais. Tal solução ou estratégia utilizada não são especificadas no documento de planejamento estratégico, mas de acordo com Montmartin (2018), a instituição utiliza para gestão do acervo digital o Dspace (BIBLIOTHÈQUE ET ARCHIVES NATIONALES DU QUÉBEC, 2021c, 2021d; MONTMARTIN, 2018).

8.1.10 Library of Congress

A *Library of Congress*, membro fundadora do IIPC, iniciou o trabalho de recolhimento da *web* em 2000, por meio de seu programa de *web archiving*, adquirindo *websites* de maneira temática, seletiva ou por eventos, formando coleções. A instituição não é obrigada a fazer o recolhimento por meio de um mandato legal, por exemplo, de depósito legal. Todavia, possui a lei de *copyright* que obriga que qualquer trabalho publicado nos EUA protegido por direitos autorais seja enviado em duas cópias para a *Library of Congress*. Em geral, as obras eletrônicas que são publicadas nos EUA e disponíveis apenas *on-line* não são sujeitas à exigência de depósito obrigatório. No entanto, um livro ou série apenas eletrônica está sujeito a depósito obrigatório se houver um pedido formal para a utilização na Biblioteca do Congresso; os *websites* estão excluídos deste mandato legal (IIPC, 2021b; U.S. COPYRIGHT OFFICE, 2021; LIBRARY OF CONGRESS, 2017; ENGLE, 2012).

Estima-se que o arquivo da *web* da *Library of Congress* contenha mais de 2 petabytes de conteúdo e crescimento exponencial de cerca de 20 a 25 terabytes por mês. O *software* de recolha utilizado é o *Heritrix*; os conteúdos são posteriormente armazenados no formato WARC. A renderização é feita por meio do *OpenWayback*, e a frequência irá depender do escopo da coleção estipulado previamente. O acesso às coleções descritas é aberto por intermédio do <https://www.loc.gov/web-archives/> ou mediante pesquisa por URL, exceto aos conteúdos cujos proprietários não concederam a permissão para exibição. A *Library of Congress* fornece recomendações para a criação de *sites* preserváveis, como as diretrizes de acessibilidade, padrões *web*, metadados, licenças de uso e formatos sustentáveis em longo prazo (LIBRARY OF CONGRESS, 2021b, 2021c).

A LC tem uma área dedica à preservação, que subdivide em preservação digital, em que são incluídos programas relacionados com o empacotamento e a ingestão de conteúdo digital, monitoramento e relatórios de armazenamento digital, formatos de arquivos digitais sustentáveis e metadados. Nos anos 2000, iniciou um Programa Nacional de Infraestrutura e Preservação da Informação Digital (*National Digital Information Infrastructure and Preservation Program* – NDIIPP) que esteve apoiando e financiando projetos de outras instituições americanas, com objetivo de salvaguardar conteúdos em risco naquele país, ainda promovendo eventos, treinamento e educação continuada no âmbito da preservação digital. Vale destacar uma ação de educação promovida pela *Library of Congress* em colaboração com *Archive-It* voltado para o arquivamento da *web*, o “*The K-12 Web Archiving Project*”¹⁰¹, no qual estudantes arquivaram *sites* ligados à sua vida ou interesse naquele momento, desenvolveram habilidades de pensamento crítico e uma consciência da natureza transitória do conteúdo da *web*; o projeto aconteceu entre 2008 e 2014. O NDIIPP publicou um relatório em 2010, no qual são mencionados os objetivos estratégicos de 2000 a 2020 do programa nas áreas de gestão de conteúdos, infraestrutura distribuída, estudo de políticas e divulgação (NATIONAL DIGITAL INFORMATION INFRASTRUCTURE AND PRESERVATION PROGRAM, 2011; LIBRARY OF CONGRESS, [2021d], [2021e], [2021h]).

A *Library of Congress* publicou o documento “*Digital Strategy for the Library of Congress 2019-2023*”, em que enumera quatro objetivos de alto nível: expandir o acesso, aprimorar serviços, otimizar recursos e medir resultados. Pretende-se alcançar tais objetivos nos próximos cinco anos. Foi complementado por outro documento institucional, “*Digital Collections Strategy Overview 2022-2026*”, publicado em outubro de 2021, no qual

¹⁰¹ <https://www.archive-it.org/k12/>

especificam os quatro objetivos da estratégia digital: expandir o acesso; aprimorar os serviços; otimizar recursos e medir o impacto. Desse modo, o escopo será em torno das seguintes áreas: nos fluxos de trabalho, políticas e práticas para recomendação, identificação, seleção, aquisição, descrição, processamento, preservação e acesso de conteúdo digital para a coleção da biblioteca; infraestrutura técnica da biblioteca para suportar todo o ciclo de vida do conteúdo digital; desenvolvimento de políticas e procedimentos de acesso adequados a diferentes categorias de conteúdo digital; treinamento e desenvolvimento de pessoal; acesso a todo o material digital relevante, independentemente de sua localização; e realinhamento de recursos para atender um ambiente crescente de coleções digitais. Pela transversalidade do trabalho em ambiente digital, foi formado um grupo de trabalho com a representação das partes interessadas da *Library of Congress* (LIBRARY OF CONGRESS, [2021f], [2021g]). Todavia, ambos os documentos não mencionam os aspectos técnicos e tecnológicos utilizados para realizar a preservação digital.

Apenas no programa de gestão de coleções digitais (Program Digital Collections Management) puderam-se conhecer os padrões utilizados pela *Library of Congress* no âmbito da preservação digital, pois contém “The Digital Collections Management Compendium (DCMC)” que visa reunir ou resumir as orientações e as políticas relevantes que formam as práticas de gerenciamento da coleção digital da *Library of Congress*. A instituição utiliza o padrão PREMIS – *PREservation Metadata: Implementation Standard*, os níveis de preservação digital do NDSA e o modelo OAIS – ISO 14721 (LIBRARY OF CONGRESS, [2021i]).

Observa-se que a *Library of Congress* possui diversos documentos que apoiam o trabalho da preservação digital, implementando programas de longa duração para que sejam garantidos o acesso e o uso em longo prazo.

8.1.11 National Library of Australia

A *National Library of Australia*, assim como as instituições apresentadas anteriormente, é membro fundador do IIPC; começou seu arquivo da *web* em 1996. A instituição possui mandato legal *Copyright Act* 1968 exigindo que os editores depositem as publicações impressas, e desde 17 de fevereiro de 2016, a lei sofreu alteração, ampliando o depósito legal para materiais *on-line*, incluindo *sites* e páginas *web* (IIPC, 2021b; NATIONAL LIBRARY OF AUSTRALIA, 2016; AUSTRALIAN GOVERNMENT, 1968).

Ao depositarem os materiais, os editores selecionam o nível de acesso ao público, que pode ser um nível restrito ao local, ou seja, apenas no espaço físico da biblioteca, ou amplo, quando qualquer pessoa poderá baixar e salvar, a partir do catálogo da biblioteca ou pelo Trove¹⁰² (serviço de descoberta nacional) (NATIONAL LIBRARY OF AUSTRALIA, [2021a]).

PANDORA¹⁰³ é a área dedicada ao arquivamento da *web* australiana, “seu acrônimo resume a missão de preservar e acessar recursos documentais em rede da Austrália” (NATIONAL LIBRARY OF AUSTRALIA, 2020). A *National Library of Australia* é a instituição responsável por uma ampla cobertura nacional com prioridade para conteúdo de relevância e significância para o país que não fica restrito ao domínio ‘.ua’, e divide a tarefa com outras nove instituições estaduais, que buscam selecionar conteúdos que reflitam suas jurisdições. O escopo do arquivo *web* é seletivo, por eventos, conteúdo governamental e domínio nacional. Para o conteúdo governamental, a instituição, em 2014, criou o programa *Australian Government Web Archive* (AGWA) para colheita em massa dos conteúdos, que pode ser acessado através do endereço (<https://webarchive.nla.gov.au/collection>). A frequência da recolha varia de acordo com o cronograma preestabelecido pelos curadores *web*. A biblioteca faz a notificação aos editores que já atenderam as exigências do depósito legal por meio de rastreadores da *web*, como *Heritrix*, HTTrack ou Webrecord; e o acesso é por meio do Trove ou pelo próprio PANDORA (NATIONAL LIBRARY OF AUSTRALIA, 2020, [2021a], [2021b]).

O arquivo da *web* tem como objetivo fornecer o acesso em longo prazo ao patrimônio australiano e utiliza a 4ª ed. da política de preservação digital da *National Library of Australia*, publicada em fevereiro de 2013, e documentos auxiliares com a estratégia de coleção 2020-2024.

8.2 POLÍTICAS, CRITÉRIOS E AÇÕES DE PRESERVAÇÃO DIGITAL DAS INSTITUIÇÕES IBERO-AMERICANAS DO IIPC

Para atender ao objetivo C, identificar as políticas, ações e critérios de preservação digital, utilizados pelas instituições ibero-americanas do IIPC, foi consultado o *website* do IIPC e os *websites* das instituições, e analisado o conteúdo da documentação disponibilizada.

¹⁰² Trove é um serviço de descoberta de coleções digitais, contendo mais de 6 bilhões de itens de diferentes instituições, incluindo bibliotecas, museus, galerias, a mídia, organizações governamentais e comunitárias e muito mais em toda a Austrália. Disponível em: <https://trove.nla.gov.au/about/what-trove>.

¹⁰³ <http://pandora.nla.gov.au/>

Além disso, foi enviado e-mail para todas as instituições ibero-americanas solicitando mais informações que não foram localizadas em seus *websites*.

8.2.1 Arquivo.pt – Portuguese Web Archive

O Arquivo.pt iniciou seu trabalho em 2008 a partir de uma ideia surgida em 2001, por meio de um projeto de pesquisa desenvolvido pela Faculdade de Ciências da Universidade de Lisboa. O objetivo da instituição é arquivar de forma periódica os “conteúdos portugueses disponibilizados publicamente na Internet e posterior disponibilização ao público através de um sistema de pesquisa integrado”. Configura-se como uma unidade orgânica da Fundação para Ciência e Tecnologia – FCT, que é a agência pública nacional de apoio à investigação em ciência, tecnologia e inovação, em todas as áreas do conhecimento, alocada no Ministério da Ciência, Tecnologia e Ensino Superior; e da unidade pertencente à FCT, a Fundação para Computação Científica Nacional – FCCN, que liga instituições e laboratórios, como uma rede nacional de educação e pesquisa (ARQUIVO.PT, 2017a, 2021a; FCT, [c2021a]; FCCN, [c2021]). De acordo com o Decreto-Lei nº 55, de 17 de abril de 2013, art. 3º: missões e atribuições da FCT, alínea 2n:

[...] Assegurar a inventariação, gestão e preservação do património científico e tecnológico nacional, bem como do acervo bibliográfico e documental à sua guarda, e promover a preservação de conteúdos disponíveis na Internet nacional, garantindo a disponibilização deste à comunidade científica e ao público em geral, sem prejuízo da coordenação exercida pela Secretaria – Geral do MEC. (PORTUGAL, 2013, p. 2258).

Tal atribuição de promover a preservação de conteúdos disponíveis na internet portuguesa é realizada pelo Arquivo.pt. No período entre 1996 e 2007, as páginas *web* eram recolhidas pelo *Internet Archive*, posterior a essa data, a instituição portuguesa passou a fazer as suas próprias recolhas da *web*. A instituição possuía mais de 4 milhões de páginas preservadas em 2017, fazendo recolhas de forma exaustiva, por domínio e de forma seletiva. Diferentemente de outros arquivos da *web*, oferece uma interface em português e inglês de pesquisas históricas de páginas e imagens, a partir de palavra-chave, possibilidade de computação automática dos dados arquivados para fins de investigação, e a mais comum entre os arquivos da *web*, a pesquisa por URL. A frequência das recolhas varia de 3 a 4 vezes ao ano, com um volume de em torno de 400 conteúdos *web* em português por dia, e a forma de acesso é *on-line* sem restrições (ARQUIVO.PT, 2017b, 2021c, 2021d).

Como o Arquivo.pt é ligado às áreas de pesquisa e tecnologia, tem como vocação ofertar: formação na área de preservação da *web*, “para que se torne consciência comum de que a informação em formato digital é um patrimônio valioso que tem de ser preservado” (ARQUIVO.PT, 2021c) dividido em módulos: a) Arquivo.pt: uma nova ferramenta para pesquisar o passado; b) Bem publicar, para bem preservar; c) Acesso e processamento automático de informação preservada da Web através de APIs; e d) Arquivar a Web: faça-você-mesmo. Além disso, publica artigos científicos e técnicos para divulgação dos seus serviços; e promove, por meio de um concurso anual, nomeado de Prêmio Arquivo.pt, trabalhos inovadores que utilizam informações preservadas da *web*. A última edição, em 2021, a premiação foi de € 10 mil para o vencedor, de € 3 mil para o segundo colocado e de € 2 mil para o terceiro lugar. O concurso está aberto a todos os interessados, de forma individual ou em grupo. Além disso, não há restrição de tema, mas é obrigatório o uso do arquivo português como fonte primordial de informação (ARQUIVO.PT, 2021b, 2021c, 2021g).

Ainda como forma de difundir o acervo *web* português, a entidade promove exposições *on-line*, agrupadas seja pelo critério institucional ou temático, uma forma de revisitar conteúdos do passado ou páginas atualmente indisponíveis (ARQUIVO.PT, 2021e).

A entidade também oferece um serviço chamado de memorial do Arquivo.pt. Este compreende-se como uma preservação de alta qualidade do conteúdo de *websites* que já não são atualizados por entidades públicas ou privadas, permitindo manter o nome do domínio original do *website*; o conteúdo do *website* pesquisável por intermédio dos motores de busca e evita *links* quebrados para páginas internas (ARQUIVO.PT, 2021f).

Vale ressaltar um vídeo institucional produzido em 2017 que menciona o custo de investimento para preservação de *websites*, que foi estimado naquele ano em € 216 milhões, superior ao PIB de Portugal em 2016, que ficou em € 181 milhões (ARQUIVO.PT, 2017b).

O Arquivo.pt preserva 8.925 milhões de conteúdos *web* recolhidos desde 1996. Outrossim, utiliza como *softwares* de recolha principalmente o *Heritrix*, *Brozzler* e/ou *arquivopatcher*, e como padrão de metadados, o Dublin Core. Em julho de 2020, eram 562 terabytes de arquivos comprimidos ARC e WARC em dois servidores diferentes. A instituição possuía um total de 84 servidores, e os dados de cada servidor são preservados em partições de discos configuradas em *Redundant Array of Inexpensive Drives*¹⁰⁴ – RAID 5 (mais rápido para identificar erros; leitura rápida e maior segurança) ou RAID 6 (possibilidade de falhar 2 HDs ao mesmo tempo sem perda de dados e maior segurança

¹⁰⁴ RAID. In: WIKIPEDIA: a enciclopédia livre. [São Francisco, CA: Wikimedia Foundation, 2021]. Disponível em: <https://en.wikipedia.org/wiki/RAID>. Acesso em: 14 dez. 2021.

comparado ao RAID 5), isso significa que utilizam subsistema de armazenamento composto por vários discos individuais, com a finalidade de ganhar segurança e desempenho, por meio da redundância de dados (BRANCO, 2020; RAID, 2021). A instituição também oferece o serviço “SavePageNow – Grava páginas no Arquivo.pt”, em que qualquer pessoa pode inserir a URL do *website* que deseja preservar (ARQUIVO.PT, 2021h).

Não foi localizado um documento de política de preservação digital, especificando quais os critérios utilizados pela instituição, mas puderam-se conhecer algumas estratégias por meio dos relatórios técnicos. Todavia, a FCT publicou seu Plano de Atividades 2021, mencionando que tem como objetivo “desenvolver, operar e manter a infraestrutura de investigação Arquivo.pt, preservando a *web* portuguesa” e que, no âmbito da comunicação institucional, assegurar “a gestão, implementação e desenvolvimento de meios e mecanismos de gestão documental e preservação digital do património documental e bibliográfico da FCT” (FTC, [2021b], p. 28 e 32). O Plano de Atividades de 2021 reforça o compromisso institucional com a preservação do património digital.

Durante a pesquisa, não foi localizado no *website* da instituição um documento de política de preservação digital, dessa forma, foi enviado e-mail para (contacto@arquivo.pt) solicitando mais informações. O Arquivo.pt afirma que o processo de preservar informação publicada na *web* é complexo e que, para organizar esse processo, são produzidos e mantidos alguns documentos técnicos internos que não são públicos e que também são compreensíveis apenas para especialistas em arquivos da *web*. Menciona que a preservação digital é uma atividade transversal a cada um dos serviços que a FCT-FCCN realiza a gestão (por exemplo, publicações científicas, dados abertos, publicações em acesso aberto, informação técnica de gestão da rede académica) porque cada tipo de dados requer abordagens de preservação digital específicas, de maneira que sejam adequadas ao contexto de cada serviço. Além disso, o Arquivo.pt dedica-se apenas à preservação digital de informação publicada na *web*.

8.2.2 Biblioteca Nacional da Espanha

A Biblioteca Nacional da Espanha – BNE, iniciou o trabalho do Arquivo da *web* espanhola¹⁰⁵ em 2009, ingressando como membro do IIPC em 2010; e o conteúdo recolhido tem a contribuição das bibliotecas das comunidades autónomas. Entre as comunidades

¹⁰⁵ <http://www.dl-e.es/openwayback/wayback/>

autônomas, a Catalunha foi pioneira na Espanha em arquivamento da *web*, iniciou em 2005 o PADICAT e, também, faz parte do IIPC (IIPC, 2021b; BNE, [c2021e]).

A BNE conta com o Real Decreto nº 635, de 10 de julho de 2015, que regula o depósito legal das publicações *on-line*, incluindo os *websites* do domínio ‘.es’, seja de acesso aberto ou restrito, assim como os conteúdos produzidos nos idiomas oficiais ou produzidos por pessoas ou entidades com permanência no território espanhol. A gestão do depósito legal dos *websites* ficará sob responsabilidade da BNE, que selecionará qual patrimônio documental *on-line* será preservado, e os editores de conteúdo facilitarão a recolha quando for solicitada (ESPANHA, 2015; BNE, [c2021b]). A BNE pode estabelecer acordo com entidades públicas e privadas para cumprir a missão do depósito legal. Em 2021, firmou o convenio com a *Agencia Estatal Boletín Oficial del Estado*, para a preservação em longo prazo desse material *on-line* (BNE, [c2021d]).

Inicialmente, a BNE contou com a ajuda do *Internet Archive*, entre 2009 e 2013, utilizando o *Heritrix* e *Wayback Machine*. Desde 2016, a instituição passou também a utilizar o *NetarchiveSuite*, construído em colaboração com as bibliotecas nacionais da Suécia, França e da Dinamarca. O arquivo *web* espanhol faz recolhas por domínio nacional e de forma seletiva, a sua interface nos idiomas inglês e espanhol, o acesso é restrito ao espaço físico da BNE e nas bibliotecas regionais (IIPC, 2021b; BNE, 2015, 2017, [c2021a], [c2021b], [c2021c]).

Foram apresentados, pelo atual diretor técnico da BNE, José Luiz Bueren, no evento da FBN, a 7ª Jornada de Pesquisadores da FBN, alguns dados referentes à preservação digital e ao arquivo *web* espanhol. A BNE conta com um sistema de preservação digital desde 2012, chamado LIBSAFE¹⁰⁶, da empresa espanhola LIBNOVA. São 900TB de objetos digitais preservados, somando em torno de 2 milhões de *websites* por ano (BUEREN, 2021; FBN, 2021).

De acordo com a LIBNOVA (c2016), o “Libsafe é um software de preservação digital concebido para a preservação a longo prazo de arquivos, dados e coleções digitais de elevado valor, tais como arquivos mestres de digitalização ou informação geológica, legal, médica, ou qualquer outro tipo de informação”, tendo por base o modelo OAIS e ISO 14721, contemplando o ciclo de vida preservando qualquer formato de dados com qualquer plano de preservação à sua escolha, pois o acervo pode ser preservado na instituição ou sob guarda

¹⁰⁶ LIBNOVA. *Libsafe: the software that makes digital preservation easy*. c2016. Disponível em: <http://www.digitalpreservationsoftware.com/digital-preservation-solutions/libsafe-digital-preservation-software/>. Acesso em: 15 fev. 2022.

externa, o módulo de recuperação é Open Source e os formatos para arquivo e documentação são padrão.

Durante a pesquisa, não foram localizados no *website* da instituição um documento de política de preservação digital ou estratégias nesse sentido; dessa forma, foi enviado *e-mail* para (archivoweb@bne.es) solicitando mais informações. A BNE confirmou que atualmente não possui um documento de política de preservação digital, contudo, estão trabalhando para que seja publicado ao longo de 2022.

Todavia, a instituição possui planejamento estratégico 2021-2025, afirmando que deve garantir a preservação das coleções digitais, por meio de sistema de preservação digital em longo prazo. Para garantir o crescimento das coleções da BNE como um centro depositário da produção cultural espanhola, esta pretende assegurar o ingresso de coleções via depósito legal com apoio e atualização da infraestrutura tecnológica; além de ampliação da colaboração com outras instituições nacionais e internacionais (BNE, [c2021f]).

8.2.3 Biblioteca da Catalunya

A Biblioteca da Catalunya compreende-se como uma biblioteca regional da Espanha e, ao mesmo tempo, como a Biblioteca nacional da Catalunya. É membro do IIPC e, também, integra como colaboradora o depósito legal espanhol, regido pelo Real Decreto nº 635/2015. A instituição criou o PADICAT – Patrimônio digital da Catalunya para capturar, preservar e disseminar a *web*, contando com o apoio tecnológico do *Consorci de Serveis Universitaris de Catalunya* (CSUC) para preservar e dar o acesso a antigas versões de páginas *web* publicadas na internet, desde 2005 (PADICAT, [c2011a]; IIPC, [c2021b]; BIBLIOTECA DA CATALUNYA, [c2021a]).

O escopo de recolha do PADICAT é por domínio regional ‘.cat’, seletivo, eventos e temáticos, com a interface em três idiomas: catalão, espanhol e inglês. O objetivo do PADICAT é fazer a coleta massiva do domínio, promover de forma sistemática o depósito dos conteúdos *web* de entidades e empresas catalãs; por fim, também promover que esses conteúdos sejam utilizados pelos usuários. Tais objetivos contam com quatro linhas de trabalho permanente: a) com a definição de estratégias de preservação digital do patrimônio provenientes da internet, por exemplo, o PADICAT fornece radiografias periódicas da *web* catalã; detecta formatos que experimentam problemas de ilegibilidade a curto prazo; identifica as línguas mais utilizadas etc.; b) a promoção de linhas de investigação baseadas na criação de coleções monográficas com a participação de especialistas em cada tema; c) a criação e a

manutenção da biblioteca de jornais digitais na internet, capturado de forma sistematizada as publicações digitais em série; d) cooperar com outros arquivos da *web* e repositórios de preservação digital, bibliotecas, arquivos e museus, visando dar uma resposta eficiente aos desafios da preservação digital e ao acesso aos recursos depositados (PADICAT, [c2011a]; IIPC, [2021a]).

O PADICAT utiliza uma combinação de *softwares* para recolher, indexar e gerenciar os conteúdos *web*, entre eles, estão: *Heritrix*, *NutchWax*, *Hadoop*, *Wayback*, *Wera* e *Web Curator Tool* (LLUECA *et al.*, 2010). O *hardware* que suporta o sistema consiste em seis HP ProLiant DL360 G4p, que são responsáveis pela recolha e indexação das páginas *web*. A pesquisa e a exibição dos resultados na interface *web* são por intermédio do *software* Linux de alta disponibilidade com características de balanceamento de carga e tolerância a falhas no caso de um desastre técnico. Uma caixa FAS3170 da NetApp apresenta 19TB de espaço em disco via NFS, ligados por meio de fibra a uma Rede de Área de Armazenamento (SAN), e o sistema é completado por um robô em que as cópias de segurança dos dados são armazenadas em fita. Em relação à preservação digital, de acordo com o planejamento estratégico de 2017-2021, a Biblioteca da Catalunya coletou 313.048 de 72.663 *websites*, somando um total de 35TB de patrimônio digital preservado. A instituição possui três linhas de trabalho de preservação digital: a preservação do arquivo da *web*, ou seja, PADICAT; COFRE (COnservado para Futuros Recursos Eletrônicos), configura-se como um repositório de preservação digital; e a preservação digital distribuída conforme Real Decreto nº 635/2015, de 10 de julho, que deve obedecer requisitos mínimos, como sistema de recuperação de dados; monitoramento de obsolescência; integridade dos dados; segurança da informação e metadados (BIBLIOTECA DA CATALUNYA, [c2021b],[c2021c]; PADICAT, [c2021b]).

Não foi localizado no *website* da instituição um documento de política de preservação digital, apenas alguns documentos, cujo *link* não estava funcionando; dessa maneira, foi enviado *e-mail* para (padicat@bnc.cat) solicitando mais informações. Foram enviados dois *links* para documentos institucionais, o relatório “*Memòria del plantejament del projecte PADICAT (Patrimoni Digital de Catalunya) Biblioteca de Catalunya*, dez. de 2005”, apresentando as questões para iniciar o projeto PADICAT; e outro relatório descritivo de 2010, “*Repositori de Preservació Digital de la Biblioteca de Catalunya*”, produzido por Karibel Pérez e Eugènia Serra, membros do grupo de preservação digital da biblioteca.

Nesse último documento, é descrito que a Biblioteca da Catalunya utiliza seu repositório de preservação com base no modelo OAIS, com o desenvolvimento personalizado pela área de tecnologia da informação em uma plataforma de servidor *web* Apache 2 e base de

dados PostgreSQL 8; a aplicação foi desenvolvida com PHP 5.3 (PÉREZ; SERRA, 2010, p. 15). O esquema de metadados é o Dublin Core para gestão dos metadados descritivos e administrativos. E para os metadados técnicos e de preservação, foi definido utilizar o mapa BC-PREMIS para criar os metadados que irão incorporar os identificadores de documentos. Com objetivo de garantir a preservação em longo prazo, foi necessário assegurar a manutenção de uma série de qualidades que, segundo as autoras, são a fixidez, a integridade e a autenticidade, qualidades que são alcançadas por meio da aplicação de uma série de rotinas e práticas de “identificação de formatos; verificação de vírus; integridade dos arquivos; autenticidade e monitoramento de acesso aos documentos e pastas; segurança de rede e *backups*”. Tais tarefas são realizadas no repositório por intermédio de áreas de gerenciamento de metadados; gestão de rotinas e ações de preservação; gerenciamento e segurança de usuários; gerenciamento de conteúdo: *uploads*; gestão do Depósito Legal de documentos nascidos digitais; gerenciamento de cópias de alta qualidade; e caixa preta, área restrita de gestão interna com controle de IP (PÉREZ; SERRA, 2010, p.17-23).

8.2.4 Biblioteca Nacional do Chile

A Biblioteca Nacional do Chile – BNC associou-se ao IIPC em 2014, sendo a primeira instituição da América do Sul a fazer parte do consórcio. Possui como marco legal a Lei nº 19.773, de 23 de dezembro de 2013, que trata “sobre liberdade de opinião e informação e prática do jornalismo”. Inserindo o suporte eletrônico no art. 14, tal legislação modificou a Lei nº 20.709/2013, que tratava de “normas sobre depósito legal de creaciones audiovisual” (BIBLIOTECA NACIONAL DE CHILE, [2021a], [2021b], IIPC, [2021b]; CHILE, 2013).

Em 2013, foram então iniciadas, por meio de um projeto-piloto, as primeiras recolhas da *web* chilena, reunindo coleções que procuram representar os acontecimentos daquele país e, também, relacionadas com a missão institucional da BNC. O arquivo da *web* chilena configura-se como um serviço oferecido pela biblioteca nacional digital com objetivo de “armazenar e preservar vários *websites* nacionais, a fim de assegurar a disponibilidade de informações e conhecimentos gerados em formato digital, no caso do seu eventual desaparecimento” (BIBLIOTECA NACIONAL DE CHILE, [2021a], [2021b], [2021c]).

A instituição utiliza o *software Heritrix*, WGET, ferramenta baseada em Linux para descarregar conteúdos em protocolos HTTP, HTTPS e FTP. O acesso aos conteúdos recolhidos é restrito aos pontos de acesso no Chile, para o cumprimento da Lei nº 17.336/2017 de propriedade intelectual que estabelece que “nenhuma obra de qualquer tipo

pode ser divulgada sem o consentimento expresso e autorização do(s) seu(s) autor(es)”. No entanto, a lei, em seu art. 71 K, estabelece que as bibliotecas e arquivos podem oferecer o acesso aos documentos eletrônicos para consulta gratuita e, simultaneamente, por um número razoável de usuários. Dessa forma, as recolhas de *websites* que não pertencem ao Serviço Nacional do Patrimônio Cultural (SNPC) só podem ser vistas a partir de uma série de pontos físicos (BIBLIOTECA NACIONAL DE CHILE, [2021a], [2021b], [2021c]; CHILE, 2017).

Em relação à preservação digital, a instituição segue as normas internacionais e as recomendações do IIPC, a carta de preservação da Unesco e a ISO 30300 (gestão de documentos eletrônicos), e documentos internos, como o Manual de procedimento para o depósito legal dos recursos eletrônicos e de Gestão de coleção digital. De acordo com Roberto Bello, do Departamento de coleções especiais e digitais, todavia, não há um documento de política de preservação digital em vigor e que estão reformulando a área digital da instituição (BIBLIOTECA NACIONAL DE CHILE, [2021d]; BELLO, [2021]).

Não foi localizado no *website* da instituição um documento de política de preservação digital, dessa maneira, foi solicitada a informação por meio do fale conosco¹⁰⁷ da biblioteca. De acordo com o Departamento de coleções especiais e digitais, o trabalho de preservação digital é realizado por esse departamento, dividido também com o Laboratório de digitalização e com a Unidade de sistemas de informação bibliográfica, que tem um setor de engenharia informática. Outrossim, pelo fato de o Arquivo da *Web* do Chile ser um projeto-piloto, ao longo dos anos, tem sofrido de vários problemas devido à sua complexidade e custo de financiamento. Por esse motivo, confirmou-se que ainda não possuem documentação escrita sobre políticas, planos, estratégias ou manuais sobre preservação digital. Mencionaram que, atualmente, estão realizando o processo de migração das salas de servidores para um serviço externo, obrigando a travar o crescimento das coleções. A BNC não recolhe o domínio ‘cl.’, apenas faz a coleta de forma seletiva de certos sítios que são definidos com base na política geral de desenvolvimento de coleções da Biblioteca.

8.3 DISCUSSÃO DOS RESULTADOS

Após identificação, caracterização e acesso aos documentos institucionais das instituições membros do IIPC, elaborou-se um quadro com intuito de sistematizar as informações por meio de categorias. A partir desse panorama e tendo acesso aos documentos,

¹⁰⁷ <http://www.bibliotecanacionaldigital.gob.cl/bnd/612/w3-propertyvalue-783237.html>

foi realizada a análise. Para estudar os *websites*, foi utilizada a técnica de análise de conteúdo e análise documental, que, segundo Bardin¹⁰⁸ (1977, p. 42-45), configuram-se como:

[...] Análise de conteúdo – um conjunto de técnicas de análise das comunicações visando obter, por procedimentos, sistemáticos e objetivos de descrição do conteúdo das mensagens, indicadores (quantitativos ou não) que permitam a inferência de conhecimentos relativos às condições de produção/recepção (variáveis inferidas) destas mensagens.

[...] Análise documental – uma operação ou conjunto de operações visando representar um conteúdo de um documento sob uma forma diferente do original, a fim de facilitar num estado ulterior, a sua consulta e referência.

Bardin (1977, p. 95) afirma que a análise se organiza em três fases. Primeiro, é realizada uma pré-análise que se “caracteriza com a escolha dos documentos, a formulação de hipóteses e a elaboração de indicadores que irão embasar a interpretação final”. Foram definidas as 11 instituições membros fundadoras e 4 instituições ibero-americanas que permanecem associadas ao IIPC.

A segunda fase é a exploração do material, em que os *websites* do IIPC e das instituições analisadas foram acessados, visando criar categorias. Glushko¹⁰⁹ (2020) afirma que, em contraste com as categorias culturais, que são criadas e utilizadas implicitamente, e as categorias individuais, que são utilizadas por pessoas de forma solitária,

[...] as categorias institucionais são criadas e utilizadas explicitamente, e na maioria das vezes por muitas pessoas em coordenação umas com as outras. As categorias institucionais são mais frequentemente criadas em domínios abstratos e de informação intensiva onde são necessárias categorias inequívocas e precisas para regular e sistematizar a atividade, para permitir a partilha e reutilização de informação, e para reduzir os custos de transação. (GLUSHKO, 2020, p. 1, tradução nossa).

Já a terceira fase é o tratamento dos resultados, quando é possível visualizar os dados em forma de quadros ou representações gráficas, sendo possível condensar e pôr em destaque as informações fornecidas pela análise.

O quadro 10 apresenta a caracterização institucional dos onze membros fundadores, somados às respectivas instituições que dividem a tarefa de arquivamento da *web* naquele país.

¹⁰⁸ BARDIN, L. *Análise de conteúdo*. Lisboa: Edições 70, c1977.

¹⁰⁹ Institutional Categories – In contrast to cultural categories that are created and used implicitly, and to individual categories that are used by people acting alone, institutional categories are created and used explicitly, and most often by many people in coordination with each other. Institutional categories are most often created in abstract and information-intensive domains where unambiguous and precise categories are needed to regulate and systematize activity, to enable information sharing and reuse, and to reduce transaction costs (GLUSHKO, 2020, p. 1).

O quadro 11 lista os quatro membros ibero-americanos. Ambos os quadros foram sistematizados em dez categorias: nome; tipo de instituição; país; legislação; ano de início do arquivamento da *web*; método de coleta; ferramenta de coleta; acesso; tipo de pesquisa e idioma de interface do arquivo *web*. Tais categorias foram consideradas a partir da pré-análise realizada.

Quadro 10 – Caracterização das Instituições Membros do IIPC

CARACTERIZAÇÃO INSTITUCIONAL DOS MEMBROS DO IIPC									
Instituição Membro do IIPC	Tipo de Instituição	País	Legislação	Ano de Início do arquivamento da web	Método de Coleta	Ferramenta de Coleta	Acesso	Tipo de Pesquisa	Idioma de interface do Arquivo web
1-Bibliothèque nationale de France e (INA*)	Biblioteca Nacional e Instituto de Audiovisual	França	DL de 2006 dividido por duas instituições Código do Patrimônio (artigos L131-1 a L133-1 e R131-1 a R133-1), complementado por decretos de 1995, 1996 e 2006. Lei nº 2006-961, de 1º agosto de 2006, e Decreto de 19 de dezembro de 2011.	2002 e 2009	Domínio nacional, em massa, seletivo, evento, temático	Heritrix, Wayback, NutchWAX, NetarchiveSuit, BCWeb.	Restrito ao Local	Pesquisa de URL, Pesquisa de palavra-chave, Pesquisa de texto completo, Coleções de tópicos	Francês
2-Internet Archive	Organização sem fins lucrativos	EUA	Não possui	1996	Domínio Nacional, Regional, Massivo, Seletivo, Evento, Temático	Heritrix, Wayback, Archive-It	Acesso on-line	Pesquisa URL, Coleções Tópicas	Inglês
3-The British Library	Biblioteca Nacional	UK	DL de 2013 divididos entre 6 instituições. A Lei das Bibliotecas de Depósito Legal (2003) alargou a Legislação existente a publicações não impressas (eletrônicas), incluindo sítios Web, sujeito a novos	2004 (UKWA)	Domínio Nacional, Seletivo, Evento, Temático, Periódico	Heritrix, Web Curator Tool, Wayback, Mirror web (https://github.com/ukwa/)	Restrito ao local	Pesquisa URL, Navegação alfabética, Pesquisa por texto completo (1996 – 2013), Assunto, Coleta especial	Inglês, Galês e Escocês.

			Regulamentos de habilitação, promulgados em 2013						
4-The National Archives, U.K.(*)	Arquivo Nacional	UK	Public Records Act 1958 and Section 163 of the Copyright Designs and Patents Act 1988	1996 (UKGW)	Seletivo, Evento, Temático	[Heritrix]	Acesso on-line	Texto completo e pesquisa de URL, navegação alfabética, navegação por assunto, coleções de tópicos	Inglês
5-Det Kongelige Bibliotek (The Royal Danish Library)	Biblioteca Nacional	Dinamarca	DL de 2004, Lei nº 1439 de 22 de dezembro de 2004 de depósito legal, do Ministério da Cultura	2005	Domínio nacional (transversal), seletiva, eventos, especiais e colaborativas	NetarchiveSuite	Restrito ao Local	Busca URL (Open Wayback), busca de texto livre (Blacklight)	Dinamarquês, Inglês, Alemão, Italiano, Francês, Espanhol
6-Kansalliskirjasto (The National Library of Finland)	Biblioteca Nacional	Finlândia	DL de 2007, Lei nº 1433, de 28 de dezembro de 2007 dividido entre 5 instituições	2006	Domínio nacional, domínio regional, evento, temático	Heritrix; Wayback	Restrito ao Local	pesquisa de URL, pesquisa de texto completo	Finlandês, Sueco, Inglês
7-Kungliga biblioteket (National Library of Sweden)	Biblioteca Nacional	Suécia	DL 2012, The Swedish Act (2012:492). No entanto, o marco legal não abrange páginas web, bases de dados, mídias sociais ou outros meios de comunicação interativos	1997	Domínio nacional, em massa, seletivo	NetarchiveSuite	Restrito ao Local	pesquisa por URL	Sueco
8-Landsbókasafn Íslands – (National and	Biblioteca Nacional e Universitária	Islândia	DL 2003, Legislation on Legal Deposits no 20/2002; Regulations 982/2003	1996 (Internet Archive) 2004	Domínio nacional, seletivo, evento	Heritrix; Wayback	Acesso limitado	Pesquisa por URL	Islandês, tradução limitada em Inglês

University Library of Iceland)									
9-Nasjonalbiblioteket (National Library of Norway)	Biblioteca Nacional	Noruega	DL 1990, atualizada em 2016 regulamentada 2018 Ministério da Cultura em 1 de julho de 2018 nos termos da Lei nº 32, de 9 de junho de 1989.	2001	Domínio nacional, evento, seletiva	"Veidemann"; Heritrix e OpenWayback	Restrito ao Local	Pesquisa por palavra-chave	Norueguês
10-Library and Archives Canada	Biblioteca e Arquivo Nacional	Canadá	Library and Archives of Canada Act (S.C. 2004, c. 11), aprovada em 22/04/2004 e o Regulamento de Depósito Legal de Publicações (SOR / 2006-337)	2005	Domínio nacional, evento, seletiva/ governamental, evento/ temático	Heritrix	Sem acesso ao público	Navegação alfabética por nome do departamento do governo federal e URL /	Inglês, Francês
11-National Library and Archives of Quebec (*)	Biblioteca Regional	Canadá	Marco legal de depósito legal sancionado em 2001, que respeita à Bibliothèque et Archives nationales du Québec, foi atualizada em 1 de outubro de 2021.	2012	governamental, seletivo, evento/temático	Heritrix; Wayback	Acesso on-line	Pesquisa por URL, Navegação alfabética	Inglês, Francês
12-Library of Congress	Biblioteca Nacional	EUA	Não possui	2000	Seletivo, Evento, Temático	Heritrix; Wayback	Acesso on-line	Pesquisa por URL, Navegação alfabética, Navegação por assunto, Coleções por tópico	Inglês

13-National Library of Australia	Biblioteca Nacional	Austrália	Copyright Act 1968 atualizada desde 17 de fevereiro de 2016	1996	Seletivo, domínio inteiro, governo de lista de sementes em massa, eventos	Heritrix, HTTrack ou Web Record	Acesso on-line ao público depende da licença do editor	Pesquisa de URL, palavras-chave, frases de texto completo, opções de pesquisa avançada (incluindo limitação à pesquisa de material governamental, domínios, tipos de arquivo e intervalos de datas), navegação em coleção com curadoria, conteúdo selecionado catalogado	Inglês
----------------------------------	---------------------	-----------	---	------	---	---------------------------------	--	--	--------

Fonte: Elaborada pela autora.

As instituições com o símbolo (*) não fundaram o IIPC, mas participam do consórcio dividindo a tarefa de preservação com membros fundadores nos seus respectivos países.

Quadro 11 – Caracterização das Instituições Ibero-Americanas do IIPC

CARACTERIZAÇÃO INSTITUCIONAL									
Instituição Membro do IIPC	Tipo de Instituição	País	Legislação	Ano de Início do arquivamento da web	Método de Coleta	Ferramenta de Coleta	Acesso	Tipo de Pesquisa	Idioma de interface do Arquivo web
Arquivo.pt – Portuguese Web Archive	Rede Nacional de Educação e Ensino	Portugal	Termos e Condições Última atualização em 19 de setembro de 2017. Decreto-Lei nº 55/2013 e DL nº 7/2004, de 07 de janeiro.	2008	Domínio nacional, seletivo	Heritrix, Brozzler e/ou arquivo patcher	Acesso on-line	Pesquisa de texto completo e de imagens, lista de histórico de versões, pesquisa avançada e interfaces de programação de aplicações (API) que facilitam o desenvolvimento de aplicações de valor acrescentado por terceiros	Português e Inglês
Biblioteca Nacional da Espanha	Biblioteca Nacional	Espanha	Real Decreto 635, de 10 de julho de 2015	2009	Domínio nacional, seletivo e temático	Heritrix, Wayback e NetarchiveSuite	Restrito ao local	Busca URL	Espanhol e Inglês
Biblioteca de Catalunya – PADICAT	Biblioteca Regional	Espanha	Real Decreto 635, de 10 de julho de 2016	2005	Domínio regional ‘.cat’, seletivo, eventos e temáticos	Heritrix, NutchWax, Hadoo, Wayback, Wera e Web Curator Tool	Acesso on-line	Pesquisa de URL, Pesquisa por palavra-chave, Navegação alfabética, Navegação por assunto, Coleções de tópicos	Catalão, Espanhol, Inglês
Biblioteca Nacional do Chile	Biblioteca Nacional	Chile	Lei 19773/ 2013 e Lei 17336/2017	2013	Seletivo	Heritrix	Restrito ao local	[Busca por URL]	Espanhol

Fonte: Elaborado pela autora.

Nos quadros 10 e 11, as instituições são caracterizadas utilizando-se dez categorias, o que possibilitou visualizar suas semelhanças e diferenças. Destaca-se que, na categoria “tipo de instituição” (quatorze) as bibliotecas nacionais são maioria (dez), sendo que seis delas dividem o trabalho com outras instituições, como bibliotecas regionais, bibliotecas universitárias ou arquivos nacionais. Observa-se que, em Portugal, o trabalho não é realizado por instituições de memória como bibliotecas e arquivos, mas por uma instituição ligada área de tecnologia e pesquisa.

Na categoria “país”, onze instituições têm sede no continente europeu, cinco no continente americano e somente uma na Oceania. Em relação à categoria “legislação”, apenas três instituições não possuem marco legal ligado ao depósito legal, as instituições dos EUA e a de Portugal.

Na categoria “ano de início do arquivamento da *web*”, em que o período abarca os anos de 1996 a 2013, a primeira instituição é a *Internet Archive* em 1996, e a mais recente, a BNC, que iniciou o trabalho em 2013. A grande maioria das instituições lançou os projetos dentro da temática nos anos 2000, impulsionados principalmente pela publicação da *Carta sobre a Preservação do Patrimônio Digital* da Unesco. Nas categorias “método de coleta” e “ferramenta de coleta”, a maioria faz a captura/recolhimento por domínio nacional, seletivo ou temático, utilizando ou baseando-se no *software Heritrix*.

Na categoria “acesso”, as instituições da França, da Inglaterra, da Noruega, da Espanha, da Dinamarca, da Finlândia, da Suécia e do Chile possuem acesso restrito ao local físico, similar ao modelo de depósito legal de publicações impressas, e uma instituição, a do Canadá, não fornece acesso ao público. As demais instituições fornecem o acesso *on-line* ou de forma limitada, de acordo com o marco legal de cada país. A categoria acesso está implicitamente ligada a mais duas categorias, “tipo de pesquisa” e “idioma de interface do arquivo *web*”; a maioria oferece pesquisa por URL, palavras-chave ou texto completo, sendo que a instituição portuguesa oferece um recurso que as outras não mencionam, a opção de buscar por imagens. O idioma de interface utilizado pelas instituições é o inglês somado a outro e oito delas só fornecem uma opção de idioma, isto é, o seu de origem.

Finalizada a análise das categorias provenientes da caracterização institucional, o passo seguinte foi analisar os documentos institucionais, em que se considerou o “Enfoque na governança institucional” relacionado com a preservação digital por cada uma delas, com intuito de responder aos objetivos específicos C e D. Tal análise baseou-se em dois de três aspectos levantados por Grácio (2012) sobre a preservação digital. Quanto ao primeiro aspecto, o organizacional, no qual consta “o objetivo institucional, que envolve uma equipe

multidisciplinar, trata também das responsabilidades, dos recursos financeiros e dos atos administrativos da instituição”, acredita-se que os documentos institucionais publicados estejam abarcados nesse universo. E o terceiro são os aspectos técnicos – “à seleção, descarte; modelos, padrões e iniciativas (metadados, OAIS), suporte, infraestrutura tecnológica (*hardware* e *software*); estratégias de preservação (migração, emulação)”. Ressalta-se que o segundo aspecto se refere “à legislação, direito do autor e/ou mandato legal”; que já foi abordado nos quadros 10 e 11 com a caracterização institucional (GRÁCIO, 2012, p. 83, 97 e 109).

Com base no quadro 11 de caracterização das Instituições Ibero-Americanas do IIPC, as quatro instituições analisadas foram contactadas via *e-mail* (vide APÊNDICE A) solicitando mais informações que não foram localizadas em seus *websites* e a fim de identificar as suas políticas, ações, critérios de preservação digital e os problemas e barreiras à preservação da *web*. Conforme quadro 12 a seguir:

Quadro 12 – Endereço de contato das instituições do IIPC

Instituições Ibero-Americanas Participantes do IIPC	e-mail / Fale conosco
Arquivo.pt (Portuguese Web Archive)	contacto@arquivo.pt
National Library of Spain	archivoweb@bne.es
Library of Catalonia (PADICAT)	padicat@bnc.cat
Chilean National Library	http://www.bibliotecanacionaldigital.gob.cl/bnd/612/w3-propertyvalue-783237.html

Fonte: Elaborado pela autora.

A partir disso, elaborou-se o quadro 13 de Enfoque na governança institucional, que discorre sobre as categorias de aspectos organizacionais e aspectos técnicos, utilizados pelas instituições ibero-americanas do IIPC; baseando-se nas respostas recebidas, nos documentos e informações dos seus *websites*.

Quadro 13 – Enfoque na Governança Institucional

ENFOQUE NA GOVERNANÇA INSTITUCIONAL			
Instituição Membro do IPC	Aspecto organizacional – Políticas/ Manuais/ Planejamento Estratégico	Aspectos Técnicos – Estratégias de Preservação	Observações
Arquivo.pt – Portuguese Web Archive	Plano de atividades 2021 da FTC menciona que tem como objetivo “desenvolver, operar e manter a infraestrutura de investigação Arquivo.pt, preservando a web portuguesa” e que no âmbito da comunicação institucional assegurar “a gestão, implementação e desenvolvimento de meios e mecanismos de gestão documental e preservação digital do património documental e bibliográfico da FCT” (FTC, [2021b], p. 28 e 32) https://www.fct.pt/docs/PA2021.pdf Durante a pesquisa não foi localizado no <i>website</i> da instituição um documento de política de preservação digital, desta forma foi enviado e-mail solicitando mais informações. A resposta da instituição foi que o processo de preservar informação publicada na web é complexo e que para organizar este processo, são produzidos e mantidos alguns documentos técnicos internos que não são públicos e que também são compreensíveis apenas para especialistas em arquivos da web.	O Arquivo.pt preserva 8.925 milhões de conteúdos web recolhidos desde 1996, utiliza como softwares de recolha principalmente o Heritrix, Brozzler e/ou arquivopatcher, utiliza o Dublin Core como padrão de metadados. Em julho de 2020, eram 562 Terabytes de arquivos comprimidos ARC e WARC em dois servidores diferentes, a instituição possuía um total de 84 servidores, e os dados de cada servidor são preservados em partições de discos configuradas em Redundant Array of Inexpensive Drives – RAID 5 (mais rápido para identificar erros; leitura rápida e maior segurança) ou RAID 6 (possibilidade de falhar 2 HDs ao mesmo tempo sem perda de dados e maior segurança comparado ao RAID 5), significa que utilizam subsistema de armazenamento composto por vários discos individuais, com a finalidade de ganhar segurança e desempenho, por meio da redundância de dados (BRANCO, 2020; RAID, 2021).	https://github.com/arquivo
<i>Biblioteca Nacional da Espanha</i>	Pretende elaborar e publicar sua política de preservação em 2022. Planejamento Estratégico 2021-2025 2.4.2 Projetar e implementar um plano de preservação digital de longo prazo para garantir a preservação e transmissão das coleções digitais (BNE, [c2021], p. 27) http://www.bne.es/export/sites/BNWEB1/webdocs/LaBNE/Transparencia/Institucional/2021-plan-estrategico-bne.pdf	LIBSAFE é utilizado como o sistema de preservação digital desde 2012. Tendo por base o Modelo OAIS e ISO 14721, contemplando do o ciclo de vida preservando qualquer formato de dados com qualquer plano de preservação à sua escolha, pois o acervo pode ser preservado na instituição ou sob guarda externa, o módulo de recuperação é Open Source e os formatos para arquivo e documentação são padrão. Gestão: de Metadados PREMIS, XML, planos de preservação, formatos, auditoria e controle de qualidade, armazenamento com redundância (LIBINOVA, 2014).	https://github.com/thoth-pub/thoth/wiki/National-Library-of-Spain
<i>Biblioteca da Catalunya – PADICAT</i>	Planejamento estratégico 2017-2021. Aumentar o espaço de armazenamento físico e virtual para ter capacidade de armazenamento suficiente para que a coleção cresça até 2025 (BC, [c2021b], p. 13). https://www.bnc.cat/Coneix-nos/Qualitat-i-estrategia/Pla-estrategic	Modelo OAIS, Dublin Core, PREMIS, identificação de formatos; verificação de vírus; integridade dos arquivos; autenticidade e monitoramento de acesso aos documentos e pastas; segurança de rede e backups (PÉREZ; SERRA, 2010, p.17- 20). Repositori de preservació digital de la Biblioteca de Catalunya: informe descriptiu i de situació: Redactat per Karibel Pérez i Eugènia Serra del Grup de Preservació Digital de la BC 2010	

		<p>http://hdl.handle.net/2072/97251</p> <p>Memòria del plantejament del projecte PADICAT (Patrimoni Digital de Catalunya) https://recercat.cat/bitstream/handle/2072/1757/IN_PADICAT_2005.pdf?sequence=1</p> <p>Infraestructura y funcionamiento https://www.padicat.cat/es/conocenos/infraestructura-y-funcionamiento</p> <p>Em relação à preservação digital, de acordo com o plano estratégico de 2017-2021, a BC coletou 313.048 de 72.663 <i>websites</i>, somando um total de 35TB de patrimônio digital preservado. A instituição possui três linhas de trabalho de preservação digital: a preservação do arquivo da web, ou seja, PADICAT (Patrimônio Digital da Catalunha); COFRE (CONservado para Futuros Recursos Eletrônicos), configura-se como um repositório de preservação digital; e a preservação digital distribuída conforme Real Decreto 635/2015, de 10 de julho, que deve-se obedecer requisitos mínimos, como sistema de recuperação de dados; monitoramento de obsolescência; integridade dos dados; segurança da informação e metadados (BIBLIOTECA DA CATALUNYA, [c2021b],[c2021c]; PADICAT, [c2021b]).</p>	
<i>Biblioteca Nacional do Chile</i>	Confirmou-se que ainda não possuem documentação escrita sobre políticas, planos, estratégias ou manuais sobre preservação digital.	<p>Não há um documento que contemple categoria, apenas são citados alguns documentos que “são parte dos esforços feitos pela instituição no âmbito da sua Política Digital”, tais documentos: ISO 30300 – uma nova abordagem para a gestão de registros; Recomendações da IFLA sobre Informação e Literacia da Mídia / IFLA.org; Depósito legal eletrônico, Biblioteca Nacional do Chile; Manual de Procedimentos para o Depósito Legal de Recursos Eletrônicos; e Gerenciamento de coleções digitais (BIBLIOTECA NACIONAL DE CHILE, [2021d]). http://www.bibliotecanacionaldigital.gob.cl/bnd/612/w3-propertyvalue-174087.html</p>	

Fonte: Elaborado pela autora.

Percebeu-se que, na categoria “Aspectos Organizacionais – Políticas / Manuais / Planejamento Estratégico”, atualmente, as quatro instituições ibero-americanas não possuem ou disponibilizam um documento de política de preservação digital. As duas instituições espanholas (BNE e BC-PADICAT) contam com um documento de planejamento estratégico recente, em que reforçam o compromisso com a preservação digital: “2.4.2 Projetar e implementar um plano de preservação digital de longo prazo para garantir a preservação e transmissão das coleções digitais” (BNE, [c2021], p. 27) e “aumentar o espaço de armazenamento físico e virtual para ter capacidade de armazenamento suficiente para que a coleção cresça até 2025” (BC, [c2021b], p. 13). A instituição portuguesa (Arquivo.pt) está subordinada à FCT e, conseqüentemente, ao seu plano de atividades. A instituição chilena (BNC) confirmou por *e-mail* que ainda não possui documentação escrita sobre políticas, planos, estratégias ou manuais referentes à preservação digital.

Já na categoria “Aspectos Técnicos – Estratégias de Preservação”, todas as instituições promovem, de acordo com suas necessidades e recursos (financeiros, pessoal, infraestrutura), ações para garantir o acesso em longo prazo dos documentos. Tal categoria remete ao quarto axioma de Owens (2018), no qual o autor salienta que “a preservação é o resultado de um trabalho contínuo das pessoas e compromissos de recursos”, pois “uma instituição que pensa em preservação digital terá que ter recursos fixos anuais alocados para a realização de um programa contínuo nesta área” (OWENS, 2018, p. 5). Por meio do *website* do Arquivo.pt, pôde-se localizar alguns dos relatórios técnicos sobre as atividades realizadas, como: uso de ferramentas de redundância, padrão de metadados Dublin Core, equipe de técnicos com perfis diversificado, além de ser a instituição que fornece o acesso livre e promove os conteúdos preservados por meio de prêmios, cursos e eventos, pontos importantes como acesso/uso/reuso tão preconizados no modelo do ciclo de vida do arquivamento da *web* Bragg e Hanna (2013).

A BNE utiliza, desde 2012, o LIBSAFE, uma ferramenta privada de sistema de preservação digital para fazer a gestão das suas coleções digitais. O sistema conta com as seguintes características técnicas: área de planos de preservação, na qual se incluem a definição de formatos de metadados, controles de ingestão, processos de auditoria, análise e conversão de formatos de arquivos para aplicação em grupo de objetos digitais; preservação de qualquer formato de objeto digital, identificando e verificando a integridade dos formatos; ingestão e controle de qualidade; inclusão e gestão de metadados, tendo a capacidade de inserir qualquer grupo de metadados encapsulado em XML e convertendo-o em qualquer esquema compatível através do XML/XSLT. No momento da recuperação de uma cópia do

objeto preservado digitalmente, gera-se o METS associado, utilizando os metadados originais ou os últimos que foram armazenados são atualizados. Além disso, faz a detecção de objetos duplicados solucionando conflitos, mesmo entre diferentes esquemas de metadados. Há também o controle de cópias, sem limitação do tipo ou localização do armazenamento. Conta com a área de auditoria e controle, realizando as tarefas de forma automática e programada; acesso, catálogo e recuperação, com uma interface de pesquisa para pré-visualização ou impressão, para a recuperação de objetos específicos; e por fim uma área contendo relatórios e estatísticas (LIBNOVA, 2014).

O armazenamento dos objetos preservados é realizado em dispositivos próprios do cliente, no caso, a BNE, para que o controle e a segurança sejam maximizados, sendo compatível com quase todos os sistemas de armazenamento, incluindo gestão de redundância e segurança (RAID) ou localização física dos dispositivos, desde que a conectividade da banda larga seja adequada; ainda permite combinar diferentes tipos de armazenamento sob o mesmo plano de preservação (LIBNOVA, 2014). Outrossim, Schisler (2020, p. 801) afirma que “a indústria no setor digital procura vender soluções para os nossos problemas, com sistemas, nuvens, *softwares* especializados” para salvaguardar dados digitais, no entanto, faz-se necessário o conjunto de recursos apropriados para realizar a preservação digital.

Já no PADICAT, utilizam em seu repositório de preservação digital diferentes áreas de gerenciamento com base no modelo OAIS; esquema de metadados Dublin Core e PREMIS; identificação de formatos; verificação de vírus; integridade dos arquivos; autenticidade e monitoramento de acesso aos documentos e pastas; segurança de rede e *backups* (PÉREZ; SERRA, 2010). Além disso, Owens (2018) salienta no primeiro axioma que “um repositório não é apenas um *software*, é a soma de recursos financeiros, *hardware*, pessoal tempo, e a implementação contínua de políticas e planejamento para assegurar a longo prazo acesso ao conteúdo” (OWENS, 2018, p. 4).

Na BNC, não há um documento que contemple a categoria “Aspectos Técnicos – Estratégias de Preservação”, apenas são citados no *website* “Acerca da BND – Biblioteca Nacional Digital do Chile”; e, especificamente, na área nomeada como “Preservação digital” alguns documentos que “são parte dos esforços feitos pela instituição no âmbito da sua Política Digital” (ISO 30300 – uma nova abordagem para a gestão de registros; Recomendações da IFLA sobre Informação e Literacia da Mídia / IFLA.org; Depósito legal eletrônico, Biblioteca Nacional do Chile; Manual de Procedimentos para o Depósito Legal de Recursos Eletrônicos; e Gerenciamento de coleções digitais) (BIBLIOTECA NACIONAL DE CHILE, [2021d]). Não se mencionam quais ações a instituição utiliza em relação à

preservação digital; além dos referidos documentos relacionados, ressalta a importância da *Carta sobre a Preservação do Patrimônio Digital* da Unesco. Schisler (2020, p. 800) concorda com Owens (2018), que “avalia a seriedade dos programas de preservação digital de uma instituição pela verba alocada anualmente para a tarefa”.

Na categoria “Observações”, destaca-se que tanto o Arquivo.pt quanto a BNE possuem perfil no Github¹¹⁰, configurando-se como “uma plataforma de hospedagem de código-fonte e arquivos com controle de versão usando o Git. Ele permite que programadores ou qualquer usuário cadastrado na plataforma contribuam em projetos privados e/ou Open Source de qualquer lugar do mundo”. Desde 2018, pertence à Microsoft, com intuito de impulsionar seu compromisso com o código aberto (GITHUB, 2022)¹¹¹. A política da plataforma envolve a promoção da inovação, colaboração e igualdade de oportunidades para que os desenvolvedores possam realizar o melhor trabalho (GITHUB INC, [c2022])¹¹². Tal iniciativa do Github de colaboração entre desenvolvedores e instituições corrobora o axioma dezesesseis de Owens (2018), que, para fazer a preservação digital, requer pensar como um futurista, ou seja, precisa-se estar atento às tendências de desenvolvimento das tecnologias digitais, considerar a evolução tecnológica futura, com diversidade de formatos e padrões, sempre preservar com olhar no futuro, pois a preservação é uma ação continuada. Vale ressaltar que todos os membros fundadores do IIPC possuem perfil nessa plataforma.

Para alcançar o último objetivo específico D, identificar os problemas e as barreiras à preservação das páginas *web*, foram realizadas reflexões sobre as referidas instituições de memória com base nos textos da pesquisa bibliográfica e documental. Destacam-se dez categorias de problemas e barreiras encontradas também nas respostas das instituições pesquisadas (vide ANEXO A), que são:

- 1- Formato – os arquivos WARC gerados devem ser incluídos no fluxo de preservação, tal como acontece com outros tipos de formatos digitais. O WARC é um tipo de formato que não é abrangido pelos fluxos de preservação habituais e necessita de consideração e alocação específica;

¹¹⁰ O GitHub é uma rede social de desenvolvedores. A primeira parte do nome, “Git”, é por causa da utilização do sistema de controle de versão e a segunda parte, “Hub”, tem a ver com a conexão entre profissionais de programação de qualquer lugar do mundo (SOUZA, 2020).

¹¹¹ GITHUB. In: WIKIPEDIA: a enciclopédia livre. [São Francisco, CA: Wikimedia Foundation, 2022]. Disponível em: <https://pt.wikipedia.org/wiki/GitHub>. Acesso em: 15 fev. 2022.

¹¹² GITHUB INC. *GitHub Policy: fighting for developers*. [c2022]. Disponível em: <https://github.com/about/developer-policy>. Acesso em: 15 fev. 2022.

- 2- Seleção – este trabalho, muitas vezes, não é realizado de forma exaustiva, e os níveis de abrangência do *website* também não o são. É preciso determinar o escopo do programa de arquivamento e preservação da *web*;
- 3- Captura – a duração em média de uma página *web* é menos de 100 dias, considerados como recursos efêmeros, sendo necessário pensar em mecanismos técnicos e recursos humanos para que o repositório preserve e ofereça um instantâneo da *web* em um determinado momento. Esse problema abrange as mídias sociais que não são capturadas corretamente, o que atualmente é um grande problema, uma vez que grande parte da informação publicada na *web* provém de mídias sociais;
- 4- Metadados – problemas em determinar qual o conjunto de metadados é mais apropriado para esse tipo de conteúdo;
- 5- Visualização – problemas em visualizar os conteúdos como foram produzidos, já que muitos recursos dinâmicos atrapalham a captura dos *websites* e, conseqüentemente, a acessibilidade e a usabilidade são prejudicadas;
- 6- Investimento financeiro – em geral, os principais problemas devem-se à falta de recursos financeiros, humanos e técnicos (desenvolvimento, design) e à complexidade da preservação, intrínseca a este tipo de material.
- 7- Especialistas – problemas em contratar e manter equipes capacitadas em arquivamento e preservação da *web*;
- 8- Disseminação – custos em disseminar, por meio de infraestrutura própria ou privada, por exemplo, *Internet Archive*;
- 9- Acesso – dar acesso em consonância com os mandatos legais, de depósito e direitos autorais desses conteúdos *web* são também barreiras à preservação digital;
- 10- Mudança de cultura organizacional – as instituições de memória estão em constante adaptação nesse cenário digital, mas ainda continuam cumprindo o depósito legal semelhante aos conteúdos em suportes analógicos.

Nunes (2021) estudou os aspectos éticos e legais no acesso e uso da informação no âmbito do arquivamento da *web*, concluindo que o acesso às informações é complexo e que uma política informacional deve ser orientada para que toda a sociedade seja beneficiada, sendo primordial o trabalho colaborativo e interdisciplinar entre as diferentes áreas. Tal aspecto legal foi ressaltado por Gracio (2012) como segundo ponto importante para a preservação digital.

Algumas dessas barreiras foram salientadas no livro *The Past Web: Exploring Web Archives*, em particular no capítulo “VI – Um olhar sobre o futuro”, que resume o livro e dá uma perspectiva, destacando as lições aprendidas, com objetivo de refletir sobre as mudanças que ocorreram na *web* e como isso tem afetado a preservação do seu conteúdo (MASANÈS; MAJOR; GOMES, 2021, p. 285). De acordo com os autores, são problemas para a preservação e arquivamento da *web* a variedade de formatos e a dinamicidade das páginas *web*; e os altos custos operacionais e de recursos humanos. Dessa forma, frisam que há necessidade de capacitação e formação de profissionais especializados em preservação e arquivamento da *web*, como uma profissão do futuro. Para Ramos e Cianconi (2019), o conhecimento produzido por uma instituição é, frequentemente, melhorado por seus colaboradores, que são os atores que vivenciam na prática os desafios organizacionais da preservação digital. A preservação é considerada um componente estratégico e transversal a fim de garantir a missão institucional. Desse modo, destacam três tipos de barreiras ao efetivo compartilhamento de conhecimento organizacional, que são semelhantes aos problemas e barreiras encontrados para a preservação digital, seja de maneira “individual, com a falta de habilidade de comunicação ou diferença cultural/técnica entre os profissionais; organizacional, com a falta de infraestrutura e recursos; e a tecnológica, com a dificuldade de utilização de ferramentas e sistemas ou até mesmo do acesso” (RAMOS JUNIOR; CIANCONI, 2019, p. 83-85). Por isso, é primordial a cooperação entre instituições congêneres para a realização de tal tarefa.

Masanès, Major e Gomes (2021) concluíram que “não existe nenhum modelo pré-existente de instituições patrimoniais (arquivos, museus, bibliotecas) que podem acomodar diretamente iniciativas de arquivamento da *web*”, afirmando que deve ser um novo tipo de organização do patrimônio cultural, com identidade, responsabilidades e métodos próprios (MASANÈS; MAJOR; GOMES, 2021, p. 289). Essa posição é diferente da de Owens (2018), que, em seu segundo axioma, afirma que são essas instituições (bibliotecas, arquivos, museus, famílias, organizações religiosas, governos etc.) que tornam possível a preservação, ou seja, são os principais facilitadores da preservação em longo prazo com a missão de garantir o acesso as informações ao longo dos anos.

Apesar de, teoricamente, Masanès, Major e Gomes (2021) acreditarem que é preciso criar instituições específicas para o trabalho de arquivamento da *web*, na prática, ao longo de mais de 25 anos, a tarefa vem sendo realizada primordialmente por instituições de memória e patrimônio desde a publicação da Unesco (c2004), a *Carta sobre a Preservação do Patrimônio Digital* – confirmado pelo segundo axioma de Owens (2018) e Schisler (2020)

que aponta que é preciso que as instituições que atravessam gerações sejam fortalecidas e tenham seu papel reconhecido na missão de preservar. Além disso, o digital remete a uma aparente impressão de solidez que faz parecer que é possível preservar tudo. Como bem lembrado por Owens (2018, p. 8) e Schisler (2020, p. 801), antes de tudo, a preservação digital é um desafio técnico, embora a preservação da *web* possua desafios no âmbito social e ético, empregando graus de esforços variados que serão colocados à prova no futuro por diferentes tipos objetos digitais complexos. Realizar o arquivamento dos fragmentos desse universo digital, especificamente da *web*, possibilitará pesquisar, acessar, usar e reutilizar conteúdos desse universo da produção digital, pois preservar páginas *web* é uma forma de capturar um momento desse imenso universo digital que a *web* proporciona.

9 CONSIDERAÇÕES FINAIS

O presente trabalho teve como objetivo geral estudar a preservação do patrimônio digital por instituições de guarda da memória das páginas *web*, que fazem parte do IIPC, com base na *Carta sobre a Preservação do Patrimônio Digital* da Unesco, com o propósito de responder à seguinte questão: onde e como estão sendo preservadas as páginas *web*? E, também, alcançar quatro determinados objetivos específicos.

O percurso da pesquisa teve como primeiro passo introduzir a temática do ambiente digital, apresentando o surgimento da internet e da *web*, bem como a forma em que ambas foram inseridas ao longo dos anos no cotidiano da sociedade, tornando-se instrumentos importantes de comunicação. Foi abordada também a importância da *Carta sobre a Preservação do Patrimônio Digital* e outros documentos produzidos pela Unesco no cenário da preservação, com a promoção de políticas institucionais no âmbito da preservação – a Carta da Unesco incluiu as páginas *web* como parte do patrimônio digital, tornando-se, então, elemento-chave para a presente pesquisa.

Foi preciso entender as questões ligadas ao arquivamento da *web*, apresentando os conceitos, os processos, os desafios e os atores no contexto nacional e internacional, com o intuito de salientar a importância da preservação das páginas *web* para as futuras gerações e, também, conectar o assunto com a preservação do patrimônio. Partiu-se do conceito de preservação ainda no ambiente analógico, mostrando as semelhanças com a preservação digital, pois ambos os conceitos são parte integrante das missões das instituições de memória e patrimônio.

A partir desse caminho, foi possível chegar às etapas para arquivamento e preservação da *web*, apresentando o modelo de referência OAIS, os princípios FAIR e TRUST, correlacionando-os com os padrões de metadados e arquivamento, o que permitiu adentrar na parte empírica, isto é, nas análises e discussão dos resultados das instituições consorciadas ao IIPC.

Foi necessário, inicialmente, compreender como tal consórcio foi formado, identificando-se e caracterizando-se as instituições membros fundadoras do IIPC, a saber: *Bibliothèque nationale de France*, *Internet Archive*, *The British Library*, *Royal Danish Library*, *National Library of Finland*, *National Library of Sweden*, *National and University Library of Iceland*, *Library and Archives Canada*, *Library of Congress*, *National Library of Norway* e *National Library of Australia*. Foi feito o mesmo procedimento com as instituições ibero-americanas do IIPC, delimitando o campo pesquisado, observando suas políticas, ações

e critérios de preservação, em Portugal, Espanha e Chile, nas instituições Arquivo.pt (*Portuguese Web Archive*), *National Library of Spain*, *Biblioteca da Catalunya* (PADICAT) e *Chilean National Library*.

Examinando as instituições do IIPC, vimos que a preservação das páginas *web* foi praticada ao longo dos últimos anos por diversas instituições, destacando-se as bibliotecas nacionais que dividem o trabalho com outros tipos de instituições, como bibliotecas regionais, bibliotecas universitárias ou arquivos nacionais. Como exceção, a tarefa foi também realizada por organizações sem fins lucrativos ou ligadas ao ensino e à pesquisa. A predominância de bibliotecas nacionais realizando o trabalho de preservação de páginas *web* explica a utilização do marco de depósito legal para o recolhimento desses tipos de conteúdos. No entanto, são repetidas as regras aplicadas nos documentos analógicos, como acesso e uso restrito ao local físico institucional.

Há também uma concentração de arquivamento e preservação da *web* em instituições localizadas em países do Hemisfério Norte, principalmente no continente europeu. Acredita-se que o fomento à pesquisa nesse âmbito da preservação digital oferecido pela União Europeia seja essencial para concentração de esforço nessa região, principalmente estimulando tal trabalho de forma contínua, com promoção de eventos, comunidade de pesquisadores e financiamento de projetos. Constatou-se que a comunidade europeia desenvolve várias ações de preservação de patrimônio cultural, com recomendações e apoio financeiro, inclusive projetos e iniciativas de *web archiving*.

Descobriu-se, durante as análises dos dados levantados e na construção dos quadros intitulados “Caracterização institucional” e “Enfoque na governança institucional”, que, apesar de muitas possuírem missões institucionais semelhantes, a execução da preservação digital é distinta, conforme a realidade de cada país, dificultando o agrupamento ou a realização de estatísticas.

Ao acessar os *websites* institucionais e analisar as respostas recebidas por *e-mails* das instituições, pode-se perceber que poucas possuem área dedicada à preservação digital com equipe de especialistas, e que apenas algumas contam com documentos institucionais sobre a temática, política de preservação digital, planejamento estratégico e legislação. Grande parte das instituições membros fundadoras possuem área dedicada à preservação digital e documentos institucionais sobre a temática, pois contam com uma experiência de anos no ofício de preservação digital.

Em relação às instituições ibero-americanas, a realidade da BNE e BC-PADICAT é parecida por serem regidas pela mesma legislação (Real Decreto), embora exista a diferença

na questão do acesso – a primeira concede apenas acesso no local físico, como a BNC, no Chile, e a segunda oferece o acesso *on-line*, como o Arquivo.pt.

Observa-se que a BNC, apesar de ser ligada a um consórcio internacional como IIPC, ainda está solidificando seu trabalho e que, atualmente, mostra interesse no assunto, embora ainda não possua um documento institucional de política de preservação digital e estratégias para arquivamento de páginas *web*. A instituição também não menciona de forma objetiva em seu *website* quais padrões utiliza, apenas ressalta a carta da Unesco como ponto de partida para a preservação digital e elenca alguns documentos que considera importantes. Isso se mostra diferente das demais, que empregam esforços para sanar as dúvidas relacionadas com o tema em seu *website* institucional, promovendo também a preservação da memória de suas atividades.

No entanto, os desafios para as quatro instituições são semelhantes, conforme mencionados nas respostas ao *e-mail* (Anexo A). Os problemas e as barreiras à preservação das páginas *web* estão ligados a: formato do arquivo WARC, captura e seleção, pois seu conteúdo é efêmero e dinâmico; problemas em visualizar os conteúdos na forma como foram produzidos, consequentemente prejudicando a acessibilidade e usabilidade; investimento financeiro, pois os principais problemas devem-se à falta de recursos financeiros, humanos e técnicos (desenvolvimento, *design*) e à complexidade da preservação digital intrínseca a esse tipo de material; o quesito especialistas, com a dificuldade de encontrar e manter equipes capacitadas em arquivamento e preservação da *web*; disseminação e acesso, os custos de disseminar, dar acesso e as questões ligadas à legislação desses conteúdos *web* são também barreiras à preservação digital; e a mudança de cultura organizacional, pois as instituições de memória estão em constante adequação ao cenário digital, mas ainda continuam cumprindo o depósito legal de modo semelhante ao que fazem em relação aos conteúdos em suportes analógicos.

Percebeu-se, dentro desse universo do IIPC, a importância de estabelecer políticas e documentar de forma institucional a adoção de modelos (OAIS), normas (ISO 28500), padrões (Dublin Core e METS), reconhecidos internacionalmente para que essas ações sustentem o trabalho de preservação digital. Outrossim, por meio das informações levantadas para cumprir os objetivos específicos, ter uma perspectiva da maneira de como cada instituição realiza a tarefa na prática.

Notou-se que apenas a instituição portuguesa oferece um prêmio anual para os pesquisadores que fazem uso dos conteúdos preservados. Tal iniciativa é uma forma de fomentar e ampliar o acesso, uso e reuso mencionados no ciclo de arquivamento da *web*,

consequentemente, promovendo os conteúdos no contexto das humanidades digitais e o sentimento preservacionista do patrimônio digital na sociedade.

Considerando que apesar de tais instituições pesquisadas estarem no IIPC, em que é proporcionado o acesso a uma comunidade colaborativa de especialistas, informações, projetos de financiamento e serviços ligados ao arquivamento e preservação da *web*, elas passam por dificuldades semelhantes em relação a investimentos em recursos humanos especializados, de infraestrutura (*software* e *hardware*), questões legais e, principalmente, em sustentar a preservação em longo prazo desses objetos complexos que são as páginas *web*, isto é, do patrimônio digital.

A pesquisa permitiu atingir os objetivos propostos, e percebeu-se que a preservação digital é um desafio tanto para os profissionais quanto para as instituições, e que nem sempre a teoria anda no mesmo ritmo que a prática. Ainda que sejam empregados todos os instrumentos necessários e reconhecidos pela comunidade internacional, há o fato de terem que garantir a mesma missão de preservar o patrimônio digital em longo prazo, a realidade e os objetivos de cada país são distintos.

Conclui-se que as instituições de memória, como bibliotecas e arquivos, devem refletir e reagir, como a Unesco bem colocou, liderando esse movimento de preservação do patrimônio digital; e embora o desafio seja imenso, deve-se considerar esses objetos complexos como parte do patrimônio nacional. É preciso incluir o Brasil nesse cenário de preservação digital de páginas *web*, seja por meio de programas que visem ao acesso aberto, ponderando que atualmente a maioria das iniciativas nacionais de preservação da *web* continua vinculada aos modelos de acesso restrito ao local físico, pelo mandato do depósito legal, que diminuem a utilidade e a visibilidade desses conteúdos – com exceção da Biblioteca Nacional da Austrália que, ao recolher, elenca quais níveis de acesso os conteúdos podem ter na tentativa de permanecer acessível à informação que nasceu aberta. Além disso, a exploração dos conteúdos arquivados requer uma compreensão da dinâmica da *web* pelos pesquisadores, usuários, profissionais, ou seja, pela sociedade que fará uso de tais conteúdos, capacitando-os também para que estes sejam úteis a todos. Ademais, considerando que a *web* é, atualmente, o principal meio de comunicação na sociedade, se as organizações de memória e patrimônio continuarem a subestimar a importância de preservá-la em longo prazo como fonte de informação, pesquisa e prova, melhor dizendo, como parte do patrimônio digital preconizado pela carta da Unesco e não criarem mecanismos de acesso, para uso e reuso desses objetos digitais complexos, as instituições de guarda da memória correm o risco de se tornarem irrelevantes no futuro e ainda contribuirão para o apagamento da memória nacional

brasileira. Além disso, é fundamental incluir nos currículos das universidades a temática da preservação digital e arquivamento de páginas *web*, para que os profissionais em formação sejam capacitados para trabalhar, gerir, organizar e dar acesso a esses conteúdos complexos para a sociedade.

Desse modo, o estudo e a análise da preservação do patrimônio digital realizada por instituições de guarda da memória das páginas *web*, pertencentes ao IIPC, baseando-se nos preceitos da *Carta sobre a Preservação do Patrimônio Digital* da Unesco, que buscou conhecer onde e como estão sendo preservadas as páginas *web*, proporcionaram-me amadurecer como pesquisadora do tema e, também, como profissional. Durante a execução desta pesquisa, surgiu o convite para compor a “*Câmara Técnica Consultiva do CONARQ – para a preservação de websites e mídias sociais*”, como um dos seus membros especialistas, representando a Biblioteca Nacional no cenário arquivístico brasileiro. Decerto que foi um rico aprendizado realizar a pesquisa, ampliando a compreensão sobre o arquivamento e preservação digital da *web*, refletindo sobre as políticas e as práticas das instituições membros fundadoras e ibero-americanas do IIPC. Portanto, está sendo útil e contribuindo para a atuação profissional desta bibliotecária e arquivista na Biblioteca Nacional do Brasil.

Cabe relatar ainda que o estudo promoveu grande aprendizado e motivações para pensar sobre o lugar que o arquivamento e a preservação da *web* ocuparão nas instituições de guarda e preservação da memória no Brasil no futuro. Percebeu-se que tais acervos são extremamente importantes e devem ser encarados como fazendo também parte do patrimônio nacional. Certamente, é impossível preservar tudo que é produzido na *web* em nossa sociedade, mas soluções devem ser buscadas tanto em termos de investimento, recursos e políticas públicas que estimulem a preservação desses conteúdos *web* em longo prazo como de políticas institucionais e capacitação profissional para preservação desses acervos que estão ameaçados por causa de todo tipo de riscos, inclusive de guerras.

REFERÊNCIAS

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS – ABNT. *NBR ISO 14721*: sistemas espaciais de transferência de dados e de informação: Sistema Aberto de Arquivamento de Informação (SAAI): modelo de referência. Rio de Janeiro: ABNT, 2021.

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS – ABNT. *NBR ISO 15472*: sistemas espaciais de dados e de informação: Modelo de referência para um sistema aberto de arquivamento de informação (SAAI): modelo de referência. Rio de Janeiro: ABNT, 2007.

ALENCAR BRAYNER, A. Programa de arquivo de páginas web no Reino Unido: uma breve história de oportunidades e desafios. *Revista Digital de Biblioteconomia e Ciência da Informação*, Campinas, v. 14, n. 2, p. 318-333, 2016. DOI: 10.20396/rdbci.v14i2.8645982.

AMERICAN LIBRARY ASSOCIATION – ALA. *Definitions of digital preservation*. Chicago, IL: ALA, 2008. Disponível em: <http://www.ala.org/alcts/resources/preserv/defdigpres0408>. Acesso em: 31 maio 2021.

ARCHIVE IT. *About Archive-It*. [2014]. Disponível em: <https://archive-it.org/blog/learn-more/>. Acesso em: 21 set. 2021.

ARCHIVE IT. *Canadian government information PLN web archive*. 2013. Disponível em: <https://archive-it.org/organizations/700>. Acesso em: 21 nov. 2021.

ARCHIVE IT. *K-12 web archiving*. [2014?]. Disponível em: <https://www.archive-it.org/k12/>. Acesso em: 21 nov. 2021.

ARQUIVO NACIONAL (Brasil). *Dicionário brasileiro de terminologia arquivística*. Rio de Janeiro: Arquivo Nacional, 2005. 232 p.

ARQUIVO.PT. *Informações gerais*. 13 abr. 2021a. Disponível em: <https://sobre.arquivo.pt/pt/ajuda/o-que-e-o-arquivo-pt>. Acesso em: 10 dez. 2021.

ARQUIVO.PT. *Prêmios Arquivo.pt*. 23 set. 2021b. Disponível em: <https://sobre.arquivo.pt/pt/colabore/premios-arquivo-pt/>. Acesso em: 10 dez. 2021.

ARQUIVO.PT. *Objetivos do Arquivo.pt*. 13 abr. 2021c. Disponível em: <https://sobre.arquivo.pt/pt/acerca/objetivos/>. Acesso em: 10 dez. 2021.

ARQUIVO.PT. *Recolha de conteúdos*. 27 jul. 2021d. Disponível em: <https://sobre.arquivo.pt/pt/ajuda/recolha-e-arquivo-de-conteudos/>. Acesso em: 10 dez. 2021.

ARQUIVO.PT. *Exposições online Arquivo.pt*. 16 set. 2021e. Disponível em: <https://sobre.arquivo.pt/pt/exemplos-2/colecoes/>. Acesso em: 10 dez. 2021.

ARQUIVO.PT. *Memorial do Arquivo.pt*. 6 dez. 2021f. Disponível em: <https://sobre.arquivo.pt/pt/colabore/memorial/>. Acesso em: 10 dez. 2021.

ARQUIVO.PT. *Formação acerca de preservação da web*. 12 out. 2021g. Disponível em: <https://sobre.arquivo.pt/pt/ajuda/formacao/#modulos>. Acesso em: 10 dez. 2021.

ARQUIVO.PT. *SavePageNow – grava páginas no Arquivo.pt*. 2021h. Disponível em: <https://arquivo.pt/services/savepagenow?l=pt>. Acesso em: 10 dez. 2021.

ARQUIVO.PT. *Termos e condições*. 19 set. 2017a. Disponível em: <https://sobre.arquivo.pt/pt/acerca/termos-e-condicoes/>. Acesso em: 10 dez. 2021

ARQUIVO.PT. *Arquivo.pt: novas formas de viajar ao passado!* 26 set. 2017b. Disponível em: <https://www.youtube.com/watch?v=YVqFey7hVJc&t=12s>. Acesso em: 10 dez. 2021.

ASCHENBRENNER, A. Kulturarw3: The Swedish Archive. *Tu Wien Informatics*, Viena, Áustria, [200?]. Disponível em: <https://www.ifs.tuwien.ac.at/~aola/publications/thesis-ando/Kulturarw3.html>. Acesso em: 18 out. 2021.

AUBRY, S. *The WARC file format celebrates its 10th anniversary*. 29 maio 2019. Disponível em: <https://netpreserveblog.wordpress.com/2019/05/29/warc-10th-anniversary/>. Acesso em: 03 set. 2021.

BANOS, V. *et al.* CLEAR: a credible method to evaluate website archivability. In: INTERNATIONAL CONFERENCE ON PRESERVATION OF DIGITAL OBJECTS (iPRES 2013), 10., 2013, Lisbon. *Proceedings* [...]. Lisbon: Instituto Superior Técnico, 2013. Disponível em: http://purl.pt/24107/1/iPres2013_PDF/CLEAR%20a%20credible%20method%20to%20evaluate%20website%20archivability.pdf. Acesso em: 31 maio 2021.

BARDIN, L. *Análise de conteúdo*. Lisboa: Edições 70, c1977.

BESSER, H. Digital Longevity. In: SITTS, M. (ed.). *Handbook for digital projects: a management tool for preservation and access*. Andover, MA: Northeast Document Conservation Center, 2000. Disponível em: <http://besser.tsoa.nyu.edu/howard/Papers/sfs-longevity.html>. Acesso em: 10 fev. 2021.

BELLOTTTO, H. L. *Arquivo: estudos e reflexões*. Belo Horizonte: Ed. UFMG, 2014. 477 p.

BELLO, R. Biblioteca nacional digital: archivo de la web chilena. In: SIMPÓSIO POLÍTICAS E ESTRATÉGIAS DE PRESERVAÇÃO DE CONTEÚDO NA WEB, TOI 2021 ONLINE, 2., 15 dez. 2021. 1 vídeo (160 min). Publicado pelo canal TOI 2021 ONLINE. Disponível em: <https://www.youtube.com/watch?v=3KhAu7OczSA>. Acesso em: 23 dez. 2021.

BIBLIOTHÈQUE ET ARCHIVES NATIONALES DU QUÉBEC. *Dépôt legal*. [2021a]. Disponível em: https://www.banq.qc.ca/services/bibliotheque_nationale/depot_legal/#avantages. Acesso em: 30 nov. 2021.

BIBLIOTHÈQUE ET ARCHIVES NATIONALES DU QUÉBEC. *Collecte de sites web*. [2021b]. Disponível em: https://www.banq.qc.ca/services/bibliotheque_nationale/depot_legal/collecte_sites/. Acesso em: 30 nov. 2021.

BIBLIOTHÈQUE ET ARCHIVES NATIONALES DU QUÉBEC. *Préservation des collections patrimoniales*. [2021c]. Disponível em: https://www.banq.qc.ca/collections/collections_patrimoniales/sauvegarde/. Acesso em: 30 nov. 2021.

BIBLIOTHÈQUE ET ARCHIVES NATIONALES DU QUÉBEC. *Plan Stratégique 2019-2022*. [2021d]. Disponível em: <https://numerique.banq.qc.ca/patrimoine/details/52327/65853?docref=6GutLBCaFtvEynAN44-NKw>. Acesso em: 30 nov. 2021.

BIBLIOTHÈQUE ET ARCHIVES NATIONALES DU QUÉBEC. *Collections*. [2021e]. Disponível em: https://www.banq.qc.ca/collections/collections_patrimoniales/archives_web/index.html?language_id=1. Acesso em: 14 dez. 2021.

BIBLIOTECA DA CATALUNYA. *Dipòsit legal*. [c2021a]. Disponível em: <https://www.bnc.cat/Professionals/Diposit-legal#enqueconsisteix>. Acesso em: 14 dez. 2021.

BIBLIOTECA DA CATALUNYA. *Pla estratègic 2017-2021*. [c2021b]. Disponível em: <https://www.bnc.cat/Coneix-nos/Qualitat-i-estrategia/Pla-estrategic>. Acesso em: 14 dez. 2021.

BIBLIOTECA DA CATALUNYA. *Preservação digital*. [c2021c]. Disponível em: <https://www.bnc.cat/Professionals/Preservacio-digital2>. Acesso em: 14 dez. 2021.

BIBLIOTECA DA CATALUNYA. *Memòria del plantejament del projecte PADICAT (Patrimoni Digital de Catalunya)*. 2005. Disponível em: https://recercat.cat/bitstream/handle/2072/1757/IN_PADICAT_2005.pdf?sequence=1. Acesso em: 31 jan. 2022.

BIBLIOTECA NACIONAL DE ESPANHA. *Depósito legal de las publicaciones en línea*. [c2021a]. Disponível em: <http://www.bne.es/es/Colecciones/Adquisiciones/DepositoLegal/DLe/index.html>. Acesso em: 14 dez. 2021.

BIBLIOTECA NACIONAL DE CHILE. *Biblioteca Nacional Digital*: archivo de la web chilena. [2021a]. Disponível em: <http://archivoweb.bibliotecanacionaldigital.cl/wayback/>. Acesso em: 23 dez. 2021.

BIBLIOTECA NACIONAL DE CHILE. *Biblioteca nacional digital*: depósito eletrônico. [2021b]. Disponível em: <http://www.bibliotecanacionaldigital.gob.cl/bnd/612/w3-propertyvalue-174014.html>. Acesso em: 23 dez. 2021.

BIBLIOTECA NACIONAL DE CHILE. *Servicio Nacional del patrimonio cultural*: depósito legal. [2021c]. Disponível em: https://www.bibliotecanacional.gob.cl/615/w3-article-7058.html?_noredirect=1. Acesso em: 23 dez. 2021.

BIBLIOTECA NACIONAL DE CHILE. *Preservación digital*. [2021d]. Disponível em: <http://www.bibliotecanacionaldigital.gob.cl/bnd/612/w3-propertyvalue-174087.html>. Acesso em: 23 dez. 2021.

BIBLIOTECA NACIONAL DE CHILE. *Contacto*. [2021e]. Disponível em: <http://www.bibliotecanacionaldigital.gob.cl/bnd/612/w3-propertyvalue-783237.html>. Acesso em: 23 dez. 2021.

BIBLIOTECA NACIONAL DE ESPANHA. *Archivo de la web española*. [c2021b]. Disponível em: <http://www.bne.es/es/Colecciones/ArchivoWeb/>. Acesso em 14 dez. 2021.

BIBLIOTECA NACIONAL DE ESPANHA. *Información técnica*. [c2021c]. Disponível em: <http://www.bne.es/es/Colecciones/ArchivoWeb/InfoTecnica/index.html>. Acesso em: 14 dez. 2021.

BIBLIOTECA NACIONAL DE ESPANHA. *Entidades colaboradoras*. [c2021d]. Disponível em: <http://www.bne.es/es/Colecciones/Adquisiciones/DepósitoLegal/DLe/entidades-colaboradoras/>. Acesso em: 14 dez. 2021.

BIBLIOTECA NACIONAL DE ESPANHA. *Comunidades autónomas*. [c2021e]. Disponível em: <http://www.bne.es/es/Colecciones/ArchivoWeb/Subcolecciones/ComunidadesAutonomas.html>. Acesso em: 14 dez. 2021.

BIBLIOTECA NACIONAL DE ESPANHA. *Plan estratégico 2021– 2025*. [c2021f]. Disponível em: <http://www.bne.es/export/sites/BNWEB1/webdocs/LaBNE/Transparencia/Institucional/2021-plan-estrategico-bne.pdf>. Acesso em: 09 fev. 2022.

BIBLIOTECA NACIONAL DE ESPANHA. *Busca*. [c2021g]. Disponível em: <http://www.dle.es/openwayback/wayback/>. Acesso em: 09 fev. 2022.

BIBLIOTECA NACIONAL DE ESPANHA. *Depósito legal de las publicaciones en línea: preguntas frecuentes*. 09 jul. 2015. Disponível em: http://www.bne.es/export/sites/BNWEB1/webdocs/Servicios/PreguntasFrecuentes/DLe_PreguntasFrecuentes.pdf. Acesso em: 14 dez. 2021.

BIBLIOTECA NACIONAL DE ESPANHA. *Preguntas frecuentes sobre el archivo de la web española*. 6 jul. 2017. Disponível em: <http://www.bne.es/export/sites/BNWEB1/webdocs/Colecciones/ArchivoWeb/PreguntasFrecuentesAWE.pdf>. Acesso em: 14 dez. 2021.

BIBLIOTHÈQUE NATIONALE DE FRANCE. *Expertise sur l'archivage du web*. 2021a. Disponível em: <https://www.bnf.fr/fr/expertise-sur-larchivage-du-web>. Acesso em: 21 set. 2021.

BIBLIOTHÈQUE NATIONALE DE FRANCE. *Code du patrimoine: articles L131-1 à L133-1 et R131-1 à R133-1*. 2021b. Disponível em: <https://www.bnf.fr/fr/code-du-patrimoine-articles-l131-1-l133-1-et-r131-1-r133-1>. Acesso em 16 mar. 2021.

BIBLIOTHÈQUE NATIONALE DE FRANCE. *Archives de l'internet*. 2021c. Disponível em: <https://www.bnf.fr/fr/archives-de-linternet>. Acesso em: 21 set. 2021.

BIBLIOTHÈQUE NATIONALE DE FRANCE. *Parcours guidés dans les archives de l'internet*. 2021d. Disponível em: https://www.bnf.fr/fr/centre-d-aide/parcours-guides-dans-les-archives-de-linternet#step01_. Acesso em: 21 set. 2021.

BIBLIOTHÈQUE NATIONALE DE FRANCE. *SPAR – Système de Préservation et d'Archivage Réparti*. 2021e. Disponível em: <https://www.bnf.fr/fr/spar-systeme-de-preservation-et-darchivage-reparti>. Acesso em: 21 set. 2021.

BIBLIOTHÈQUE NATIONALE DE FRANCE. *Infrastructure matérielle de SPAR (Système de Préservation et d'Archivage Réparti)*. 2021f. Disponível em: <https://www.bnf.fr/fr/infrast-structure-materielle-de-spar-systeme-de-preservation-et-darchivage-reparti>.

BIBLIOTHÈQUE NATIONALE DE FRANCE. *Dépôt légal des sites web: mode d'emploi*. 2021g. Disponível em: <https://www.bnf.fr/fr/centre-d-aide/depot-legal-des-sites-web-mode-demploi#step01>. Acesso em: 21 set. 2021.

BIBLIOTHÈQUE NATIONALE DE FRANCE. *Consultation des archives de l'internet de la BnF*. 2021h. Disponível em: <https://www.bnf.fr/fr/consultation-des-archives-de-linternet-de-la-bnf>. Acesso em: 21 set. 2021.

BIBLIOTHÈQUE NATIONALE DE FRANCE. *Le dépôt légal*. 2021i. Disponível em: <https://www.bnf.fr/fr/quest-ce-que-le-depot-legal> . Acesso em: 21 set. 2021.

BLUMENTHAL, K. The stack: An introduction to the WARC file. 01 abr. 2021. *Archive-It Blog*, 2021a. Disponível em: <https://archive-it.org/blog/post/the-stack-war-c-file/>. Acesso em: 03 set. 2021.

BLUMENTHAL, K. Archive-It APIs and integrations. *Archive It Help Center*, [2021b]. Disponível em: <https://support.archive-it.org/hc/en-us/articles/360001231286>. Acesso em: 21 set. 2021.

BOULTON, J. The Nexus browser. *Blog digital Archaeology*, Mar. 2014. Disponível em: <https://digital-archaeology.org/the-nexus-browser/>. Acesso em: 16 abr. 2021.

BRANCO, I. Capacity management study of Arquivo.pt. 7 July 2020. *Arquivo.pt*, 2020. Disponível em: https://sobre.arquivo.pt/wp-content/uploads/arquivo_pt_2020_capacity_management.pdf. Acesso em: 14 dez. 2021.

BRASIL. Decreto nº 8.539, de 8 de outubro de 2015. Dispõe sobre o uso do meio eletrônico para a realização do processo administrativo no âmbito dos órgãos e das entidades da administração pública federal direta, autárquica e fundacional. *Diário Oficial União*: seção 1, Brasília, DF, 9 out. 2015. Disponível em: http://www.planalto.gov.br/ccivil_03/_Ato2015-2018/2015/Decreto/D8539.htm. Acesso em: 29 nov. 2020.

BRASIL. Câmara dos Deputados. *PL 2.431/2015*: projeto de lei. 2015b. Disponível em: <https://www.camara.leg.br/proposicoesWeb/fichadetramitacao?idProposicao=1594241>. Acesso em: 06 set. 2021.

BRAGG, M.; HANNA, K. *The web archiving life cycle model*. [S. l.: s. n.], 2013. Disponível em: http://ait.blog.archive.org/files/2014/04/archiveit_life_cycle_model.pdf. Acesso em: 16 mar. 2021.

BROWN, A. *Archiving websites: a practical guide for information management professionals*. [S. l.]: Facet, 2006. DOI: 10.29085/9781856049009.

BRÜGGER, N. *Archiving websites: general considerations and strategies*. Århus: The Centre for Internet Research, 2005. Disponível em: https://cfi.au.dk/fileadmin/www.cfi.au.dk/publikationer/archiving_underside/archiving.pdf. Acesso em: 29 nov. 2020.

BRÜGGER, N. Web archiving: between past, present, and future. In: CONSALVO, M.; ESS, C. (ed.). *The handbook of internet studies*. Oxford: Blackwell Publishing, 2011. p. 24-42.

BRÜGGER, N. *The archived web: doing history in the digital age*. Cambridge, MA: The MIT Press, 2018. 200 p.

BUEREN, J. L. Memória e preservação digital. In: FUNDAÇÃO BIBLIOTECA NACIONAL (Brasil). 19 out. 2021. 66 slides. Disponível em: <https://nuvem.bn.gov.br/index.php/s/FzfdooTdRMkErqO>. Acesso em: 14 dez. 2021.

BYRNE, H. Launching the UK Web archive 2020 annual domain crawl. *UK Web archive blog*, Reino Unido, 2020. Disponível em: <https://blogs.bl.uk/webarchive/2020/09/launching-the-uk-web-archive-2020-annusl-domain-crawl.html>. Acesso em: 20 set. 2021.

CAMPELLO, B. *Introdução ao controle bibliográfico*. 2. ed. Brasília, DF: Briquet de Lemos Livros, 2006.

CASSARES, N. C. *Como fazer conservação preventiva em arquivos e bibliotecas*. São Paulo: Arquivo do Estado e Imprensa Oficial, 2000. Disponível em: http://www.arquivoestado.sp.gov.br/site/assets/publicacao/anexo/como_fazer_conservacao_preventiva_em_arquivos_e_bibliotecas.pdf. Acesso em: 09 mar. 2021.

CASEY, C. The cyberarchive: a look at the storage and preservation of web sites. *College & Research Libraries*, Chicago, v. 59, n. 4, p. 304-310, July 1998. Disponível em: <https://crl.acrl.org/index.php/crl/article/view/15217>. Acesso em: 29 nov. 2020.

CASTELLS, M. *A galáxia da internet: reflexões sobre a internet, os negócios e a sociedade*. Rio de Janeiro: Ed. Zahar, 2003.

CAVALCANTE, L. E. Patrimônio digital e informação: política, cultura e diversidade. *Encontros Bibli*, Florianópolis, v. 12, n. 23, p. 152-170, 2007. DOI: 10.5007/1518-2924.2007v12n23p152.

CHILE. Ministerio de Educación Pública. Ley 17336 propiedad intelectual. *Biblioteca del Congreso Nacional de Chile*, 03 nov. 2017 Disponível em: <https://www.bcn.cl/leychile/navegar?idNorma=28933>. Acesso em: 23 dez. 2021.

CHILE. Ministerio Secretaría General de Gobierno. Ley 19733 sobre libertades de opinion e informacion y ejercicio del periodismo. *Biblioteca del Congreso Nacional de Chile*, 23 dec. 2013. Disponível em: <https://www.bcn.cl/leychile/navegar?idNorma=186049&buscar=19.733>. Acesso em: 23 dez. 2021.

CODEC. In: WIKIPEDIA: a enciclopédia livre. [São Francisco, CA: Wikimedia Foundation, 2021]. Disponível em: Disponível em: <https://pt.wikipedia.org/wiki/Codec>. Acesso em: 23 ago. 2021.

COLLEX-PERSÉE. *Who are we?* [c2021a]. Disponível em: <https://www.collexpersee.eu/a-propos/who-are-we/>. Acesso em: 15 maio 2021.

COLLEX-PERSÉE. *Projets*: ResPaDon. [c2021b]. Disponível em: <https://www.collexpersee.eu/projet/respadon/>. Acesso em: 15 maio 2021.

COMITÊ GESTOR DA INTERNET NO BRASIL. CGI.BR. Preservação do conteúdo web brasileiro. [Workshop] 2 out. 2019. In: FÓRUM DA INTERNET NO BRASIL, 9., 1-4 out. 2019, Manaus. [S. l.]: CGI.BR, 2019. Disponível em: <https://forumdainternet.cgi.br/2019/programacao/detalhe/2/1577/>. Acesso em: 01 jun. 2021.

CONSULTATIVE COMMITTEE FOR SPACE DATA SYSTEM – CCSDS. *Reference model for an open archival information system (OAIS)*: recommended practice, CCSDS 650.0-M-2: magenta book. Washington, DC: CCSDS, 2012. Disponível em: <https://public.ccsds.org/pubs/650x0m2.pdf>. Acesso em: 31 maio 2021.

CONWAY, P. Archival preservation practice in a nationwide context. *American Archivist*, Chicago, IL, v. 53, n. 2, p. 204-222, Mar./June 1990. Special Preservation Issue. Disponível em: <https://www.jstor.org/stable/40293442>. Acesso em: 31 maio 2021.

CONWAY, P. *Preservation in the digital world*. Alexandria: CLIR Publications, 1996. Disponível em: <https://www.clir.org/pubs/reports/conway2/index/>. Acesso em: 01 jun. 2021.

CONWAY, P. *Preservação no universo digital*. 2. ed. Rio de Janeiro: Projeto Conservação Preventiva em Bibliotecas e Arquivos: Arquivo Nacional, 2001. Disponível em: <http://www.arqsp.org.br/cpba>. Acesso em: 31 maio 2021.

CONSELHO NACIONAL DE ARQUIVOS – CONARQ. Câmara Técnica de Documentos Eletrônicos. *Glossário documentos arquivísticos digitais*. versão 8. [S. l.]: CONARQ, 2020. Disponível em: http://antigo.conarq.gov.br/images/ctde/Glossario/glosctde_2020_08_07.pdf. Acesso em: 31 maio 2021.

CONSELHO NACIONAL DE ARQUIVOS – CONARQ. Câmara Técnica de Documentos Eletrônicos. *Carta para a preservação do patrimônio arquivístico digital*. [S. l.]: CONARQ, 2005. Disponível em: http://conarq.gov.br/images/publicacoes_textos/Carta_preservacao.pdf. Acesso em: 31 de maio de 2021.

CONSELHO NACIONAL DE ARQUIVOS – CONARQ. Resolução nº 37, de 19 de dezembro de 2012. Aprova as diretrizes para a presunção de autenticidade de documentos arquivísticos digitais. *Diário Oficial da União*: seção 1, Brasília, DF, 20 dez. 2012. Disponível em: https://www.gov.br/conarq/pt-br/centrais-de-conteudo/publicacoes/conarq_presuncao_autenticidade_completa.pdf. Acesso em: 31 de maio de 2021.

CONSELHO NACIONAL DE ARQUIVOS – CONARQ. Resolução nº 43, de 04 de setembro de 2015. Altera a redação da Resolução do CONARQ nº 39, de 29 de abril de 2014, que estabelece diretrizes para a implementação de repositórios digitais confiáveis para a transferência e recolhimento de documentos arquivísticos digitais para instituições arquivísticas dos órgãos e entidades integrantes do Sistema Nacional de Arquivos – SINAR. *Diário Oficial da União*: seção 1, Brasília, DF, 08 set. 2015. Disponível em: <https://www.gov.br/conarq/pt-br/legislacao-arquivistica/resolucoes-do-conarq/resolucao-no-43-de-04-de-setembro-de-2015>. Acesso em: 03 set. 2021.

CONSELHO NACIONAL DE ARQUIVOS – CONARQ. Câmara Técnica Consultiva do Conarq sobre Preservação de websites e mídias sociais inicia seus trabalhos. *CONARQ Notícias*, 8 dez. 2021. Disponível em: <https://www.gov.br/conarq/pt-br/assuntos/noticias/camara-tecnica-consultiva-do-conarq-sobre-preservacao-de-websites-e-midias-sociais-inicia-seus-trabalhos>. Acesso em: 07 jan. 2022.

CONSELHO NACIONAL DE DESENVOLVIMENTO CIENTÍFICO E TECNOLÓGICO. Estudos e práticas de preservação digital. [S. l.]: CNPq, [c2021]. [Grupo de pesquisa]. Disponível em: dgp.cnpq.br/dgp/espelhogrupo/3997875180380796. Acesso em: 15 maio 2021.

COSTA, M.; GOMES, D.; SILVA, M. The evolution of web archiving. *International Journal on Digital Libraries*, p. 1-15, 2016. Disponível em: <https://sobre.arquivo.pt/wp-content/uploads/the-evolution-of-web-archiving.pdf>. Acesso em: 08 set. 2021.

CRIPPA, G. Patrimônio cultural: a cidade como documento. In: CRIPPA, G.; MOSTAFA, S. P. (org.). *Ciência da Informação e documentação*. Campinas: Alínea, 2011.

CUNHA, M. B.; CAVALCANTI, C. R. *Dicionário de biblioteconomia e arquivologia*. Brasília, DF: Brinquet de Lemos, 2008. 451 p.

CUNHA, J. A.; LIMA, M. G. Preservação digital: o estado da arte. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 8., 2007, Salvador. *Anais[...]*. São Paulo: ANCIB, 2007. Pôster. Disponível em: <https://www.researchgate.net/publication/279472540>. Acesso em 18 abr. 2021.

DANTAS, C. G. *Criptografias da memória: um estudo teórico-prático sobre o arquivamento da web no Brasil*. 2015. 227 f. Tese (Doutorado em Memória Social) – Universidade Federal do Estado do Rio de Janeiro, Rio de Janeiro, 2015. Disponível em: <http://www.memoriasocial.br/documentos/Teses/Tese42.pdf>. Acesso em: 29 nov. 2020.

DAY, M. Preserving the fabric of our lives: a survey of web preservation initiatives. In: EUROPEAN CONFERENCE ON RESEARCH AND ADVANCED TECHNOLOGY FOR DIGITAL LIBRARIES, 7., 2003, Trondheim. *Proceedings [...]*. Trondheim, Norway: Springer Verlag, 2003a. Disponível em: <http://www.ukoln.ac.uk/metadata/presentations/ecdl2003-day/day-paper.pdf>. Acesso em: 29 nov. 2020.

DAY, M. *Collecting and preserving the World Wide Web: a feasibility study undertaken for the JISC and Wellcome Trust*. United Kingdom: JISC: The Wellcome Trust, 2003b. Version 1.0, 25 feb. 2003. Disponível em: <https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.64.388&rep=rep1&type=pdf>. Acesso em: 18 abr. 2021.

DIGITAL CURATION CENTER. *Digital Curation 101*. [2010]. Disponível em: <https://www.dcc.ac.uk/sites/default/files/documents/DC%20101%20Curation%20and%20Preservation.pdf>. Acesso em: 20 abr. 2021.

DIGITAL PRESERVATION COALITION – DPC. *Glossary*. 2021. Disponível em: <https://www.dpconline.org/handbook/glossary#P>. Acesso em: 31 maio 2021.

DILLO, I. FAIR and TRUST a perfect couple. *UCLAC Webinar Series*, Oct. 2020. 54 slides. Disponível em: https://www.cepal.org/sites/default/files/news/files/fair_and_trust_ingrid_dillo_webinar_10-2020.pdf. Acesso em: 03 set. 2021.

DINAMARCA. Ministério da Cultura. *Lei 1439 de 22 de dezembro de 2004*. Disponível em: <https://www.retsinformation.dk/eli/lta/2004/1439> . Acesso em: 24 set. 2021.

DODEBEI, V. Contribuições das teorias da memória para o estudo do patrimônio na web. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 7., 2006, Marília, SP. *Anais [...]*. Marília: UNESP, 2006. Disponível em: <http://repositorios.questoesemrede.uff.br/repositorios/handle/123456789/494?show=full>. Acesso em: 27 maio 2021.

DOOLEY, J; BOWERS, K. *Descriptive metadata for web archiving*: recommendations of the OCLC research library partnership web archiving metadata working group. Dublin, OH: OCLC Research, 2018. DOI: <https://doi.org/10.25333/C3005C>.

DOMBROWSKI, Q.; KIJAS, A.; MAJSTOROVIC, S. *Saving Ukrainian Cultural Heritage Online (SUCHO)*. 2022. Disponível em: <https://www.sucho.org/>. Acesso em: 12 abr. 2022.

DOUGHERTY, M. *et al. Researcher engagement with web archives*: state of the art. London: JISC, 2010. Disponível em: https://ecommons.luc.edu/communication_facpubs/14/. Acesso em: 28 fev. 2021.

DUBLIN CORE METADATA INITIATIVE. *The Dublin Core*. [c1995-2021]. Disponível em: https://www.dublincore.org/resources/glossary/dublin_core/. Acesso em: 03 set. 2021.

ENGLE, E. Legal issues in web archiving. *The Signal*, 30 maio 2012. Disponível em: <https://blogs.loc.gov/thesignal/2012/05/legal-issues-in-web-archiving/>. Acesso em: 01 dez. 2021.

DURANTI, L. The long-term preservation of the digital heritage: a case study of universities institutional repositories. *JLIS.it.*, v. 1, n. 1, p. 157-168, giugno 2010. DOI: <http://dx.doi.org/10.4403/jlis.it-12>.

DURANTI, L. *InterPARES Project*: the future of our digital memory. [200?]. 17 slides. Disponível em: http://www.unesco.org/new/fileadmin/MULTIMEDIA/HQ/CI/CI/pdf/mow/inter pares_en.pdf. Acesso em: 01 jun. 2021.

EUROPEAN COMMISSION. *European digital heritage*. [c2021]. Disponível em: <https://ec.europa.eu/culture/cultural-heritage/cultural-heritage-eu-policies/european-digital-heritage>. Acesso em: 31 maio 2021.

ENGAGING WITH WEB ARCHIVES – EWA. (#EWAVirtual), 21-22 Sept. 2020, County Kildare, Ireland. *Anais eletrônicos [...]*. County Kildare, Ireland: Maynooth University Arts and Humanities Institute, 2020. Disponível em: <https://ewaconference.com/>. Acesso em: 31 maio 2021.

ENGAGING WITH WEB ARCHIVES – EWA. (#EWAVirtual), 21-22 Sept. 2020, County Kildare, Ireland. ‘Opportunities, challenges and potentialities’. Published by EWA Conference channel. Disponível em: https://www.youtube.com/channel/UCxT4WqoDaO3B_Hvhr6rpB6Q. Acesso em: 01 jun. 2021.

ENGAGING WITH WEB ARCHIVES – EWA. EWA Conference, [c2021]. [S. l.: s. n.], [c2021]. Disponível em: <https://ewaconference.com/>. Acesso em: 01 jun. 2021.

EUROPEAN COMMISSION. *Horizon 2020*. Disponível em: <https://ec.europa.eu/programmes/horizon2020/en/what-horizon-2020>. Acesso em: 23 ago. 2021.

ESPAÑA. Ministerio de Educación, Cultura y Deporte. *Real Decreto 635/2015, de 10 de julio, por el que se regula el depósito legal de las publicaciones en línea*. Disponível em: <https://www.boe.es/eli/es/rd/2015/07/10/635>. Acesso em: 14 dez. 2021.

FAIR data principles. 17 nov. 2016. In: WIKIMIDIA commons: a midiateca livre. Flórida: Wikimedia Foundation, 2022. Disponível em: https://upload.wikimedia.org/wikipedia/commons/a/aa/FAIR_data_principles.jpg. Acesso em: 11 abr. 2022.

FARIAS, J. P.; BOMFIM, K. C. V. A produção científica sobre preservação de websites em língua portuguesa. *Encontros Bibli: Revista Eletrônica de Biblioteconomia e Ciência da Informação*, v. 24, n. 55, p. 1-15, 2019. DOI: 10.5007/1518-2924.2019.e58207.

FAZANO, I. F.; FONSECAV. L. Web archiving como ferramenta de preservação digital: colaboração da Library of Congress, Rio Office no projeto dos Jogos Olímpicos de 2016. *Ciência da Informação*, Brasília, DF, v. 48, n. 3, 20 mar. 2020. Disponível em: <http://revista.ibict.br/ciinf/article/view/4993>. Acesso em: 11 mar. 2021.

FERREIRA, M. *Introdução a preservação digital: conceitos, estratégias e actuais consensos*. Guimarães, Portugal: Escola de Engenharia da Universidade do Minho, 2006.

FERREIRA, L. B. *Arquivamento da web e mídias sociais: preservação digital de vídeos da campanha presidencial brasileira de 2018*. 2019. 105 f. Dissertação (Mestrado em Comunicação e Informação) – Faculdade de Biblioteconomia e Comunicação, Universidade Federal Rio Grande do Sul, Porto Alegre, 2019. Disponível em: <https://lume.ufrgs.br/handle/10183/194617>. Acesso em: 29 nov. 2020.

FINLÂNDIA. Ministerio da Educação. *Lei 1433 de 28 de dezembro de 2007*. Disponível em: <https://finlex.fi/fi/laki/alkup/2007/20071433#Lidm45237817194960>. Acesso em: 29 set. 2021.

FORMENTON, D. *et al.* Os padrões de metadados como recursos tecnológicos para a garantia da preservação digital. *Biblios: Journal of Librarianship and Information Science*, Tacna, Peru, n. 68, p. 82-95, dez. 2017. Disponível em: <http://biblios.pitt.edu/ojs/index.php/biblios/article/view/414/311>. Acesso em: 08 set. 2021.

FORMENTON, D. *Identificação de padrões de metadados para preservação digital*. 2015. 101 f. Dissertação (Mestrado em Ciência, Tecnologia e Sociedade, do Centro de Educação e Ciências Humanas) – Universidade Federal de São Carlos, São Paulo, 2015. Disponível em: <https://repositorio.ufscar.br/bitstream/handle/ufscar/7221/DissDF.pdf?sequence=1>. Acesso em: 8 set. 2021.

FORMENTON, D.; GRACIOSO, L. S. Padrões de metadados no arquivamento da Web: recursos tecnológicos para a garantia da preservação digital de websites arquivados. *RDBCi: Revista Digital de Biblioteconomia e Ciência da Informação*, Campinas, v. 20, n. 00, p. e022001, 2022. DOI: 10.20396/rdbci.v20i00.8666263. Disponível em: <https://periodicos.sbu.unicamp.br/ojs/index.php/rdbci/article/view/8666263>. Acesso em: 25 jan. 2022.

FLORES, D. Preservação de páginas web e redes sociais em cadeia de custódia: identificação, seleção e arquivamento. In: ENCONTRO NACIONAL DE MEMÓRIA DO PODER JUDICIÁRIO, 1., 2021, Brasília, DF. Palestra on-line, 43 slides. Brasília, DF: CNJ, 2021. Disponível em: <http://documentosdigitais.blogspot.com>. Acesso em: 18 maio 2021.

FLORES, D. Preservação de páginas web e redes sociais em cadeia de custódia: identificação, seleção e arquivamento. In: ENCONTRO NACIONAL DE MEMÓRIA DO PODER JUDICIÁRIO, 1., 2021, Brasília, DF. Palestra on-line, 18 maio 2021. Brasília, DF: CNJ, 2021. 1 vídeo (160 min). Publicado pelo canal Conselho Nacional de Justiça. Disponível em: <https://www.youtube.com/watch?v=RkR2RjaSjC0>. Acesso em: 31 maio 2021.

FONSECA, M. O. *Arquivologia e a Ciência da Informação*. Rio de Janeiro: Ed. FGV, 2005. 124 p.

FOUNDATION FOR ADVANCEMENT IN CONSERVATION. AIC definitions of conservation terminology. *WAAC Newsletter*, v. 18, n. 2, May 1996. Disponível em: <https://cool.culturalheritage.org/waac/wn/wn18/wn18-2/wn18-202.html>. Acesso em: 31 maio 2021

FRANÇA. *Loi n° 2006-961 du 1 août 2006*. Relative au droit d'auteur et aux droits voisins dans la société de l'information. 2006. Disponível em: <https://www.legifrance.gouv.fr/loda/id/JORFTEXT000000266350/>. Acesso em: 16 mar. 2021.

FRANÇA. Collecte des sites web français liés à l'audiovisuel français. 2019. Disponível em: <https://www.data.gouv.fr/fr/datasets/collecte-des-sites-web-francais-lies-a-laudiovisuel-francais/>. Acesso em: 16 mar. 2021.

FUNDAÇÃO BIBLIOTECA NACIONAL (Brasil). JORNADA DE PESQUISADORES, 7., 2021. Conferência de abertura do Diretor Técnico da Biblioteca Nacional da Espanha, Sr. José Luis Bueren Gómez-Acebo. 19 out. 2021. Disponível em: <https://www.bn.gov.br/acontece/noticias/2021/10/jornada-pesquisadores-conferencia-abertura-7a-edicao>. Acesso em: 14 dez. 2021.

FUNDAÇÃO BIBLIOTECA NACIONAL (Brasil). *Portaria FBN n° 8, de 8 de fevereiro de 2021*. Disponível em: <https://www.in.gov.br/web/dou/-/portaria-fbn-n-8-de-8-de-fevereiro-de-2021-304722636>. Acesso em: 23 ago. 2021.

FUNDAÇÃO BIBLIOTECA NACIONAL (Brasil). *Portaria FBN/PRESI n° 34 de 27 de agosto de 2019*. Disponível em: https://sei.bn.gov.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0. Acesso em: 23 ago. 2021.

FUNDAÇÃO BIBLIOTECA NACIONAL (Brasil). Servidoras da BN participam de curso sobre atuais desafios do gerenciamento de arquivos. *BN Notícias*, dez. 2019. Disponível em: <https://www.bn.gov.br/acontece/noticias/2019/12/servidoras-bn-participam-curso-sobre-atuais-desafios>. Acesso em: 23 ago. 2021.

FUNDAÇÃO DE AMPARO À PESQUISA DO ESTADO DE SÃO PAULO. *Open Science @ FAPESP*, [c2021]. Disponível em: <https://www.fapesp.br/openscience/>. Acesso em: 03 Set. 2021.

FUNDAÇÃO PARA COMPUTAÇÃO CIENTÍFICA NACIONAL – FCT. *Quem é a FCCN?* [c2021]. Disponível em: <https://www.fcn.pt/quem-somos/>. Acesso em: 12 dez. 2021.

FUNDAÇÃO PARA A CIÊNCIA E A TECNOLOGIA – FCT. *Sobre a FCT*. [c2021]. Disponível em: <https://www.fct.pt/fct>. Acesso em: 12 dez. 2021.

FUNDAÇÃO PARA A CIÊNCIA E A TECNOLOGIA – FCT. *Plano de Atividades 2021*. [c2021b]. Disponível em: <https://www.fct.pt/docs/PA2021.pdf>. Acesso em: 10 dez. 2021.
FUNDAÇÃO PRO-MEMÓRIA DE SÃO CARLOS. *Dicionário da Pró-Memória*. São Carlos, SP: Pró-Memória, c2021. Disponível em: https://www.promemoria.saocarlos.sp.gov.br/index.php?option=com_content&view=article&id=145&Itemid=199. Acesso em: 31 maio 2021.

FUNDACION OLGA GALLEGU. *Fallo da IV edición do Premio Olga Gallego de Investigación en Arquivos*. 08 mar. 2021. Disponível em: <https://www.fundacionolga.gallego.gal/actualidade/fallo-da-iv-edicion-do-premio-olga-gallego-de-investigacion-en-arquivos.htm>. Acesso em: 08 jun. 2021.

GARRETT, J. An evaluation of web archiving programs in the us relevant to international and area studies: the example of Latin America and the Caribbean. *Center for Research Libraries, Global Resources Network*, Feb. 2019. DOI:10.13140/RG.2.2.23094.34885. Disponível em: <https://www.crl.edu/reports/web-archiving-programs-latin-america-and-caribbean>. Acesso em: 29 nov. 2020.

GIL, A. C. *Métodos e técnicas de pesquisa social*. São Paulo: Atlas, 1989.

GLUSHKO, R. J. The What and why of categories. In: GLUSHKO, R. J. (ed.). *The discipline of organizing*. 4nd. ed. Berkeley, CA: University of California, 2020. Disponível em: <https://human.libretexts.org/@go/page/73608>. Acesso em: 7 dez. 2021.

GO FAIR. FAIR Principles. [2021]. Disponível em: <https://www.go-fair.org/fair-principles/>. Acesso em: 11 abr. 2022.

GOETHALS, K.; AGUIAR, A.; ALMEIDA, E. *História da internet*. Trabalho de Disciplina (Gestão e Organização da Informação na Internet) – Pós-Graduação em Gestão da Informação, Faculdade de Engenharia, Universidade do Porto, Porto, 2000. Disponível em: <https://web.fe.up.pt/~mgi99015/projects/Goii/M1/final.doc>. Acesso em: 29 nov. 2020.

GOEL, V. Defining web pages, web sites and web captures. *Internet Archive Blogs*, 23 out. 2016. Disponível em: <https://blog.archive.org/2016/10/23/defining-web-pages-web-sites-and-web-captures/>. Acesso em: 31 maio 2021.

GOMES, D.; MIRANDA, J.; COSTA, M. A survey on web archiving initiatives. In: INTERNATIONAL CONFERENCE ON THEORY AND PRACTICE OF DIGITAL LIBRARIES, 1., 2011, Berlim. *Proceedings* [...]. Berlim: [s. n.], 2011. Disponível: <https://sobre.arquivo.pt/wp-content/uploads/a-survey-on-web-archiving-initiatives.pdf>. Acesso em: 06 jul. 2019.

GOMES, D.; COSTA, M. The importance of web archives for humanities. *International Journal of Humanities and Arts Computing*, Edinburgh, v. 8, n. 1, p. 106-123, Apr. 2014. Disponível em: <https://pdfs.semanticscholar.org/d36c/1a2e5fd2cd18357de2a24357d5612773cd75.pdf>. Acesso em: 31 maio 2021.

GOODING, P.; TERRAS, M. *Electronic legal deposit: shaping the library collections of the future*. 2nd. ed. London: Facet Publishing, 2020. 256 p.

GOVERNMENT OF CANADA. *Library and Archives of Canada Act (S.C. 2004, c. 11)*. 22/04/2004. Atualizado em 11/11/2021. Disponível em: <https://laws-lois.justice.gc.ca/eng/acts/L-7.7/index.html>. Acesso em: 17 nov. 2021.

GOVERNMENT OF CANADA. *Legal Deposit of Publications Regulations SOR/2006-337*. 12/12/2006. Disponível em: <https://laws-lois.justice.gc.ca/eng/regulations/SOR-2006-337/page-1.html>. Acesso em: 17 nov. 2021.

GOUVERNEMENT DU QUÉBEC. *Loi Sur Bibliothèque et Archives Nationales du Québec, cap. B-1.2, 2001*. Atualizada 01/10/2021. Disponível em: <http://legisquebec.gouv.qc.ca/fr/document/lc/B-1.2>. Acesso em: 30 nov. 2021.

GRÁCIO, J. C. A. *Preservação digital na gestão da informação: um modelo processual para as instituições de ensino superior: um modelo processual para as instituições de ensino superior*. São Paulo: Cultura Acadêmica, 2012. Disponível em: <https://repositorio.unesp.br/bitstream/handle/11449/113727/ISBN9788579833335.pdf?sequence=1&isAllowed=y>. Acesso em: 09 mar. 2021.

GUERRA dos navegadores. In: WIKIPÉDIA: a enciclopédia livre. [São Francisco, CA: Wikimedia Foundation, 2021]. Disponível em: https://pt.wikipedia.org/wiki/Guerra_dos_navegadores. Acesso em: 16 abr. 2021.

GITHUB. In: WIKIPEDIA: a enciclopédia livre. [São Francisco, CA: Wikimedia Foundation, 2022]. Disponível em: Disponível em: <https://pt.wikipedia.org/wiki/GitHub>. Acesso em: 15 fev. 2022.

GITHUB INC. GitHub policy: fighting for developers. [c2022]. Disponível em: <https://github.com/about/developer-policy>. Acesso em: 15 fev. 2022.

HEDSTROM, M. Digital preservation: a time bomb for digital libraries. *Computers and the Humanities*, v. 31, p. 189-202, 1998. Disponível em: <https://www.researchgate.net/publication/30843346>. Acesso em: 31 maio 2021.

HIELMCRUNE, H. *Pligtaflevering af dansk materiale, offentliggjort på Internettet*. 2005. Disponível em: <http://www.pligtaflevering.dk/online/vejledning.pdf>. Acesso em: 24 set. 2021.

HIRTLE, P. B. Archival authenticity in a digital age. *In: AUTHENTICITY in a digital environment*. May 2000. Washington, DC: Council on Library and Information Resources, 2000. p. 8-23. Disponível em: <https://www.clir.org/wp-content/uploads/sites/6/pub92.pdf>. Acesso em: 31 maio 2021.

HOCKX-YU, H. The past issue of the web. *In: INTERNATIONAL WEB SCIENCE CONFERENCE*, 3., 2011, New York. *Proceedings* [...]. New York, NY: Association for Computing Machinery, 2011. p. 1-8. DOI: <https://doi.org/10.1145/2527031.2527050>.

HOCKX-YU, H. Scaling up to archive the UK web. *UK Web archive blog*, 4 set. 2013. Disponível em: <https://blogs.bl.uk/webarchive/2013/09/scaling-up-to-archive-the-uk-web.html>. Acesso em: 20 set. 2021.

INNARELLI, H. C. Preservação digital e seus dez mandamentos. *In: SANTOS, V.; INNARELLI, H.; SOUSA, R. Arquivística: temas contemporâneos: classificação, preservação digital, gestão do conhecimento*. 3. ed. Brasília, DF: SENAC, 2009. p. 21-75.

INTERNET ARCHIVE. *About the internet archive*. 2020. Disponível em: <https://web.archive.org/web/20200602101307/https://archive.org/about/>. Acesso em: 29 nov. 2020.

INTERNET ARCHIVE. *Building a digital library for the future: our mission*. 26 jan. 1997. Disponível em: <https://web.archive.org/web/19970126045828/http://www.archive.org/>. Acesso em: 21 set. 2021.

INTERNET ARCHIVE. *About the internet archive*. [2021a]. Disponível em: <https://archive.org/about/>. Acesso em: 21 set. 2021.

INTERNET ARCHIVE. *Celebrate the Internet Archive's 25th Anniversary*. [2021b]. Disponível em: <https://anniversary.archive.org/#welcome>. Acesso em: 21 set. 2021.

INTERNET ARCHIVE. *Web & data service*. [2021c]. Disponível em: <https://webservices.archive.org/>. Acesso em: 21 set. 2021.

INTERNET ARCHIVE. *Building a digital library for the future: our mission*. 26 jan. 1997. Disponível em: <https://web.archive.org/web/19970126045828/http://www.archive.org/>. Acesso em: 31 maio 2021.

INSTITUT NATIONAL DE L'AUDIOVISUEL. *Rapport d'activités 2020*. Disponível em: <https://www.ina.fr/institut-national-audiovisuel/organisation/rapport-d-activites>. Acesso em: 05 jan. 2022.

INSTITUT NATIONAL DE L'AUDIOVISUEL. *Recherchez: dans le catalogue du dépôt legal: web media*. [2021a]. Disponível em: <http://inatheque.ina.fr/index/WEB-MEDIA/>. Acesso em: 03 set. 2021.

INSTITUT NATIONAL DE L'AUDIOVISUEL. *Le web média*. [2021b]. Disponível em: <https://www.ina.fr/institut-national-audiovisuel/collections-audiovisuelles/le-web-media>. Acesso em: 03 set. 2021.

INSTITUT NATIONAL DE L'AUDIOVISUEL. *Legal deposit*. [c2022]. Disponível em: <https://www.ina.fr/offres-et-services/depot-legal>. Acesso em: 05 jan. 2022.

INTERNATIONAL FEDERATION OF LIBRARY ASSOCIATIONS AND INSTITUTIONS – IFLA. *IFLA Statement on Legal Deposit (2011)*. Última atualização em 4 set. 2013. Disponível em: <https://www.ifla.org/publications/ifla-statement-on-legal-deposit>. Acesso em: 31 maio 2021.

INTERNATIONAL INTERNET PRESERVATION CONSORTIUM – IIPC. UNT Digital Library. University of North Texas Libraries. Disponível em: <https://digital.library.unt.edu/explore/partners/IIPC/>. Acesso em: 7 jan. 2022.

INTERNATIONAL INTERNET PRESERVATION CONSORTIUM – IIPC. *The British Library*. [2019]. Disponível em: <https://netpreserve.org/about-us/members/british-library/>. Acesso em: 20 set. 2021.

INTERNATIONAL INTERNET PRESERVATION CONSORTIUM – IIPC. *About us*. c2020a. Disponível em: <https://web.archive.org/web/20201203154753/https://netpreserve.org/about-us/>. Acesso em: 01 jun. 2021.

INTERNATIONAL INTERNET PRESERVATION CONSORTIUM – IIPC. *About us*. c2021a. Disponível em: <https://netpreserve.org/about-us/>. Acesso em 02 mar. 2021.

INTERNATIONAL INTERNET PRESERVATION CONSORTIUM – IIPC. *Members*. c2021b. Disponível em: <https://netpreserve.org/about-us/members/>. Acesso em 02 mar. 2021.

INTERNATIONAL INTERNET PRESERVATION CONSORTIUM – IIPC. *IIPC General Assembly & Web Archiving Conference*. c2021c. Disponível em: <https://netpreserve.org/ga2021/>. Acesso em: 17 jun. 2021.

INTERNATIONAL INTERNET PRESERVATION CONSORTIUM – IIPC. *Press Release*. 5 maio 2004. Capturada em 3 jun. 2004, 05:43:12. Disponível em: <http://web.archive.org/web/20040603054312/http://netpreserve.org/press/pr20040505.php>. Acesso em 02 mar. 2021.

INTERNATIONAL INTERNET PRESERVATION CONSORTIUM – IIPC. *IIPC strategic plan*. 2016. Disponível em: <https://netpreserve.org/resources/IIPC-Strategic-Plan.pdf>. Acesso em: 1 jun. 2021.

INTERNATIONAL INTERNET PRESERVATION CONSORTIUM – IIPC. *2021-2025 consortium agreement of the International Internet Preservation Consortium (IIPC)*. 2021c. Disponível em: <https://netpreserve.org/resources/IIPC-Consortium-Agreement-2021-2025.pdf>. Acesso em: 01 jun. 2021.

INTERNATIONAL INTERNET PRESERVATION CONSORTIUM – IIPC. *The WARC Format 1.1*. 2021d. Disponível em: <https://iipc.github.io/warc-specifications/specifications/warc-format/warc-1.1/>. Acesso em: 03 set. 2021.

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION – ISO. *BS ISO 28500: 2009: information and documentation: WARC file format*. Switzerland: ISO, 2009.

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION – ISO. *BS ISO 28500*: 2017: Information and documentation: WARC file format. Switzerland: ISO, 2017. Disponível em: <https://www.iso.org/obp/ui/#iso:std:iso:28500:ed-2:v1:en>. Acesso em: 10 fev. 2021.

INTERNET. *In*: WIKIPEDIA: a enciclopédia livre. [São Francisco, CA: Wikimedia Foundation, 2021]. Disponível em: <https://pt.wikipedia.org/wiki/Internet#:~:text=A%20Internet%20%C3%A9%20um%20sistema,progressivamente%20usu%C3%A1rios%20no%20mundo%20inteiro>. Acesso em: 31 maio 2021.

INTERNET LIVE STATS. [c2021]. Disponível em: <https://www.internetlivestats.com/>. Acesso em: 10 fev. 2021.

INTERNET MEMORY FOUNDATION. *Archivethe.net*. 2015. Disponível em: <https://web.archive.org/web/20150206044401/http://internetmemory.org/en/index.php/projects/atn>. Acesso em: 31 maio 2021.

INTERNATIONAL WEB ARCHIVING WORKSHOPS. *BnF*, 2006. Disponível em: <http://bibnum.bnf.fr/ecdl/>. Acesso em: 31 maio 2021.

INTERNATIONAL WEB ARCHIVING WORKSHOPS. *IWAW 2010*: The 10th Intl Web Archiving Workshop. Disponível em: <http://www.wikicfp.com/cfp/servlet/event.showcfp?eventid=9651>. Acesso em: 31 maio 2021.

INTERNET MEMORY FOUNDATION. The foundation. *Internet Archive*, 29 dec. 2014, 06:46:45 GMT. Disponível em: <https://web.archive.org/web/20141229064645/http://internetmemory.org/en/>. Acesso em: 31 maio 2021.

INTERNET MEMORY FOUNDATION. Internet memory foundation: about the project. *Internet Archive*, 26 dec 2014, 15:26:34 GMT. Disponível em: <https://web.archive.org/web/20141226152634/http://internetmemory.org/en/index.php/projects>. Acesso em: 31 maio 2021.

INTERPARES 2. *Terminology Database*. 2002. Disponível em: http://www.interpares.org/ip2/ip2_terminology_db.cfm. Acesso em: 01 jun. 2021.

INTERPARES 2 project. *Project summary*. [c2021a]. Disponível em: http://www.interpares.org/ip2/ip2_index.cfm. Acesso em: 09 jun. 2021.

INTERPARES project: International Research on Permanent Authentic Records in Electronic Systems. [c2021b]. Disponível em: <http://www.interpares.org/>. Acesso em: 31 maio 2021.

JARDIM, J. M.; FONSECA, M. O. Arquivos. *In*: CAMPELLO, B.; CALDEIRA, P. T. (org.). *Introdução às fontes de informação*. 2. ed. Belo Horizonte: Autêntica, 2008. p. 121-139.

KHAN, M.; RAHMAN, A. U. A systematic approach towards web preservation. *Information Technology and Libraries*, v. 38, n. 1, p. 71-90, Mar. 2019. DOI: <https://doi.org/10.6017/ital.v38i1.10181>.

KAHLE, B. Reflections as the Internet Archive turns 25. *Internet Archive Blogs*, 21 jul. 2021. Disponível em: <https://blog.archive.org/2021/07/21/reflections-as-the-internet-archive-turns-25/>. Acesso em: 21 set. 2021.

LARIVIÈRE, J. *Guidelines for legal deposit legislation*. Paris: UNESCO, 2000. Disponível em: <https://www.ifla.org/files/assets/national-libraries/publications/guidelines-for-legal-deposit-legislation-en.pdf>. Acesso em: 27 maio 2021.

LEINER, B. M. *et al.* Brief history of the internet. *Internet Society*, 1997. Disponível em: <https://www.internetsociety.org/internet/history-internet/brief-history-internet/>. Acesso em: 07 mar. 2021.

LIBNOVA. *Libsafe*: digital preservation plan design and management. Madrid: LIBNOVA, 2014. Disponível em: <http://www.digitalpreservationsoftware.com/wp-content/uploads/documents/20140401-libsafe-features-requirements-eng.pdf>. Acesso em: 15 fev. 2022.

LIBNOVA. *Libsafe*: the software that makes digital preservation easy. [c2016]. Disponível em: <http://www.digitalpreservationsoftware.com/digital-preservation-solutions/libsafe-digital-preservation-software/>. Acesso em: 15 fev. 2022.

LIBRARY OF CONGRESS (U.S.). *Web archiving* [program]: glossary. [Washington, DC, 2021a]. Disponível em: <https://www.loc.gov/programs/web-archiving/about-this-program/glossary/>. Acesso em: 31 maio 2021.

LIBRARY OF CONGRESS (U.S.). *Program web archiving*. [Washington, DC, 2021b]. Disponível em: <https://www.loc.gov/programs/web-archiving/about-this-program/>. Acesso em: 01 Dez. 2021.

LIBRARY OF CONGRESS (U.S.). *Frequently asked questions*. [Washington, DC, 2021c]. Disponível em: <https://www.loc.gov/programs/web-archiving/about-this-program/frequently-asked-questions/>. Acesso em: 01 dez. 2021.

LIBRARY OF CONGRESS (U.S.). *Digital preservation*. [Washington, DC, 2021d]. Disponível em: <https://www.digitalpreservation.gov/>. Acesso em: 03 dez. 2021.

LIBRARY OF CONGRESS (U.S.). *Preservation home*. [Washington, DC, 2021e]. Disponível em: <https://www.loc.gov/preservation/>. Acesso em: 03 dez. 2021

LIBRARY OF CONGRESS (U.S.). *Digital strategy*. [Washington, DC, 2021f]. Disponível em: <https://www.loc.gov/digital-strategy>. Acesso em: 03 dez. 2021.

LIBRARY OF CONGRESS (U.S.). *Digital collections strategy overview 2022-2026*. Washington, DC, Oct. 2021g. Disponível em: https://www.loc.gov/acq/devpol/Digital%20Collections%20Strategy%20Overview_final.pdf. Acesso em: 03 dez. 2021.

LIBRARY OF CONGRESS (U.S.). *The K-12 Web Archiving Project*. [Washington, DC, 2021h]. Disponível em: <https://www.digitalpreservation.gov/series/k12archiving.html>. Acesso em: 03 dez. 2021.

LIBRARY OF CONGRESS (U.S.). *Program digital collections management*. [2021i]. Disponível em: <https://www.loc.gov/programs/digital-collections-management/about-this-program/frequently-asked-questions/>. Acesso em: 03 dez. 2021.

LIBRARY OF CONGRESS (U.S.). *Web archiving supplementary guidelines*. Sept. 2017. Disponível em: <https://www.loc.gov/acq/devpol/webarchive.pdf>. Acesso em: 01 dez. 2021.

LIBRARY AND ARCHIVES CANADA. *Strategy for a digital preservation program*. Nov. 2017. Disponível em: <https://www.bac-lac.gc.ca/eng/about-us/publications/Documents/LAC-Strategy-Digital-Preservation-Program.pdf>. Acesso em: 24 nov. 2021.

LIBRARY AND ARCHIVES CANADA. *Government of Canada web archive*. Aug. 20, 2018a. Disponível em: <https://wayback.archive-it.org/7778/20180820180352/http://www.bac-lac.gc.ca/eng/discover/archives-web-government/Pages/web-archives.aspx>. Acesso em: 17 nov. 2021.

LIBRARY AND ARCHIVES CANADA. *Digital preservation program*. Jan. 09, 2018b. Disponível em: <https://www.bac-lac.gc.ca/eng/services/digital-preservation-program/Pages/default.aspx>. Acesso em: 08 fev. 2022.

LIBRARY AND ARCHIVES CANADA. *Web archiving*. Dec. 31, 2019. Disponível em: <https://www.bac-lac.gc.ca/eng/transparency/briefing/2019-transition-material/Pages/digital-web-archiving.aspx>. Acesso em: 24. nov. 2021.

LIN, D. The TRUST principles for trustworthy data repositories: an update. *Research Data Alliance*, Sept. 12, 2019. Disponível: <https://www.rd-alliance.org/trust-principles-trustworthy-data-repositories-%E2%80%93-update>. Acesso em: 03 set. 2021.

LIST of web archiving initiatives. In: WIKIPEDIA: a enciclopédia livre. [São Francisco, CA: Wikimedia Foundation, 2021]. Disponível em: https://en.wikipedia.org/wiki/List_of_Web_archiving_initiatives. Acesso em: 10 fev. 2021.

LIVING INTERNET. *TCP/IP Internet Protocol*, [2021]. Disponível em: https://www.livinginternet.com/internet/i/ii_tcpip.htm. Acesso em: 07 mar. 2021.

LLUECA, C. *et al.* *CAT (Curator Archiving Tool)*: millorant l'accés als arxius web, presentat a International Internet Preservation Consortium meeting (Viena 2010). Disponível em: https://www.padicat.cat/sites/default/files/ponencies/padicat_iipc_2010_cat.pdf. Acesso em: 14 dez. 2021.

LOHNDORF, J. What is web archiving? *Archive It Help Center*, [2021a]. Disponível em: <https://support.archive-it.org/hc/en-us/articles/360041674111-What-is-web-archiving>. Acesso em: 21 set. 2021.

LOHNDORF, J. Guide for new Archive-It users. *Archive It Help Center*, [2021b]. Disponível em: <https://support.archive-it.org/hc/en-us/articles/360041250172-Guide-for-new-Archive-It-users>. Acesso em: 21 set. 2021.

LOR, P. J.; SONNEKUS, E. A. S. *Guidelines for legislation for national libraries services*. Paris: UNESCO, 1997. 86 p. Disponível em: <https://unesdoc.unesco.org/ark:/48223/pf0000109568>. Acesso em: 27 maio 2021.

LOTT, Y. M.; CIANCONI, R. Vigilância e privacidade, no contexto do big data e dados pessoais: análise da produção da Ciência da Informação no Brasil. *Perspectivas em Ciência da Informação*, Belo Horizonte, v. 23, n. 04, p. 117-132, 2018. DOI: <https://doi.org/10.1590/1981-5344/3313>.

LUZ, A. *Comunicação pública e memória comunicacional: revelações e apagamentos sobre o governo da presidenta Dilma Rousseff*. 2021. 253 f. Tese (Doutorado em Comunicação e Informação) – Faculdade de Biblioteconomia e Comunicação, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2021. Disponível em: <http://hdl.handle.net/10183/235405>. Acesso em: 23 fev. 2022.

LYMAN, P. Archiving the World Wide Web. In: COUNCIL ON LIBRARY AND INFORMATION RESOURCES; Library of Congress. *Building a national strategy for digital preservation: issues in digital media archiving*. Washington, DC: Council on Library and Information Resources: Library of Congress, 2002. p. 38-51. Disponível em: <https://www.clir.org/pubs/reports/pub106/web/>. Acesso em: 10 fev. 2021.

MARCIAL, E.; VIEIRA, J. S. Memória institucional em risco. *Revista Ibero-Americana de Ciência da Informação*, Brasília, DF, v. 14, n. 1, p. 150-170, 2021. DOI: 10.26512/rici.v14.n1.2021.31252.

MÁRDERO ARELLANO, M. A. Preservação de documentos digitais. *Ciência da Informação*, Brasília, DF, v. 33, n. 2, p. 15-27, maio/ago. 2004. DOI: <https://doi.org/10.18225/ci.inf.v33i2.1043>.

MÁRDERO ARELLANO, M. A. *Critérios para a preservação digital da informação científica*. 2008. 354 f. Tese (Doutorado em Ciência da Informação) – Departamento de Ciência da Informação e Documentação, Universidade de Brasília, Brasília, DF. Disponível em: <https://repositorio.unb.br/handle/10482/1518>. Acesso em: 31 maio 2021.

MÁRDERO ARELLANO, M. A. A preservação de repositórios digitais dentro do modelo OAIS. In: SEMINÁRIO DE REPOSITÓRIOS DIGITAIS DE ACESSO ABERTO: PRÁTICAS E DESAFIOS. Rio de Janeiro: FGV, 2017. 63 slides. Disponível em: <https://bibliotecadigital.fgv.br/dspace/handle/10438/19050>. Acesso em: 31 maio 2021.

MÁRDERO ARELLANO, M. A.; SANTOS, G. C. (org.). *Bibliografia sobre preservação digital: um levantamento nos diversos suportes informacionais*. Campinas: BCCL/UNICAMP, 2021. E-book (Coleção Preservação digital, v. 1). DOI: <https://doi.org/10.20396/ISBN9786588816110>.

MARTINS, M. R. *Mapeamento de públicos estratégicos para iniciativas brasileiras universitárias de arquivamento da web no âmbito acadêmico: uma projeção para Universidade Federal do Rio Grande do Sul e seu Programa de Pós-Graduação em Comunicação*. 2019. 124 f. Dissertação (Mestrado em Comunicação e Informação) – Universidade Federal Rio Grande do Sul, Porto Alegre, 2019. Disponível em: <https://lume.ufrgs.br/handle/10183/189170>. Acesso em: 12 jul. 2019.

MASANÈS, J. *Web archiving*. Berlin: Springer, 2006. 234 p.

MASANÈS, J.; MAJOR, D.; GOMES, D. The past web: a look into the future. In: GOMES, D. *et al.* (ed.). *The past web: exploring web archives*. Suíça: Springer, 2021. p. 285-291. Disponível em: <https://www.springer.com/gp/book/9783030632908>. Acesso em: 31 maio 2021.

MELO, J. F. *Arquivamento dos websites do governo federal brasileiro: preservação do domínio GOV.BR*. 2020.133 f. Dissertação (Mestrado em Comunicação e Informação) – Faculdade de Biblioteconomia e Comunicação, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2020. Disponível em: <http://hdl.handle.net/10183/210671>. Acesso em: 22 nov. 2020.

MELO, J.; ROCKEMBACH, M. Arquivabilidade de websites para preservação digital: estudo a partir da área da saúde. *Revista Eletrônica de Comunicação, Informação e Inovação em Saúde*, Rio de Janeiro, v. 14, n. 3, 2020. DOI: <https://doi.org/10.29397/reciis.v14i3.2116>. Disponível em: <https://www.reciis.iciet.fiocruz.br/index.php/reciis/article/view/2116>. Acesso em: 01 jun. 2021.

MIRANDA, M.; BORBA, V. Metadados de preservação: definições e aplicações. In: SIEBRA, S; BORBA, V (org.). *Preservação digital e suas facetas*. São Carlos: Pedro & João Editores, 2021. 348 p. *E-book*. Disponível em: https://www.ufpe.br/documents/39626/3547884/Ebook_Preservac%C2%B8a~o_Digital_e_suas_Facetas_2021.pdf/128192be-71fe-4b9b-b685-28fbee9d928d. Acesso em: 8 set. 2021.

MOORE, R. Towards a theory of digital preservation. *The International Journal of Digital Curation*, v. 3, n. 1, 2008. Disponível em: <http://www.ijdc.net/article/view/63>. Acesso em: 16 abr. 2021.

MONTMARTIN, P. *La préservation numérique à BAnQ*. 2018. Disponível em: http://www.carl-abrc.ca/wp-content/uploads/2018/12/Montmartin_atscale_AtRN2_FR.pdf. Acesso em: 30 nov. 2021.

MOZILLA.ORG. *A história dos navegadores web*. 2021. Disponível em: <https://www.mozilla.org/pt-BR/firefox/browsers/browser-history/>. Acesso em: 16 abr. 2021.

MURRAY, K. R.; HSIEH, I. K. Archiving web-published materials: a needs assessment of librarians, researchers, and content providers. *CiterSeerx*, 2008. Disponível em: <http://citerseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.461.1774&rep=rep1&type=pdf>. Acesso em: 29 nov. 2020.

MURRAY, K.; HSIEH, I. *Web at risk: preserving our nation's cultural heritage*. [S. l: s. n.], May 31, 2006. Disponível em: https://digitalpreservation.gov/partners/documents/webatrisk_cpg_final_31may2006.pdf. Acesso em: 31 maio 2021.

NATIONAL DIGITAL STEWARDSHIP ALLIANCE. *Glossary*. Alexandria, VA, 2021. Disponível em: <https://ndsa.org/glossary/>. Acesso em: 31 maio 2021.

NATIONAL DIGITAL INFORMATION INFRASTRUCTURE AND PRESERVATION PROGRAM. *Preserving our digital heritage: The National Digital Information Infrastructure and Preservation Program 2010 Report: a collaborative initiative of the Library of Congress*. [S. l.: s. n.], 2011. Disponível em: https://www.digitalpreservation.gov/multimedia/documents/NDIIPP2010Report_Post.pdf. Acesso em: 03 dez. 2021.

NATIONAL LIBRARY OF AUSTRALIA. *Guidelines for the preservation of digital heritage*. [S. l.]: UNESCO, 2003. Disponível em: <https://unesdoc.unesco.org/ark:/48223/pf0000130071>. Acesso em: 25 maio 2021.

NATIONAL LIBRARY OF FINLAND. *Preservation policy*. 21 set. 2009. Disponível em: https://www.doria.fi/bitstream/handle/10024/94315/NLF_Preservation_Policy.pdf?sequence=2. Acesso em: 10 jan. 2022.

NATIONAL LIBRARY OF FINLAND. *Duties and strategy*. 3 mar. 2020. Disponível em: <https://www.kansalliskirjasto.fi/en/about-us/duties-and-strategy>. Acesso em: 25 jan. 2022.

NATIONAL LIBRARY OF FINLAND. *Finnish web archive*. [2021b]. Disponível em: <https://verkkoarkisto.kansalliskirjasto.fi/va/>. Acesso em: 29 set. 2021.

NATIONAL LIBRARY OF FINLAND. *Legal deposit office*. [2021a]. Disponível em: <https://www.kansalliskirjasto.fi/en/legal-deposit-office#online-material>. Acesso em: 29 set. 2021.

NATIONAL LIBRARY OF NORWAY. *Nettarkivering*. [c2021a]. Disponível em: <https://www.nb.no/samlingen/nettarkivet/nettarkivering/>.

NATIONAL LIBRARY OF NORWAY. *Digitizing at the National Library*. [c2021b]. Disponível em: <https://www.nb.no/en/digitizing-at-the-national-library/>. Acesso em: 16 nov. 2021.

NATIONAL LIBRARY OF NORWAY. *Strategy*. [c2021c]. Disponível em: <https://www.nb.no/en/strategy/>. Acesso em: 16 nov. 2021.

NATIONAL AND UNIVERSITY LIBRARY OF ICELAND. *Legal Deposits*. [c2021a]. Disponível em: <https://landsbokasafn.is/index.php?page=legal-deposits>. Acesso em: 16 nov. 2021.

NATIONAL AND UNIVERSITY LIBRARY OF ICELAND. *The Icelandic Web Archive*. [c2021b]. Disponível em: <https://vefsafn.is/>. Acesso em: 16 nov. 2021.

NETPRESERVE.ORG. *Members*. Disponível em: <https://web.archive.org/web/20040603043437/http://netpreserve.org/about/members.php>. Acesso em: 16 mar. 2021.

NETPRESERVE.ORG. *Members list*. Disponível em: <https://web.archive.org/web/20120722084143/http://www.netpreserve.org/about/memberList.php>. Acesso em: 16 mar. 2021.

NETPRESERVE.ORG. *About IIPC members*. Disponível em: <https://web.archive.org/web/20131113223031/http://netpreserve.org/about-us/members>. Acesso em: 16 mar. 2021.

NIU, J. An overview of web archiving. *D-Lib Magazine*, v. 18, n. 3/4, Mar./Apr. 2012. Disponível em: <http://www.dlib.org/dlib/march12/niu/03niu1.html>. Acesso em: 25 maio 2021.

NORONHA, D. P.; FERREIRA, S. M. S. P. Revisões de literatura. In: CAMPELLO, B. S.; CENDÓN, B. V.; KREMER, J. M. (org.). *Fontes de informação para pesquisadores e profissionais*. Belo Horizonte: UFMG, 2000. Cap. 14, p. 191-198.

NORUEGA. Ministério da Cultura. [*Obrigação de entregar documentos publicamente disponíveis (o dever de entregar a lei)*]. LOV-2015-06-19-72 de 01.01.2016. Disponível em: <https://lovdata.no/dokument/NL/lov/1989-06-09-32>. Acesso em: 16 nov. 2021.

NORUEGA. Ministério da Cultura. [*Regulamentos sobre a obrigação de entregar documentos publicamente disponíveis*]. 01.07.2018. Disponível em: <https://lovdata.no/dokument/SF/forskrift/2018-07-01-1139>. Acesso em: 16 nov. 2021.

NUNES, L. *Arquivamento da web: aspectos éticos e legais no acesso e uso da informação*. 2021.162 f. Dissertação (Mestrado em Comunicação e Informação) – Faculdade de Biblioteconomia e Comunicação, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2021. Disponível em <http://hdl.handle.net/10183/225481>. Acesso em: 22 fev. 2022.

ONLINE COMPUTER LIBRARY CENTER – OCLC. *About*. [c2021]. Disponível em: <https://www.oclc.org/en/about.html>. Acesso em: 01 jun. 2021.

ORGANIZACIÓN DE ESTADOS IBEROAMERICANOS PARA LA EDUCACIÓN, LA CIENCIA Y LA CULTURA – OEI. Sobre la OEI. [c2022]. Disponível em: <https://oei.int/quienes-somos/oei>. Acesso em: 16 mar. 2021.

OWENS, T. *The theory and craft of digital preservation*. Baltimore: Johns Hopkins University Press, 2018. 226 p.

PADICAT. *Misión y objetivos*. [c2011a]. Disponível em: <https://www.padicat.cat/es/conocenos/que-es-padicat/mision-y-objetivos>. Acesso em: 14 dez. 2021.

PADICAT. *Preservación del patrimonio digital*. [c2011b]. Disponível em: <https://www.padicat.cat/es/conocenos/que-es-padicat/preservacion-del-patrimonio-digital>. Acesso em: 14 dez. 2021.

PAES, M. L. *Arquivo: teoria e prática*. Rio de Janeiro: Ed. FGV, 2004. 228 p.

PANDORA. *Australian web archive*. [2022]. Disponível em: <http://pandora.nla.gov.au/>. Acesso em: 15 abr. 2021.

PENNOCK, M. *Web archiving*. DPC technology watch report 13-01 March 2013. Great Britain: DCP, 2013. DOI: <http://dx.doi.org/10.7207/twr13-01>. Disponível em: <https://www.dpconline.org/docs/technology-watch-reports/865-dpctw13-01-pdf/file>. Acesso em: 9 jun. 2021.

PINTO, M. R. S. *Preservação de publicações eletrônicas: a questão do depósito legal*. 2011. 131 f. Dissertação (Mestrado em Memória Social) – Centro de Ciências Humanas e Sociais, Universidade Federal do Estado do Rio de Janeiro, Rio de Janeiro, 2011. Disponível em:

<http://www.memoriasocial.pro.br/documentos/Disserta%C3%A7%C3%B5es/Diss275.pdf>. Acesso em: 9 jun. 2021.

PÉREZ, K.; SERRA, E. *Repositori de preservació digital de la Biblioteca de Catalunya*: informe descriptiu i de situació. Catalunya: Biblioteca de Catalunya, 2010. Disponível em: <http://hdl.handle.net/2072/97251>. Acesso em: 31 jan. 2022.

PORTUGAL. Ministério da Educação e Ciência. Decreto lei 55, de 17 de abril de 2013. *Diário da República*, 1ª série, n. 75. Disponível em: https://www.fct.pt/documentos/Lei_Organica_2013.pdf. Acesso em: 10 dez. 2021.

RAID. In: WIKIPEDIA: a enciclopédia livre. [São Francisco, CA: Wikimedia Foundation, 2021]. Disponível em: <https://en.wikipedia.org/wiki/RAID>. Acesso em: 14 dez. 2021.

RAMOS JÚNIOR, M. A. C.; CIANCONI, R. B. O uso de ontologias no registro de lições aprendidas em projetos gerenciados com scrum. *Perspectivas em Gestão & Conhecimento*, João Pessoa, v. 9, n. 1, p. 82-100, 2019. DOI: 10.21714/2236-417X2019v9n1.

REDE CARINIANA. *Glossário*. Brasília: IBICT, [2021]. Disponível em: <https://cariniana.ibict.br/index.php/referencias/2016-02-19-18-56-47>. Acesso em: 31 maio 2021.

REDE CARINIANA. World Digital Preservation Day 2020. *Cariniana*, 22 out. 2020. Disponível em: <https://cariniana.ibict.br/index.php/noticias/414-world-digital-preservation-day-2020>. Acesso em: 01 jun. 2021.

REDE CARINIANA. Webinar arquivamento e preservação da web. 2 fev. 2021. [S. l.: s. n.], 2021. 1 vídeo (125 min). Publicado pelo canal Rede Cariniana. Disponível em: <https://youtu.be/QPZuLwWR9y8>. Acesso em: 01 jun. 2021.

REDE CARINIANA. V SINPRED. [S. l.: s. n.], 2021. Publicado pelo canal Rede Cariniana. Disponível em: <https://www.youtube.com/channel/UC7pJ4CrMcF7WVYsWjic-nsg>. Acesso em: 8 jun. 2021.

RESAW. A research infrastructure for the study of archived web materials. [2021]. Disponível em: <https://resaw.eu/>. Acesso em: 01 jun. 2021.

RESEARCH DATA ALLIANCE. *The TRUST Principles*: an RDA community effort. 18 maio 2020. Disponível em: <https://www.rd-alliance.org/trust-principles-rda-community-effort>. Acesso em: 11 abr. 2022.

LIBRARIES GROUP. *Trusted digital repositories*: attributes and responsibilities an RLG-OCLC Report. Mountain View, CA: RLG, 2002. Disponível em: <https://www.oclc.org/content/dam/research/activities/trustedrep/repositories.pdf>. Acesso em: 03 set. 2021.

REYES AYALA, B. *Web archiving bibliography*. Texas: UNT Digital Libraries, 2013. Version 1.0. Disponível em: <https://digital.library.unt.edu/ark:/67531/metadc172362/>. Acesso em: 10 fev. 2021.

REZENDE, L. V. R.; MARTINS, D. L. Iniciativas científicas de arquivamento e preservação de conteúdos em mídias sociais: panorama atual. *Revista Ibero-Americana de Ciência da Informação*, Brasília, DF, v. 11, n. 1, p. 219-236, jan./abr. 2018. DOI: 10.26512/rici.v11.n1.2018.8538.

RIBEIRO, L. M. *A história da internet*. [S. l.: s. n.], 1998. Disponível em: <https://paginas.fe.up.pt/~mgi97018/historia.html>. Acesso em: 07 mar. 2021.

RILEY, J. *Understanding metadata what is metadata, and what is it for?* Baltimore: NISO, 2017. ISBN: 978-1-937522-72-8. Disponível em: https://groups.niso.org/apps/group_public/download/17446/Understanding%20Metadata.pdf. Acesso em: 01 jun. 2021.

ROCKEMBACH, M. Arquivamento da web: estudos de caso internacionais e o caso brasileiro. *RDBCI: Revista Digital de Biblioteconomia e Ciência da Informação*, Campinas, v. 16, n. 1, p. 07-24 jan./abr. 2018. DOI 10.20396/rdbci.v16i1.8648747. Disponível em: <https://periodicos.sbu.unicamp.br/ojs/index.php/rdbci/article/view/8648747>. Acesso em: 29 nov. 2020.

ROCKEMBACH, M. Arquivamento da web no contexto das humanidades digitais: da produção a preservação digital. *Liinc em Revista*, Rio de Janeiro, v.15, n.1, p. 131-139, maio 2019. DOI: <https://doi.org/10.18617/liinc.v15i1.4578>. Disponível em: <http://revista.ibict.br/liinc/article/view/4578/4142>. Acesso em: 29 nov. 2020.

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. e5713, 2021a. DOI: 10.18617/liinc.v17i1.5713.

ROCKEMBACH, M. Brazilian elections web archive. In: GENERAL ASSEMBLY & WEB ARCHIVING CONFERENCE, 14-16 June, 2021, Luxemburgo. *Abstracts [...]*. Luxemburgo: The National Library of Luxembourg: International Internet Preservation Consortium, 2021b. Disponível em: <https://netpreserve.org/ga2021/wac/abstracts/#208>. Acesso em: 17 jun. 2021.

ROCKEMBACH, M.; PAVÃO, C. Políticas e tecnologias de preservação digital no arquivamento da web. *Revista Ibero-Americana de Ciência da Informação*, Brasília, DF, v. 11, n. 1, p. 168-182, jan./abr. 2018. DOI: <https://doi.org/10.26512/rici.v11.n1.2018.8473>.

RONDINELLI, R. C. *O documento arquivístico ante a realidade digital: uma revisitação conceitual necessária*. Rio de Janeiro: Ed. FGV, 2013. 280 p.

ROYAL DANISH LIBRARY. *Legal deposit*. [2021a]. Disponível em: <https://www.kb.dk/en/about-us/legal-deposit>. Acesso em: 24 set. 2021.

ROYAL DANISH LIBRARY. *Netarkivet*. [2021b]. Disponível em: <https://www.kb.dk/find-materiale/samlinger/netarkivet>. Acesso em: 24 set. 2021.

ROYAL DANISH LIBRARY. *Research access to Netarkivet*. [2021c]. Disponível em: <https://www.kb.dk/en/find-materials/collections/netarkivet/research-access>. Acesso em: 24 set. 2021.

ROYAL DANISH LIBRARY. *Digital preservation policy*. 9 jan. 2019. Disponível em: http://id.kb.dk/documents/DB_Politik_20190109_UK.pdf. Acesso em: 24 set. 2021.

SALES, L. F. *Integração semântica de publicações científicas e dados de pesquisa: proposta de modelo de publicação ampliada para a área de Ciências Nucleares*. 2014. 266 f. Tese (Doutorado em Ciência da Informação) – Escola de Comunicação, Universidade Federal do Rio de Janeiro, IBICT, Rio de Janeiro, 2014. Disponível em: <http://repositorio.ibict.br/bitstream/123456789/874/1/LUANA%20SALES%20D.pdf>. Acesso em: 31 maio 2021.

SAMOUELIAN, M.; DOOLEY, J. *Descriptive metadata for web archiving: review of harvesting tools*. Dublin, OH: OCLC Research, 2018. DOI: <https://doi.org/10.25333/C37H0T>.

SANTOS, G. C. Ensaio sobre arquivamento de páginas web: foco na experiência do Portal de Periódicos da UNICAMP, utilizando o Conifer (Rhizome). In: SEMINÁRIO INTERNACIONAL DE PRESERVAÇÃO DIGITAL, 5., 2021, Campinas. *Resumos [...]*. Campinas: UNICAMP: IBICT, 2021. Disponível em: <http://eventoscariniana.ibict.br/index.php/sinpred/issue/view/5>. Acesso em: 31 maio 2021.

SANTOS, G. C. et al. *Pesquisa bibliográfica mediatizada pelas novas tecnologias*. In: CURSO DE ESPECIALIZAÇÃO EM GESTÃO EDUCACIONAL, Módulo 8 – Tecnologia de Informação e Comunicação, 2005. Campinas: UNICAMP, 2005. Disponível em: <http://www.bibliotecadigital.unicamp.br/document/?down=53409>. Acesso em: 17 jun. 2021.

SANTOS, V. B. Arquivamento web: legislação correlata. *Revista Brasileira de Preservação Digital*, Campinas, v. 1, p. 1-11, 2020. Disponível em: <https://econtents.bc.unicamp.br/inpec/index.php/rebpred/article/view/14800>. Acesso em: 10 fev. 2021.

SANTOS, H. M.; FLORES, D. Nuevos rumbos de la preservación digital: de las estrategias a los sistemas informatizados. *Biblios: Journal of Librarianship and Information Science*, Tacna, Peru, n. 70, p. 31-43, feb. 2018. doi:<https://doi.org/10.5195/BIBLIOS.2017>. Disponível: <http://biblios.pitt.edu/ojs/index.php/biblios/article/view/326>. Acesso em: 08 abr. 2021.

SÁ-SILVA, J. R.; ALMEIDA, C. D.; GUINDANI, J. F. Pesquisa documental: pistas teóricas e metodológicas. *Revista Brasileira de História & Ciências Sociais*, São Leopoldo, v. 1, n.1, p. 1-15, jul. 2009. Disponível em: <https://periodicos.furg.br/rbhcs/article/view/10351>. Acesso em: 29 nov. 2020.

SCHÄFER, M. B.; FLORES, D. Preservação da informação arquivística digital: repercussões para o patrimônio cultural. *Em Questão*, Porto Alegre, v. 19, n.1, p. 173-186, jan./jun. 2013. Disponível em: <https://seer.ufrgs.br/EmQuestao/article/view/27024/31549>. Acesso em: 15 maio 2021.

SCHISLER, M. The theory and craft of digital preservation: comentários sobre os 16 axiomas de preservação digital. *Revista Eletrônica de Comunicação, Informação e Inovação em Saúde*, v. 14, n. 3, 2020. DOI: <https://doi.org/10.29397/reciis.v14i3.2155>. Disponível em: <https://www.reciis.icict.fiocruz.br/index.php/reciis/article/view/2155>. Acesso em: 10 fev. 2021.

SAYÃO, L. F. Uma outra face dos metadados: informações para a gestão da preservação digital. *Encontros Bibli*, Florianópolis, v. 15, n. 30, p. 1-31, 2010. DOI: 10.5007/1518-2924.2010v15n30p1.

SILVA, W. *Softwares para repositórios digitais: avaliação para la gestión y preservación de los materiales bibliográficos y de los documentos archivísticos de la Biblioteca Nacional de Brasil*. 2019. 82 f. Dissertação (Máster em Sistemas de Informação Digital) – Universidade de Salamanca, Salamanca, Espanha, 2019. Disponível em: <https://www.researchgate.net/publication/343123807>. Acesso em: 18 abr. 2021.

SIMPÓSIO POLÍTICAS E ESTRATÉGIAS DE PRESERVAÇÃO DE CONTEÚDO NA WEB, 1., 2020. [S. l.]: FEBAB, 2020. 1 vídeo (240 min) Publicado no canal FEBAB. Disponível em: <https://www.youtube.com/watch?v=X6SmBmbD7fg>. Disponível em: 01 jun. 2021.

SIMPÓSIO POLÍTICAS E ESTRATÉGIAS DE PRESERVAÇÃO DE CONTEÚDO NA WEB, TOI 2021 ONLINE, 2., 15 dez. 2021. 1 vídeo (160 min). Publicado pelo canal TOI 2021 ONLINE. Disponível em: <https://www.youtube.com/watch?v=3KhAu7OczSA>. Acesso em: 23 dez. 2021.

SLAVIN, T. Web browser history. *Beanz Magazine*, Woodmere, NY, Aug. 2016. History. Issue: Back to School. Disponível em: <https://www.kidscodecs.com/web-browser-history/>. Acesso em: 16 abr. 2021.

SOCIETY OF AMERICAN ARCHIVISTS – SAA. *About SAA*. Chicago, IL, [c2021a]. Disponível em: <https://www2.archivists.org/aboutsaa>. Acesso em: 31 maio 2021.

SOCIETY OF AMERICAN ARCHIVISTS – SAA. Digital preservation. *SAA Dictionary of Terminology Archivists*, Chicago, IL, [c2021b]. Disponível em: <https://dictionary.archivists.org/entry/digital-preservation.html>. Acesso em: 31 maio 2021.

SOCIETY OF AMERICAN ARCHIVISTS – SAA. Web page. *SAA Dictionary of Terminology Archivists*, Chicago, IL, [c2021c]. Disponível em: <https://dictionary.archivists.org/entry/web-page.html>. Acesso em: 25 ago. 2021.

SOCIETY OF AMERICAN ARCHIVISTS – SAA. *Web archiving section*. Chicago, IL, [c2021d]. Disponível em: <https://www2.archivists.org/groups/web-archiving-section/web-archiving-section>. Acesso em: 01 jun. 2021.

SOCIETY OF AMERICAN ARCHIVISTS – SAA. Website. *SAA Dictionary of Terminology Archivists*, Chicago, IL, [c2021e]. Disponível em: <https://dictionary.archivists.org/entry/website.html>. Acesso em: 14 mar. 2022.

SØRENSEN, M.; HAVE, U. *NetarchiveSuite*. Jan. 2020. Disponível em: <https://sbforge.org/display/NAS/NetarchiveSuite>. Acesso em: 24 set. 2021.

SOUZA, I. Entenda de uma vez o que é Github e a importância dele num negócio. *Rock Content*, 11 fev. 2020. Disponível em: <https://rockcontent.com/br/blog/o-que-e-github/>. Acesso em: 15 fev. 2022.

SUMMERS, E. Appraisal practices in web archives. *SocArXiv*, College Park, MD, 15 Mar. 2018. DOI: <https://doi.org/10.31235/osf.io/75mjp>. Disponível em: <https://osf.io/preprints/socarxiv/75mjp/>. Acesso em: 29 nov. 2020.

TERRADA, G. A. F.; FONSECA, V. M. M. Perfil e lugar dos arquivos privados institucionais em entidades custodiadoras cariocas. *Informação & Informação*, Londrina, v. 24, n. 1, p. 383-402, mar. 2019. DOI: <http://dx.doi.org/10.5433/1981-8920.2019v24n1p383>.

THE BRITISH LIBRARY. About legal deposit. Reino Unido, [c2021a]. Disponível em: <https://www.bl.uk/legal-deposit/about-legal-deposit>. Acesso em: 20 set. 2021.

THE BRITISH LIBRARY. Legal deposit and web archiving. Reino Unido, [c2021b]. Disponível em: <https://www.bl.uk/legal-deposit/web-archiving>. Acesso em: 20 set. 2021.

THE BRITISH LIBRARY. UK Web Archive. Reino Unido. Reino Unido, [c2021c]. Disponível em: <https://www.bl.uk/collection-guides/uk-web-archive#>. Acesso em: 20 set. 2021.

THE BRITISH LIBRARY. Digital preservation. Reino Unido, [c2021d]. Disponível em: <https://www.bl.uk/digital-preservation>. Acesso em: 20 set. 2021

THE BRITISH LIBRARY. *Joint collecting framework for UK legal deposit, 2015-2020*. Reino Unido, 2015. Disponível em: <https://www.bl.uk/help/how-to-deposit-your-digital-publications>. Acesso em: 20 set. 2021.

THE BRITISH LIBRARY. The Preservation Advisory Centre: Building a preservation policy. Revisado 2013. Disponível em: <https://www.bl.uk/britishlibrary/~media/bl/global/conservation/pdf-guides/building-a-preservation-policy.pdf>. Acesso em: 07 jan. 2022.

THE BRITISH LIBRARY. *Sustaining the value: The British Library digital preservation strategy 2017-2020*. 2017. Disponível em: <https://www.bl.uk/digital-preservation/british-library/~media/c0068e4bb25c4ac39d1f4c5a020a21c8.ashx>. Acesso em: 7 jan. 2022.

THE NATIONAL ARCHIVES. Operational selection policy OPS27: UK central government web estate. *The National Archives*, Reino Unido, 2014. Disponível em: <https://cdn.nationalarchives.gov.uk/documents/information-management/osp27.pdf>. Acesso em: 20 set. 2021.

THE NATIONAL ARCHIVES. Records collection policy. *The National Archives*, Reino Unido, 2012. Disponível em: <https://cdn.nationalarchives.gov.uk/documents/records-collection-policy-2012.pdf>. Acesso em: 19 nov. 2021.

THE NATIONAL ARCHIVES. The UK government web archive: guidance for digital and records management teams. *The National Archives*, Reino Unido, 2017. Disponível em: <https://cdn.nationalarchives.gov.uk/documents/web-archiving-technical-guidance.pdf>. Acesso em: 20 set. 2021.

THIBODEAU, K. Overview of technological approaches to digital preservation and challenges in coming years. In: COUNCIL ON LIBRARY AND INFORMATION RESOURCES – CLIR. *The state of digital preservation: an international perspective. Proceedings [...]*. Washington, DC: CLIR, 2002. Disponível em: <https://www.clir.org/pubs/reports/pub107/thibodeau/>. Acesso em 18 abr. 2021.

THOMAS, A. *et al. Researcher engagement with web archives: challenges and opportunities for investment*. London: JISC: Oxford Internet Institute, 2010. Disponível em: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1715000. Acesso em: 28 fev. 2021.

TROVE. *What it is Trove*. [2009?]. Disponível em: <https://trove.nla.gov.au/about/what-trove>. Acesso em: 20 mar. 2021.

UNITED KINGDOM. *Public Records Act 1958*. [c2021a]. Disponível em: <https://www.legislation.gov.uk/ukpga/Eliz2/6-7/51/contents>. Acesso em: 19 nov. 2021.

UNITED KINGDOM. *Copyright, designs and patents act 1988*. [c2021b]. Disponível em: <https://www.legislation.gov.uk/ukpga/1988/48/section/163>. Acesso em: 19 nov. 2021.

UNITED KINGDOM. *The Legal Deposit Libraries (Non-Print Works) Regulations 2013*. Disponível em: <https://www.legislation.gov.uk/uksi/2013/777/contents/made>. Acesso em: 19 nov. 2021.

UK PARLIAMENT WEB ARCHIVE. *About*. [c2021]. Disponível em: <http://webarchive.parliament.uk/about>. Acesso em: 20 set. 2021.

UK WEB ARCHIVE. *Technical information*. [c2021a]. Disponível em: <https://www.webarchive.org.uk/en/ukwa/info/technical>. Acesso em: 20 set. 2021.

UK WEB ARCHIVE. *Topics and Themes*. [c2021b]. Disponível em: <https://www.webarchive.org.uk/en/ukwa/collection>. Acesso em: 20 set. 2021.

UK WEB ARCHIVE. *Web archives and digital preservation*. [c2021c]. Disponível em: <https://www.webarchive.org.uk/en/ukwa/collection/527>. Acesso em: 20 set. 2021.

UNESCO. *Convenção para a protecção do património mundial, cultural e natural*. Paris, 16 nov. 1972. Disponível em: <https://whc.unesco.org/archive/convention-pt.pdf>. Acesso em: 31 maio 2021.

UNESCO. Carta sobre la preservación del patrimonio digital. In: UNESCO. *Records of the General Conference*, 32nd sesión, 29 september to 17 october 2003. Paris: UNESCO, [c2004]. p. 74-77. v. 1. [Paris], Disponível em: <https://unesdoc.unesco.org/ark:/48223/pf0000133171.page=80>. Acesso em: 08 jun. 2021. Acesso em: 8 jun. 2021.

UNESCO. *Glossary of world heritage terms*. [Paris]: UNESCO, 1997. Disponível em: <https://unesdoc.unesco.org/ark:/48223/pf0000112889/PDF/112889eng.pdf.multi>. Acesso em: 31 maio 2021.

UNESCO. *Concept of digital preservation*. [c2021a]. Disponível em: <https://en.unesco.org/themes/information-preservation/digital-heritage/concept-digital-preservation#:~:text=Digital%20preservation%20consists%20of%20the,hardware%20tools%20acting%20on%20data>. Acesso em: 01 jun. 2021.

UNESCO. *Noción de patrimonio digital*. [c2021b]. Disponível em: <https://es.unesco.org/themes/information-preservation/digital-heritage/concept-digital-heritage>. Acesso em: 31 maio 2021.

UNESCO. *Proyecto de recomendación relativa a la preservación del patrimonio documental, comprendido el patrimonio digital, y el acceso al mismo*. Paris: UNESCO, 2015. Disponível em: https://unesdoc.unesco.org/ark:/48223/pf0000233916_spa?locale=en. Acesso em: 31 maio 2021.

UNESCO. *Directrices UNESCO/PERSIST sobre selección del patrimonio digital para su conservación a largo plazo*. [S. l.]: UNESCO, 2016a. Disponível em: https://www.ifla.org/files/assets/hq/topics/cultural-heritage/documents/persist-content-guidelines_es.pdf. Acesso em: 31 maio 2021.

UNESCO. *Recommendation concerning the preservation of, and access to, documentary heritage including in digital form*. Paris: UNESCO, 2016b. Disponível em: <https://unesdoc.unesco.org/ark:/48223/pf0000244675>. Acesso em: 27 jan. 2022.

UNESCO. *UNESCO PERSIST Programme*. [c2021c]. Disponível em: <https://unesco.persist.org/>. Acesso em: 31 maio 2021.

UNESCO. *Memoria del mundo: directrices para la salvaguardia del patrimonio documental*. Paris: UNESCO, 2002. 71p. Disponível em: http://www.unesco.org/new/fileadmin/MULTIMEDIA/FIELD/Quito/pdf/Directrices_Salvaguarda_Patrimonio_Documental.pdf. Acesso em: 25 maio 2021.

UNESCO. *Declaração UNESCO/UBC Vancouver: a memória do mundo na era digital: digitalização e preservação*. 26 a 28 de setembro de 2012 Vancouver, Colúmbia Britânica, Canadá. 2012. Disponível em: http://wayback.archive-it.org/10611/20170212200056/http://www.unesco.org/new/fileadmin/MULTIMEDIA/HQ/CI/CI/images/mow/unesco_abc_vancouver_declaration_pt.pdf. Acesso em: 25 maio 2021.

UNESCO. *The memory of the world in the digital age: digitization and preservation*. 26-28 September 2012, Vancouver, British Columbia, Canada. 2012. Disponível em: <http://wayback.archive-it.org/10611/20170204061808/http://www.unesco.org/new/en/communication-and-information/events/calendar-of-events/events-websites/the-memory-of-the-world-in-the-digital-age-digitization-and-preservation/>. Acesso em: 25 maio 2021.

UNESCO. *Memory of the world*. [c2017]. Disponível em: <http://www.unesco.org/new/en/communication-and-information/memory-of-the-world/register/access-by-region-and-country/br/>. Acesso em: 1 jun. 2021.

UNESCO TV. *The memory of the world in the digital age: digitization and preservation*. 2012. Disponível em: <https://www.unesco.org/archives/multimedia/document-2460>. Acesso em: 01 jun. 2021.

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL. *Núcleo de pesquisa em arquivamento da web e preservação digital*. Porto Alegre: UFRGS, [2020]. Disponível em: <https://www.ufrgs.br/nuaweb/>. Acesso em: 29 nov. 2020.

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL. *Núcleo de pesquisa da Fabico participa de evento internacional de preservação da web*. Porto Alegre, UFRGS, [c2021]. Disponível em: <http://www.ufrgs.br/fabico/nucleo-de-pesquisa-da-fabico-participa-de-evento->

internacional-de-preservacao-da-web. Acesso em: 17 jun. 2021.

U.S. COPYRIGHT OFFICE. *Mandatory deposit of copies or phonorecords for the Library of Congress*. In: LIBRARY OF CONGRESS. Circular 7d Revised: 11/2021. Disponível em: <https://www.copyright.gov/circs/circ07d.pdf>. Acesso em: 01 dez. 2021.

VAILLANT CALLOL, M. *Biodeterioração do patrimônio histórico documental: alternativas para eliminação e controle*. Rio de Janeiro: MAST: FCRB, 2013. 139 p. Disponível em: <https://www.gov.br/turismo/pt-br/entidades-vinculadas/casa-de-rui-barbosa/atuacao/museu-e-jardim/acervos-preservacao>. Acesso em: 31 de maio de 2021.

VERMA, P. Meet the 1,300 librarians racing to back up Ukraine's digital archives. *The Washington Post*, 8 Apr. 2022. Disponível em: <https://www.washingtonpost.com/technology/2022/04/08/ukraine-digital-history/>. Acesso em: 12 abr. 2022.

VIOLAWWW. In: WIKIPEDIA: a enciclopédia livre. [São Francisco, CA: Wikimedia Foundation, 2021]. Disponível em: <https://en.wikipedia.org/wiki/ViolaWWW>. Acesso em: 18 mar. 2021.

VLASSENROOT, E. *et al.* Web archives as a data resource for digital scholars. *International Journal of Digital Humanities*, London, v. 1, p. 85-111, 2019. DOI: <https://doi.org/10.1007/s42803-019-00007-7>. Disponível em: <https://link.springer.com/article/10.1007/s42803-019-00007-7>. Acesso em: 02 fev. 2021.

WEBBER, J. About this blog. *UK Web archive blog*, Reino Unido, [c2021]. Disponível em: https://blogs.bl.uk/webarchive/about-this-blog.html?_ga=2.112266817.492970548.1631918187-639698563.1631918187. Acesso em: 20 set. 2021.

WEBSTER, P. Users, technologies, organisations: towards a cultural history of world web Archiving. In: BRÜGGER, N. *Web 25: histories from the first 25 years of the world wide web*. New York: Peter Lang, 2017, 257 p. ISBN 978-1-4331-3270-4. Disponível em: <https://hcommons.org/deposits/item/hc:26187/>. Acesso em: 27 set. 2021.

WORLD WIDE WEB CONSORTIUM. *W3C Escritório Brasil*. c2021. Disponível em: <http://www.w3c.br/Sobre/>. Acesso em: 10 fev. 2021.

WORLD WIDE WEB CONSORTIUM – W3C. *W3C Escritório Brasil*. c2021. Disponível em: <http://www.w3c.br/Sobre/>. Acesso em: 10 fev. 2021.

WORLD WIDE WEB FOUNDATION – W3C. *History of the web*. [S. l.], c2021. Disponível em: <https://webfoundation.org/about/vision/history-of-the-web/>. Acesso em: 07 jun. 2021.

XIE, Z. *et al.* Archiving the relaxed consistency web. In: ASSOCIATION OF COMPUTING MACHINERY INTERNATIONAL CONFERENCE ON INFORMATION & KNOWLEDGE MANAGEMENT, 22., 2013, San Francisco, CA. *Proceedings [...]*. San Francisco: ACM, 2013. p. 2119-2128. Disponível em: <https://dl.acm.org/citation.cfm?id=2505551>. Acesso em: 02 fev. 2021.

YAMAOKA, E. J.; GAUTHIER, F. O. Objetos digitais: em busca da precisão conceitual. *Informação & Informação*, Londrina, v. 18, n. 2, p. 77-97, ago. 2013. DOI: <http://dx.doi.org/10.5433/1981-8920.2013v18n2p77>. Disponível em: <http://www.uel.br/>

revistas/uel/index.php/informacao/article/view/16162. Acesso em: 19 abr. 2021.
 YEO, G. Concepts of record (2): prototypes and boundary objects. *The American Archivist*, v. 71, n. 1, p. 118-143, Apr. 2008. DOI: <https://doi.org/10.17723/aarc.71.1.p0675v40tr14q6w2>. Disponível em: <https://meridian.allenpress.com/american-archivist/article/71/1/118/24085/Concepts-of-Record-2-Prototypes-and-Boundary>. Acesso em: 1 jun. 2021.

OBRAS CONSULTADAS

BRAGA, K. S. Aspectos relevantes para seleção de metodologia adequada à pesquisa social em Ciência da Informação. In: MUELLER, S. P. M. (org.). *Métodos para a pesquisa em Ciência da Informação*. Brasília: Thesaurus, 2007, p. 17-38.

BUCKLAND, M. K. Information as thing. *Journal of the American Society for Information Science*, New York, v. 42, n. 5, p. 351-360, June 1991. DOI: [https://doi.org/10.1002/\(SICI\)1097-4571\(199106\)42:5%3C351::AID-ASI5%3E3.0.CO;2-3](https://doi.org/10.1002/(SICI)1097-4571(199106)42:5%3C351::AID-ASI5%3E3.0.CO;2-3).

CASTELLS, M.; CARDOSO, G. (org.). *A sociedade em rede: do conhecimento à acção política*. [S. l.]: Imprensa Nacional, 2005. 439 p.

CERF, V. G.; KAHN, R. E. A protocol for packet network intercommunication. *IEEE Transactions on Communications*, v. 22, n. 5, May 1974. Disponível em: <https://www.cs.princeton.edu/courses/archive/fall06/cos561/papers/cerf74.pdf>. Acesso em: 07 mar. 2021.

CONSELHO NACIONAL DE JUSTIÇA. Encontro Nacional de Memória do Poder Judiciário, 1., 18 maio 2021. [Palestras]. 1 vídeo (160 min). Publicado pelo canal do Conselho Nacional de Justiça. Disponível em: <https://www.youtube.com/watch?v=RkR2RjaSjC0>. Acesso em: 01 jun. 2021.

CONSEIL EUROPÉEN POUR LA RECHERCHE NUCLÉAIRE. *Home of the first website*. Disponível em: <http://info.cern.ch>. Acesso em: 10 fev. 2021.

GERHARDT, T. E.; SILVEIRA, D. T. (org.). *Métodos de pesquisa*. Porto Alegre: UFRGS, 2009. 120 p.

INTERNET SOCIETY. *Brief history of internet*, 1997. Disponível em: <https://www.internet-society.org/internet/history-internet/brief-history-internet/>. Acesso em: 07 mar. 2021.

INTERPARES TRUST. Disponível em: <https://interparestrust.org/>. Acesso em: 01 jun. 2021.
 INTERNATIONAL INTERNET PRESERVATION CONSORTIUM – IIPC. *Training materials*. c2022. Disponível em: <https://netpreserve.org/web-archiving/training-materials/>. Acesso em: 01 abr. 2022.

LAKATOS, E. M.; MARCONI, M. A. *Fundamentos de metodologia científica*. São Paulo: Atlas, 2003. 311 p.

MINAYO, M. C. S. O desafio da pesquisa social. In: MINAYO, M. C. S. *Pesquisa social: teoria, método e criatividade*. 30. ed. Petrópolis: Vozes, 2011. Cap. 1, p. 9-29.

MINAYO, M. C. S. *Pesquisa Social: teoria, método e criatividade*. 21. ed. Petrópolis: Vozes,

2002.

NATIONAL INFORMATION STANDARDS ORGANIZATION. *Understanding Metadata*. Bethesda: NISO, 2004. ISBN: 1-880124-62-9. Disponível em: https://www.lter.uaf.edu/metadata_files/UnderstandingMetadata.pdf. Acesso em: 01 jun. 2021.

SARACEVIC, T. Ciência da informação: origem, evolução e relações. *Perspectivas em Ciência da Informação*, Belo Horizonte, v. 1, n. 1, p. 41-62, jan./jun. 1996. Disponível em: <http://portaldeperiodicos.eci.ufmg.br/index.php/pci/article/view/235/22>. Acesso em: 29 nov. 2020.

VENLET, J. *et al.* *Descriptive metadata for web archiving: literature review of user needs*. Dublin, OH: OCLC Research, 2018. DOI: <https://doi.org/10.25333/C33P7Z>.

APÊNDICE

APÊNDICE A – E-MAIL PARA AS INSTITUIÇÕES IBERO-AMERICANAS DO IIPC

Prezados Senhores do/da _____,

Sou Gabriela Ayres Ferreira Terrada, aluna de mestrado do Programa de Pós-Graduação em Ciência da Informação da Universidade Federal Fluminense – UFF, no Brasil. Estou realizando uma pesquisa de mestrado sobre preservação do patrimônio digital por instituições de guarda da memória das páginas *web*, que fazem parte do International Internet Preservation Consortium – IIPC, sob orientação da Prof. Dra Regina de Barros Cianconi.

Não localizamos no *website* institucional algumas informações e por esse motivo faço contato, solicitando algumas informações.

- 1- A instituição possui documento de política de preservação digital ou documento que aborde a temática (Planejamento Estratégico, Planos, Manuais)? Se a resposta for positiva, poderia enviar cópia do(s) documento(s)?
- 2- Poderia descrever como é realizada a preservação digital dos *websites* recolhidos?
- 3- A instituição possui área exclusiva dedicada à preservação digital? Em caso positivo, qual é? E, em caso negativo, qual área da instituição a realiza?
- 4- Qual o perfil profissional que atua nesta atividade?
- 5- Poderia descrever quais são os problemas e barreiras à preservação da web para a instituição?
- 6- A instituição promove o acesso e a utilização de conteúdos web preservados? Em caso afirmativo, de que forma?

Aguardo sua resposta e desde já agradeço a atenção.

Gabriela Ayres Ferreira Terrada

Bibliotecária e Arquivista

<http://lattes.cnpq.br/5675043405444867>

ANEXO

ANEXO A – RESPOSTAS DAS INSTITUIÇÕES IBERO-AMERICANAS

Membro do IIPC	Pergunta 1 –	Pergunta 2 –	Pergunta 3 –	Pergunta 4 –	Pergunta 5 –	Pergunta 6 –
Arquivo.pt – Portuguese Web Archive	<p>O processo de preservar informação publicada na web é complexo.</p> <p>Para organizar este processo, produzimos e mantemos dezenas de documentos técnicos internos que não são públicos e que também são compreensíveis apenas para especialistas em arquivos da web. No entanto, em https://arquivo.pt/publica poderá encontrar documentos públicos que lhe poderão ser úteis.</p>	<p>No geral, este relatório técnico dá uma ideia do processo:</p> <p>Daniel Gomes, David Cruz, João Miranda, Miguel Costa e Simão Fontes, Creating a searchable web archive, Relatório Técnico, Lisboa, Portugal, Maio de 2012 (bibtex) http://sobre.arquivo.pt/wp-content/uploads/creating-a-searchable-web-archive-relatorio.pdf</p>	<p>A preservação digital é uma actividade transversal a cada um dos serviços que a FCT-FCCN gere (ex. publicações científicas, dados abertos, publicações em acesso aberto, informação técnica de gestão da rede académica) porque cada tipo de dados requer abordagens de preservação digital específicas para serem adequadas ao contexto de cada serviço. O Arquivo.pt dedica-se apenas à preservação digital de informação publicada na web.</p>	<p>Depende dos serviços.</p> <p>No caso do Arquivo.pt pode ver os perfis da equipa em: https://sobre.arquivo.pt/pt/acerca/equipa/</p>	<p>Sugiro que consulte estas referências:</p> <p>The forum on the archiving of the Polish Internet Day 1 Speech: Daniel Gomes Portugal - https://www.youtube.com/watch?v=-sCjHMv5-KY</p> <p>Em particular, a partir de 15:45: https://youtu.be/-sCjHMv5-KY?t=930</p> <p>- The Past Web: Exploring Web Archives https://arquivo.pt/livro</p> <p>Em particular, a "Part VI A Look into the Future"</p>	<p>Prémio, exposições, coleções, memorial https://sobre.arquivo.pt/pt/exemplos-2/colecoes/</p> <p>https://sobre.arquivo.pt/pt/exemplos-2/exemplos/</p> <p>https://sobre.arquivo.pt/pt/colabore/divulgue-o-arquivo-pt-pela-sua-comunidade/</p> <p>https://sobre.arquivo.pt/pt/colabore/premios-arquivo-pt/</p>
<i>Biblioteca Nacional da Espanha</i>	<p>Actualmente no tenemos publicado un documento con la política de preservación digital de la BNE. Estamos trabajando en la elaboración del mismo y lo publicaremos en nuestra web a lo largo del año 2022.</p>	<p>Lo sitios web recolectados generan un tipo de archivo comprimido conocido como WARC. WARC es un tipo de formato normalizado ISO 28500:2017 creado por</p>	<p>La BNE no tiene un área exclusiva dedicada a la preservación digital. Es una sección que depende del Área de Colecciones Digitales, área responsable de la</p>	<p>En la BNE la responsabilidad del proyecto de preservación digital recae principalmente en personal bibliotecario. Sin embargo, el personal</p>	<p>Los archivos WARC generados deben incluirse en el flujo de preservación, al igual que otros tipos de formatos digitales. El formato WARC es un tipo de formato no</p>	<p>En el caso del Archivo de la Web Española el acceso se produce a través: http://www.bne.es/es/Colecciones/ArchivoWeb/</p> <p>La herramienta que permite la visualización</p>

		Internet Archive, totalmente comprobado en su seguridad y fiabilidad, que permite comprimir la información de los HTTP, conservados junto a sus metadatos técnicos generados en el guardado y creación de estos.	gestión de los fondos digitalizados y de todo lo nacido digital.	bibliotecario realiza esta labor con el apoyo del departamento informático, que da soporte técnico al proyecto.	contemplado en los flujos habituales de preservación y necesita un estudio y acomodo específico.	es OpenWayback, debido a que los contenido almacenados tienen derechos de autor vigentes y se encuentran protegidos por la ley de propiedad intelectual española, sólo es posible su visualización dentro de la Biblioteca Nacional de España y los centros de conservación con punto de acceso, los cuales se pueden ver en el siguiente enlace: http://www.bne.es/es/Colcciones/ArchivoWeb/CentrosConservacionCCA/index.html El usuario desde fuera de estas instituciones sólo puede consultar que sitios web y con qué frecuencia se han guardado, pero no consultar los contenidos. Por si quieren más información sobre lo que hacemos, le dejo algunos enlaces al sitio web de la BNE: <ul style="list-style-type: none"> • El depósito legal de las publicaciones en línea • El archivo de la web española • Información técnica • Preguntas frecuentes
--	--	--	--	---	--	--

						sobre el archivo web
<i>Biblioteca da Catalunya-PADICAT</i>	<p>En estos dos enlaces puede consultar dos documentos con la política de preservación digital de la Biblioteca:</p> <p>https://recercat.cat/bitstream/handle/2072/1757/IN_PADICAT_2005.pdf?sequence=1</p> <p>https://www.recercat.cat/handle/2072/97251</p>	<p>En este enlace del repositorio PADICAT podrá estudiar cómo se realiza el proceso para la preservación de los sitios webs:</p> <p>https://www.padicat.cat/es/conocenos/infraestructura-y-funcionamiento</p>	<p>En una etapa inicial, cuando se creó PADICAT, sí que había una unidad de la Biblioteca dedicada a este fin. Actualmente, sólo hay una persona —que trabaja en el Área de Dirección— dedicada a la coordinación, pero no se dedica exclusivamente a este tarea. La Biblioteca cuenta con la colaboración del Consorci de Serveis Universitaris de Catalunya (CSUC), para las cuestiones más técnicas</p>	Bibliotecarios e informáticos	<p>Los problemas más comunes son:</p> <ul style="list-style-type: none"> - La captura no es exhaustiva y tampoco los niveles. Una web de promedio dura menos de 100 días. Son recursos efímeros. Tenemos que hacer mucha pedagogía para explicar que el repositorio ofrece una foto fija de la web en un momento determinado. No funciona como repositorio institucional de preservación. - Problemas de visualización de algunas capturas. - Falta de recursos. - Las redes sociales no se capturan correctamente. <p>Actualmente, este es un problema importante, ya que mucha de la información en la red proviene de estos sitios. Por ejemplo, Padicat publica monográficos sobre temas concretos como son las elecciones; parte importante de la información de los candidatos está en las</p>	<p>Padicat es un repositorio de acceso abierto y libre desde cualquier sitio, por tanto, sí que fomenta el uso y el acceso.</p>

					redes sociales.	
<i>Biblioteca Nacional de Chile</i>	<p>El Archivo de la web chilena, es un proyecto piloto que ha sufrido de varios problemas por su complejidad y financiamiento. Lamentablemente no tenemos documentación escrita sobre políticas, planes estrategias o manuales que se le pueda enviar, actualmente, estamos en proceso de migración de nuestra sala de servidores a un servicio externo lo que nos ha obligado a detener el crecimiento de nuestras colecciones. La BN no hace cosecha del dominio, si no que de ciertos sitios de manera selectiva que se definen en base a política general de desarrollo de colecciones.</p>	<p>Los sitios predefinidos se cosechan periódicamente, son almacenados en formato warc los que se ingresan a los servidores de almacenamiento institucionales que cuentan con sus políticas de preservación y respaldo de información. Como comenté anteriormente en éste momento se está llevando a cabo un proceso de migración y en el nuevo hosting se han solicitado mantener los mismos sistemas de seguridad y respaldo de la información.</p>	<p>No hay un área exclusiva de preservación digital, ésta se lleva a cabo desde el Departamento de Colecciones Especiales y Digitales el Laboratorio de digitalización en estrecha colaboración por la Unidad de Sistemas de Información Bibliográfica que tienen una sección de ingeniería informática.</p>	<p>Conservadores, bibliotecólogos e ingenieros informáticos.</p>	<p>En general los principales problemas están dados por falta de recursos económicos, humanos y técnicos y por la complejidad de preservación intrínseca a éste tipo de materiales. Se es parte del IIPC donde se conocen y comparten estrategias y buenas prácticas, sin embargo son difíciles de aplicar dada la escasez de recursos.</p>	<p>Tenemos publicado en Biblioteca Nacional Digital, el Archivo de la Web chilena que muestra de manera restringida por ley de propiedad intelectual, algunos de las colecciones cosechadas. Se ha participado en conferencias y congresos nacionales e internacionales, dando a conocer nuestras actividades y se publican esporádicamente en las redes sociales institucionales.</p>