

UNIVERSIDADE DE SÃO PAULO
ESCOLA DE COMUNICAÇÕES E ARTES
DEPARTAMENTO DE INFORMAÇÃO E CULTURA
MIRELLA DEL MAZZA

404 - Not Found: Uma análise do arquivamento da web no Brasil,
suas deficiências e potencialidades.

São Paulo
2024

MIRELLA DEL MAZZA

404 - Not Found: Uma análise do arquivamento da web no Brasil,
suas deficiências e potencialidades.

Trabalho de Conclusão de Curso apresentado ao
Departamento de Informação e Cultura da Escola de
Comunicações e Artes da Universidade de São Paulo
para obtenção do título de:
404 - Not Found: Uma análise do arquivamento da web
no Brasil, suas deficiências e potencialidades..
Orientado: Prof. Dr.Martin Grossmann.

São Paulo
2024

Nome: Mirella Del Mazza

Título: **404 - Not Found**: Uma análise do arquivamento da web no Brasil, suas deficiências e potencialidades.

Trabalho de conclusão de curso apresentado ao Departamento de Informação e Cultura da Escola de Comunicações e Artes da Universidade de São Paulo, como parte dos requisitos para obtenção do grau de bacharel em Biblioteconomia.

Aprovado em: 29/08/20224

Banca examinadora

Nome: Prof. Dr. Martin Grossmann (Orientador)

Instituição: ECA-USP

Nome: Prof. Dr. Eugênio Bucci

Instituição: ECA-USP

Nome: Prof. Dr. Demi Getschko

Instituição: NIC.br (Núcleo de Informação e Coordenação do Ponto BR)

Agradecimentos

Agradeço à minha família e amigos, vocês sempre foram meu norte e me ensinaram o valor do aprendizado.

Ao meu marido, Bruno, que esteve ao meu lado desde a contagem do gabarito do vestibular até a defesa de meu TCC. Seu amor é minha grande inspiração.

Aos professores que me acompanharam e ensinaram tanto ao longo da graduação, engrandecendo minha formação e abrindo portas. Agradeço especialmente meu orientador, professor Martin Grossmann, que acreditou em meu potencial e tanto contribuiu para a realização do presente trabalho.

Resumo

O presente trabalho analisa aspectos centrais relacionados à memória digital, em particular a atividade de arquivamento da web a partir da comunicação científica presente em bases de dados nacionais e internacionais. Para tanto, verifica como o tema arquivamento da web é tratado na literatura especializada. A partir da análise dos resultados obtidos nas bases de dados, aborda o desenvolvimento do arquivamento da web a partir da análise de três iniciativas internacionais: O Internet Archive, O International Internet Preservation Consortium, e o Arquivo.pt, e dos procedimentos desenvolvidos a partir dos anos de 1990 para viabilizar o arquivamento na web no mundo, delimitando os critérios mínimos para a realização do arquivamento da web. Compara o desenvolvimento internacional do arquivamento da web ao brasileiro para verificar qual o papel dos agentes governamentais e das instituições de memória na preservação da web brasileira para, por fim, tecer um comentário sobre a ligação entre a constituição da memória social digital e o arquivamento da web.

Palavras-chave: Arquivamento da web. Memória sócio-cultural digital. Arquivos da web brasileira. Preservação da web

Abstract

The present work analyzes central aspects related to digital memory, in particular the activity of web archiving based on scientific communication present in national and international databases. To do so, it checks how the topic of web archiving is treated in specialized literature. From the analysis of the results obtained in the databases, it addresses the development of web archiving based on the analysis of three international initiatives: The Internet Archive, The International Internet Preservation Consortium, and Arquivo.pt, and the procedures developed based on of the 1990s to make web archiving viable around the world, defining the minimum criteria for carrying out web archiving. It compares the international development of web archiving to the Brazilian one to verify the role of government agents and memory institutions in the preservation of the Brazilian web and, finally, to comment on the link between the constitution of digital social memory and the archiving of web.

Keywords: Web archiving. Digital socio-cultural memory. Brazilian web archives. Web preservation

Sumário

1 Introdução	8
1.1 Problemática	11
1.2 Objetivos	12
1.2.1 Objetivo Geral	12
1.2.2 Objetivos específicos	12
1.3 Justificativa	12
1.4 Metodologia	13
2 Internet, seus websites e a memória social digital	20
2.1 Breve Histórico	24
2.2 Como funciona o arquivamento da web	27
2.2.1 Seleção	28
2.2.2 Definição de metadados	29
2.2.3 Captura de dados	29
2.2.4 Controle de qualidade	30
2.2.5 Armazenamento	30
2.2.6 Acesso ao conteúdo	31
2.3 Iniciativas relevantes para o arquivamento da web	31
2.3.1 Internet Archive	34
2.3.2 Consórcio Internacional de Preservação da Internet	36
2.3.3 O Arquivo.pt	39
2.4 Internet e a Web no Brasil	43
3. Arquivamento da web brasileira	47
3.1 Desenvolvimento de coleções da web	49
3.2 Arquivamento web institucional e governamental	54
3.3 Políticas de arquivamento web no Brasil	59
4 Considerações finais: o documento web e a memória social digital	65
Referências	70

1 Introdução

Ao longo dos séculos, a humanidade vem produzindo registros sobre si mesma e a natureza, é o que indicam as pinturas rupestres, cujos registros remontam ao período Paleolítico. Entretanto, não é possível precisar o momento em que a humanidade sentiu a necessidade de organizar sua produção de registros, o que não nos impede de pensar que a capacidade de produzir estes registros fez com que surgisse também a necessidade de guardá-los e ordená-los. Não se sabe o momento exato em que a humanidade pensou na necessidade de constituir uma biblioteca, no entanto, segundo Ortega (2004), há um marco importante descoberto em 1975, quando foi comprovada a existência das primeiras coleções organizadas de documentos (ou “biblioteca primitiva”). Datada do terceiro milênio a.C., trata-se da Biblioteca de Ebla, na Síria, uma coleção de textos administrativos, literários e científicos, registrados em 15 mil tábuas de argila organizadas em estantes segundo o tema abordado, além de 15 tábuas pequenas com resumos do conteúdo de documentos. O início da escrita propriamente dita também é estimado em meados do terceiro milênio a.C., o que denota que a escrita e a organização sistemática de documentos escritos são coincidentes. Para Milanesi (1998), a história da humanidade e sua evolução é concomitante ao desenvolvimento da ideia de biblioteca:

A história da biblioteca é a história do registro da informação, sendo impossível destacá-la de um conjunto amplo: a própria história do homem. Na medida da produção do registro informativo, o homem engendrou sistemas – tão rudimentares quanto a informação registrada – para não dispersá-la. (MILANESI, 1998, p. 16).

Um registro tem como objetivo a permanência da informação, se esta não pode ser acessada, torna-se inexistente. Ao longo dos séculos, a existência das bibliotecas se consolidou, desde a Biblioteca da Antiguidade, passando pela Idade Média, até chegar ao Renascimento e a criação de Universidades, o que possibilitou o surgimento do profissional dedicado ao acervo e conseqüentemente à preservação da informação. Essas necessidades e intencionalidades geraram processos sistêmicos de organização, conservação e preservação, que por sua vez geraram especializações, como é o caso da própria biblioteconomia, como também da documentação, arquivística e museologia, todas profissões coletoras de cultura,

como menciona Smit (1993) ao se referir às “3 Marias”. Assim sendo, a Biblioteconomia surgiu como disciplina formadora de profissionais dedicados ao estudo e desenvolvimento de técnicas de manejo e preservação da informação, seja ela em suportes físicos como livros, documentos em papel ou até mesmo obras de arte. A catalogação e a classificação desse tipo de informação sempre foram feitas preservando o suporte e levando em consideração as particularidades desses suportes.

Quando se pensa na produção de conhecimento e informação atual, nota-se que ela é marcada especialmente pelo suporte digital, ou seja, a ideia de “suporte” perpassa o entendimento do que é a Internet, a qual hospeda os websites que trazem em si informações em diferentes formatos, sendo textos, músicas, animações, entre outros. Nota-se que estamos adentrando em um universo novo, onde o suporte é imaterial. O tema fica mais complexo ao percebermos que cada vez mais existe uma tendência a encararmos a preservação da memória como responsabilidade das empresas detentoras da tecnologia de desenvolvimento dessas redes e websites. Se anteriormente a biblioteca preservava a produção humana impressa em livros, arquivos, revistas, entre outros suportes, agora não há muita clareza de como se deve preservar a produção de informação pública e virtual, afinal as empresas de tecnologia, ou big techs, não costumam ser transparentes em seus processos de desenvolvimento de produtos. Há um incentivo enorme na produção cada vez maior de informações que geram dados para estas empresas, mas não há da parte das big techs uma contrapartida na colaboração para iniciativas de preservação desses conteúdos, já que estes atualmente são vistos como um bem de consumo.

Ao se observar a produção ininterrupta de informações em meios digitais, a ideia de preservação da web impacta conceitos importantes como o da opinião pública. Como argumenta o professor Eugênio Bucci (2019), os usuários de redes sociais não entendem como as plataformas funcionam e, ao mesmo tempo, nelas consomem informações de modo acrítico, o que cria o cenário ideal para a disseminação de notícias fraudulentas, esse consumo sem criticidade ainda influencia para uma interdição dos fatos através do que o autor chama de “apagões de real”, um conceito que demonstra como a tecnologia tem como procedimento a substituição dos fatos propriamente ditos por dados digitais, bloqueando o acesso do debate público aos fatos. Para o autor, fatos estão sendo representados pelos

dados, o que inviabiliza o contato das pessoas com a verdade factual e, consequentemente, empobrece a opinião pública.

É certo também que as tecnologias digitais têm um impacto no que se considera como memória, aquilo que persiste na Internet não necessariamente tem a ver com a sua relevância cultural, mas muitas vezes pode estar atrelado a interesses econômicos de corporações. Some-se a este fato a aparente “maior autonomia” de qualquer cidadão possui para produzir informação virtual, sem que esse documento digital seja devidamente tratado, tanto no que diz respeito à sua preservação, quanto ao controle de sua reprodução e disseminação, lembrando que os softwares e aplicativos que permitem essa produção serem de código fechado, proprietário, o que torna a situação ainda mais complexa. Todo esse cenário mostra-se desafiador para o profissional de informação, seja ele um bibliotecário, museólogo ou arquivista. O atual cenário denota a necessidade de ampliar o debate acerca de questões relacionadas às ferramentas de preservação da memória, assim como quais são os vieses culturais que estão presentes na eleição de um conteúdo como relevante para preservação.

Há países e iniciativas internacionais que já estão há alguns anos buscando trazer soluções para a problemática exposta acima, conforme apontam os autores Moisés Rockembach e Caterina Marta Groposo Pavão (2024), desde que a World Wide Web passou a operar, em meados dos anos 1990, já existia a preocupação com a preservação dessas informações. A primeira resposta ao problema chegou já em 1996, com o Internet Archive (Estados Unidos), Pandora (Austrália) e KulturaW3 (Suécia). Atualmente, há projetos para a preservação da web focados em resguardar coleções relevantes para cada país, dentre estes, destacam-se as iniciativas de países como Estados Unidos, com cerca de 30 iniciativas de preservação da web em andamento, além do continente europeu, com quase 50 iniciativas. Porém, o Brasil está bastante atrasado neste assunto, tendo publicado duas resoluções sobre o tema arquivamento da web apenas em 2023. Como comparação, temos países como Reino Unido e Estados Unidos que possuem projetos consolidados de preservação da web em seus respectivos arquivos nacionais. No primeiro, o arquivamento de websites ocorre desde 2003.¹

¹ GALF, Renata. Caso Bolsonaro expõe travas para preservar rede social como documento público. **Folha de São Paulo**, São Paulo, 2024. Disponível em: <https://www1.folha.uol.com.br/poder/2024/01/brasil-engatinha-para-preservar-registros-virtuais-e-caso-bolsonaro-expoe-fragilidades.shtml>. Acesso em: 29 abr. 2024.

O atraso brasileiro em buscar iniciativas efetivas de arquivamento da web pode trazer consequências não apenas no futuro, mas também para o momento presente, uma vez que materiais potencialmente relevantes, e que só estão disponíveis na Internet, podem ser perdidos em um curto intervalo de tempo, isso se dá pois a efemeridade dos links para páginas da web é uma realidade que nenhuma instituição consegue reverter. O único modo de preservar uma página da web é fazendo um esforço ativo de preservação em repositório específico, contar com a permanência de um link no ambiente virtual não é uma opção, já que estudos demonstraram que a vida útil de uma página da web é curta, por exemplo, em 2004, os autores Ntoulas; Cho; Olston, (2004) publicaram artigo defendendo que após um ano, apenas 20% dos endereços da web conseguem se manter válidos. Levando em consideração que cada vez mais as informações da web são relevantes para a compreensão da sociedade em que vivemos, há a necessidade de que a Biblioteconomia e a Documentação enfrentem com seriedade e atenção o desafio de compreender a produção de informação na web como um ambiente chave para a preservação da memória social.

1.1 Problemática

As considerações levantadas na introdução do presente trabalho nos conduzem ao seguinte questionamento: por que o Brasil não preserva sua memória na web? Essa pergunta pode ser desdobrada em tópicos como:

- Por que o Brasil ainda não tem ações relevantes, tanto em iniciativas globais quanto locais, para o arquivamento da web brasileira?
- Qual o papel dos agentes governamentais e das instituições de memória para a preservação da web brasileira?
- Quais iniciativas passadas ou em curso existem e o que falta para estas acelerarem de modo significativo o arquivamento da web no Brasil?

Nesse sentido, há também o importante debate acerca da memória digital em tempos de disseminação massiva de informação via Internet. Todas essas questões levam a uma busca de maior entendimento do arquivamento da web, como ocorreu

seu desenvolvimento desde as primeiras iniciativas em diversos locais do mundo, sua relevância atual, e quais mecanismos possíveis para que este se torne parte da memória digital em um contexto social.

1.2 Objetivos

Para buscar respostas às questões postas na problemática, o presente trabalho investigará o assunto a partir de hipóteses que apontam para quais são os parâmetros mínimos, dentro das teorias da Biblioteconomia, para concepção de uma política de preservação da web e seu entendimento como patrimônio digital brasileiro. Para isso, foi feita uma revisão da literatura acerca do tema, assim como a análise de iniciativas nacionais e internacionais voltadas para o arquivamento da web.

1.2.1 Objetivo Geral

O objetivo do presente trabalho é identificar deficiências no trabalho de arquivamento da web brasileira e apontar caminhos para superá-las.

1.2.2 Objetivos específicos

- Definir como o tema arquivamento da web é tratado na literatura especializada.
- Definir os critérios mínimos para a realização do arquivamento da web.
- Elencar iniciativas relevantes para o arquivamento da web, nacionais e internacionais.
- Tecer um comentário sobre a ligação entre a constituição da memória social digital e o arquivamento da web.

1.3 Justificativa

A memória é um conceito importante para as Ciências Sociais como um todo, em se tratando de uma área de estudo como a Biblioteconomia e Documentação, que tem como objetivo a preservação da informação em instituições de memória, parece natural que este conceito perpassasse os estudos técnicos que fornecem

mecanismos eficientes para a manutenção da informação. É importante salientar que o conceito de informação diz respeito a “um fenômeno humano e social, que não apenas dá materialidade a conceitos subjetivos como também permite a troca efetiva e interação de ideias entre pessoas” (Silva, 2006, p. 150), com esse conceito em mente, a Biblioteconomia mostra-se uma disciplina apta a realizar a interseção entre informação, mecanismos tecnológicos e manutenção da memória social. No entanto, a observação de publicações especializadas na área demonstra que o Brasil vem falhando em suas potencialidades para formar um corpus relevante na preservação da memória por meio de sua produção na web brasileira, denotando um risco ao futuro da preservação de sua história. Isso posto, o presente trabalho visa iniciar um debate sobre a falha brasileira no que tange à preservação da web e os principais impactos dessa deficiência, especialmente no que tange à concepção de uma memória social digital.

1.4 Metodologia

Para a realização do presente trabalho, a metodologia escolhida foi de natureza qualitativa e exploratória. O objetivo foi, primeiramente, realizar um levantamento bibliográfico sobre o tema arquivamento da web, o qual contemplou livros, teses, dissertações, artigos e publicações na imprensa, privilegiando o recorte cronológico entre 2012 a 2023, a escolha de haver esse recorte dos últimos onze anos deve-se à velocidade com que as novas tecnologias alteraram a relação entre informação digital e impacto na vida de usuários e pensadores da área. Após o levantamento, foi realizada uma breve análise da literatura escolhida, com o objetivo de elaborar uma revisão bibliográfica acerca do tema.

O levantamento bibliográfico teve como objetivo buscar estudos sobre o tema no Brasil, para isso, foram usadas as fontes de informação Brapci, BDTD e Dedalus. Para o levantamento sobre o tema em publicações internacionais, foi usada a base de dados Scopus, as pesquisas nas bases mencionadas estão relacionadas aos objetivos específicos do trabalho, os quais foram citados anteriormente. O intuito da pesquisa buscou efetuar uma revisão bibliográfica capaz de apontar tendências na área de Biblioteconomia e Documentação acerca do tema e buscar compreender qual a relevância do arquivamento web para a área.

A revisão bibliográfica gerou um quadro que busca resumir o modo como o tema do arquivamento da web é tratado no Brasil e no mundo, a partir desse quadro foram escolhidas três iniciativas internacionais mais relevantes, as quais serão melhor analisadas ao longo do presente trabalho. O objetivo de se debruçar nas iniciativas internacionais é ter um melhor entendimento do tema e também buscar caminhos para a consolidação dos melhores procedimentos para que o Brasil tenha uma participação mais efetiva no arquivamento da web,

A estratégia de busca de literatura iniciou-se com a busca dos seguintes conceitos nas bases de dados: “arquivamento da web”, “preservação da web”, “arquivos da web brasileira”, “web archiving”, “web preservation”, “internet archive”, “International Internet Preservation Consortium”. Os resultados podem ser observados no quadro a seguir:

Quadro 1: Busca de Literatura

(continua)

Características das bases consultadas	Biblioteca Digital Brasileira de Teses e Dissertações	BRAPCI	Dedalus	Scopus
URL	bdtb.ibict.br	www.brapci.inf.br	dedalus.usp.br	http://www.scopus.com
Tipo de busca	Avançada	Simples	Avançada	Simples
Campos	Todos	Todos	Todos	Todos
Período	2004-2023	2004-2023	2004-2023	2004-2023
Termos de Busca				
“arquivamento da web”	10 resultados	31 resultados	6 resultados	-
“preservação da web”	2 resultados	1 resultado	1 resultado	-
“arquivos da web brasileira”	1 resultado	31 resultados	0 resultados	-
“web archiving”	-	-	-	1.128 resultados

Quadro 1: Busca de Literatura

(conclusão)

Características das bases consultadas	Biblioteca Digital Brasileira de Teses e Dissertações	BRAPCI	Dedalus	Scopus
URL	bdtb.ibict.br	www.brapci.inf.br	dedalus.usp.br	http://www.scopus.com
Tipo de busca	Avançada	Simples	Avançada	Simples
Campos	Todos	Todos	Todos	Todos
Período	2004-2023	2004-2023	2004-2023	2004-2023
Termos de Busca				
"web preservation"	-	-	-	64 resultados
"internet archive"				810 resultados
"International Internet Preservation Consortium"				58 resultados

Fonte: elaborado pela autora.

Quadro 2: Análise da Literatura Mais Relevante em Bases Internacionais

(continua)

Busca geral		
Artigo	Objetivos	Métodos
Web archiving: techniques, challenges, and solutions Adoghe Anthony 14 Set 2013	Coletar conteúdo da web em formato de arquivo para preservação. Técnicas, desafios e soluções em arquivamento na web.	Técnicas usadas no arquivamento da web Desafios enfrentados e soluções propostas em arquivamento na web

Quadro 2: Análise da Literatura Mais Relevante em Bases Internacionais

(conclusão)

Busca geral		
Artigo	Objetivos	Métodos
An Overview of Web Archiving Jinfang Niu 29 Fev 2012	O arquivamento na Web envolve selecionar, adquirir, descrever e acessar recursos da Web. Os métodos incluem conceitos tradicionais de gerenciamento de arquivos aplicados à organização de conteúdo da web.	Revisão de literatura realizada para o desenvolvimento de cursos de arquivamento web. Os métodos incluem seleção, aquisição, organização, descrição e acesso.
Where Did the Web Archive Go Mohamed Aturban, Michael L. Nelson, Michele C. Weigle 11 Ago 2021	4 arquivos da web alteraram os URIs básicos sem métodos legíveis por máquina. 537 lembranças impactadas, 20 lembranças não foram encontradas.	Investigação longitudinal de arquivos da web Redescoberta manual devido a mudanças nos URIs básicos
ArchiveNow: Simplified, Extensible, Multi-Archive Preservation Mohamed Aturban 22 Maio 2018	Módulo Python para preservar páginas da web em vários arquivos da web. Permite que os usuários gerem arquivos WARC locais para arquivos pessoais.	O ArchiveNow captura páginas da web, fornece links para cópias arquivadas. Os usuários podem gerar arquivos WARC locais para arquivos pessoais.
The evolution of web archiving Miguel Costa 31 Ago 2017	Os arquivos da Web preservam informações digitais historicamente valiosas. Crescimento significativo em iniciativas, países, volume de dados e conteúdos preservados.	Pesquisas realizadas em 2010 e 2014 Contato direto com a comunidade de arquivamento da web

Fonte: elaborado pela autora.

Quadro 3: Análise da Literatura Mais Relevante em Bases Nacionais

(continua)

Busca em Português-BR		
Artigo	Objetivos	Métodos
História do arquivamento da web no Brasil: um percurso entre a academia, o legislativo e executivo brasileiro 10 Jun 2023 Jonas Ferrigolo Melo	O arquivamento da web no Brasil ainda não está integrado às práticas institucionais. Iniciativas em andamento para desenvolver o arquivamento web no Brasil.	Pesquisa baseada em fontes bibliográficas e documentais Revisão da literatura sobre arquivamento governamental na web, iniciativas brasileiras, leis.
A web brasileira na Covid-19: arquivamento da web e preservação digital Moisés Rockembach 16 Jun 2021	Concentra-se no arquivamento web do conteúdo brasileiro da COVID-19. Enfatiza a importância de preservar o conteúdo da web para acesso futuro.	Analisa iniciativas internacionais de preservação da web, especialmente o IIPC. Coletou 16,4 GB de arquivos da web relacionados ao COVID-19
Preservação digital: desafios, requisitos, estratégias e produção científica Danilo Formenton 07 Jun 2020	O arquivamento da web brasileira carece de estudos em Ciência da Informação. O Brasil se destaca na produção científica internacional sobre preservação digital.	Pesquisa quantitativa-qualitativa e exploratório-descritiva Revisão da literatura nacional e internacional sobre preservação digital
Políticas E Tecnologias De Preservação Digital No Arquivamento Da Web Moisés Rockembach 31 Jan 2018	O Brasil carece de iniciativas de arquivamento na web, mas pode contribuir internacionalmente. Políticas de seleção e tecnologias de código aberto são cruciais para a preservação.	Pesquisa bibliográfica e documental sobre iniciativas internacionais de arquivamento na web Pesquisa quantitativa e qualitativa sobre políticas e tecnologias de arquivamento na web

Quadro 3: Análise da Literatura Mais Relevante em Bases Nacionais

(conclusão)

Busca em Português-BR		
Artigo	Objetivos	Métodos
A divulgação do patrimônio documental dos arquivos universitários das IFES da região Sul do Brasil em Websites Eliseu dos Santos Lima 19 Out 2015	Poucos arquivos universitários no sul do Brasil usam sites para divulgação. Apenas quatro instituições identificadas têm presença on-line para informações arquivísticas.	Abordagem qualitativa descritiva Revisão de literatura em sites de instituições arquivísticas do IFES

Fonte: elaborado pela autora.

A quantidade de registros voltados para a definição do que é arquivo da web ou preservação da web se mostrou de volume baixo de publicações no Brasil, nota-se que o país com mais relevância no quesito, tanto em número de publicações quanto de citações, são os EUA. Dentre os autores que se destacaram na produção científica sobre os temas sugeridos na busca, destacou-se o autor norte-americano Michael L. Nelson, da área de Ciência da Computação. No Brasil, o autor Moisés Rockembach destaca-se como um dos mais citados e também com mais artigos publicados sobre o assunto.

A busca de literatura especializada sobre o tema arquivamento da web levou a algumas escolhas para o presente trabalho, primeiramente, foi possível estabelecer o protagonismo do Internet Archive na discussão acerca de instituições relevantes para o trabalho de arquivamento da web, além de ser a primeira iniciativa de que se tem notícia no mundo dedicada a esta atividade. Seguido pelo Internet Archive está o International Internet Preservation Consortium, uma organização internacional de bibliotecas e outras organizações, criada com o intuito de coordenar esforços para preservar o conteúdo da Internet para o futuro. Já quando se buscam estudos sobre o arquivamento de páginas em da web em língua portuguesa, a iniciativa de maior relevância é o arquivo da Web de Portugal, conhecido como

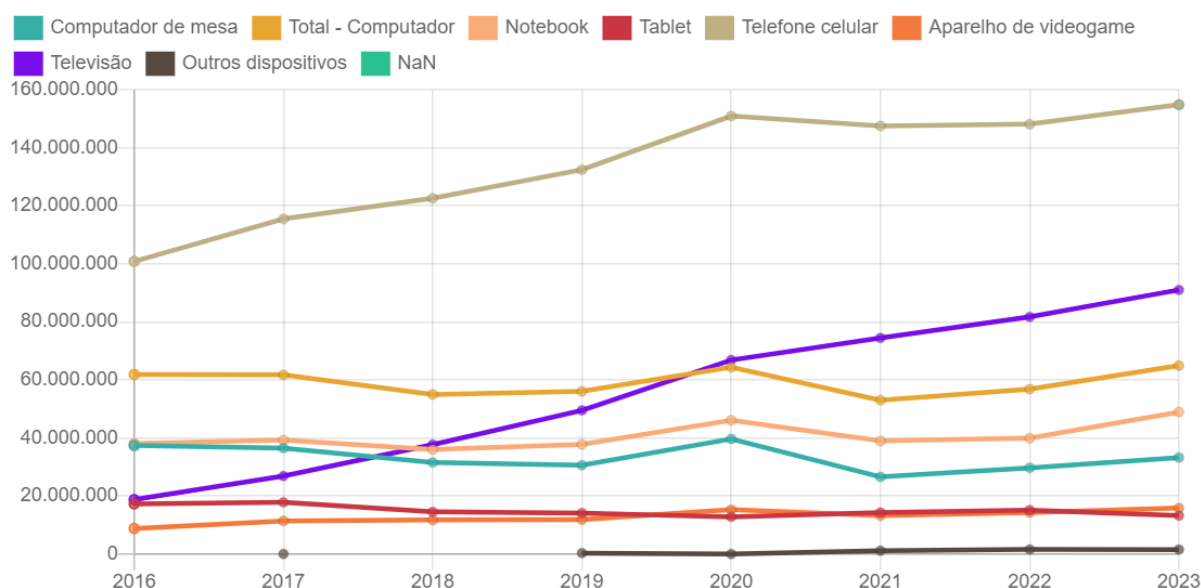
Arquivo.pt, o projeto visa arquivar e preservar conteúdos da Web relevantes para a comunidade portuguesa, mas também possui muitas informações relevantes sobre outros países lusófonos, como o Brasil.

A análise da bibliografia levantada também demonstrou que os estudos do arquivamento da web voltados para a área das Ciências Sociais levam em consideração não apenas aspectos técnicos, ou seja, quais métodos e requisitos tecnológicos são necessários para um profissional executar o arquivamento da web, mas também os estudos visam desenvolver uma análise filosófica do impacto da web no debate público, assim como os modos de desenvolvimento de informação diante da multiplicidade de vozes presente no ambiente digital. Por esse motivo, o presente trabalho irá também apresentar uma breve análise acerca do impacto da web na construção da memória social digital.

2 Internet, seus websites e a memória social digital

O impacto da tecnologia da Internet para a sociedade capitalista é inegável, nesses mais de 30 anos de uso comercial da ferramenta, é possível observar diversas mudanças no cotidiano das pessoas, tais como a substituição das cartas por correio eletrônico (e mais recentemente os aplicativos de mensagens rápidas), a expansão do e-commerce em lugar das compras em lojas físicas, e, especialmente, na onipresença de smartphones em detrimento dos computadores pessoais, essa última mudança no modo como se acessa a Internet, parece ser a mais aguda ao se pensar no impacto social, afinal, em um mundo conectado via smartphones, não existe mais o ritual de “entrar na Internet”, estamos todos o tempo todo on-line e aptos a receber informações mesmo que não solicitadas.

Gráfico 1: Usuários de Internet, por dispositivo utilizado



Fonte: CGI.br/NIC.br, Centro Regional de Estudos para o Desenvolvimento da Sociedade da Informação (Cetic.br), Pesquisa sobre o Uso das Tecnologias de Informação e Comunicação nos domicílios brasileiros.

O gráfico acima demonstra o impacto do smartphone no Brasil nos últimos 8 anos, são quase 160 milhões de brasileiros cotidianamente, como preconizou Teixeira Coelho (2015), com os cérebros na mão, capazes de acessar em qualquer momento e com a rapidez de alguns cliques, tudo que a humanidade foi capaz de produzir, ou ao menos tudo que a humanidade produziu e está

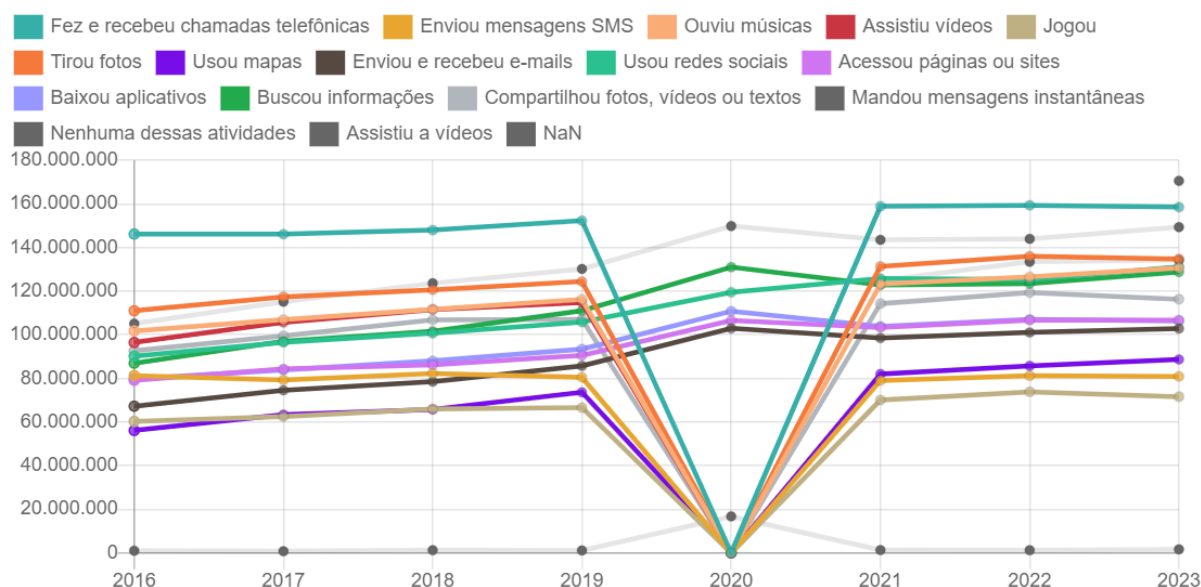
salvaguardado na web, ou seja, está rastreável e acessível. Neste ponto é importante retomar outro trecho do livro de Teixeira Coelho em que ele ressalta que estamos com nossos smartphones o tempo todo em mãos, enquanto nosso cérebro dentro da cabeça desconhece o que o este pequeno cérebro vê ou pensa, isso passa uma ideia de neutralidade da tecnologia, mas sabe-se que é impossível haver neutralidade em um meio de comunicação, a informação está dada e pode ser interpretada de modos diversos, como ressalta o autor:

Os novos meios técnicos são entregues vazios ou neutros de conteúdo a seus usuários, que deles se servem segundo suas tendências de comportamento, verdadeiro conteúdo desses meios — e uma boa parte desse conteúdo está sendo preenchido pelos porcos-espinhos, num movimento conservador, ou reacionário, de intensidade inédita. (COELHO, 2015, p. 51)

É perceptível que estamos diante de uma importante mudança social decorrente do fato de que a principal fonte de informações para grande parte da população mundial seja oriunda de ferramentas digitais. No gráfico a seguir, também apresentado na Pesquisa TIC Domicílios 2023, do Centro Regional de Estudos para o Desenvolvimento da Sociedade da Informação (Cetic.br), é apresentado o resultado da pesquisa realizada todos os anos entre 2016 e 2023 (exceto 2020, devido à pandemia de COVID-19), nas pesquisas anuais, foi feita uma pergunta aos usuários se eles poderiam apontar quais as principais atividades por eles desempenhadas com a ajuda do celular nos últimos três meses, as respostas demonstram que a população brasileira usa o telefone celular primeiramente para fazer e receber chamadas telefônicas, mas a conexão à Internet é especialmente usada para buscar informações e usar redes sociais, ou seja, a Internet mostra-se mais uma vez onipresente no cotidiano das populações, impondo um novo modo de se relacionar com a informação, uma vez que o usuário não tem mais o papel proativo de buscar e selecionar a informação de que precisa, no lugar da busca ativa, há o constante bombardeamento de informações vindas de uma curadoria de grandes empresas de tecnologia, estas usam robôs para ranquear a relevância de uma informação e apresentá-la ao usuário segundo um perfil pré-estabelecido por meio de análises como curtidas em redes sociais, tempo de tela, entre outros dados

que as “big techs” capturam da nossa atividade on-line sem que esse processo de captura de dados de usuários nos seja devidamente informados.²

Gráfico 2: Usuários de telefone celular, por atividades realizadas no telefone celular nos últimos três meses:



Fonte: CGI.br/NIC.br, Centro Regional de Estudos para o Desenvolvimento da Sociedade da Informação (Cetic.br), Pesquisa sobre o Uso das Tecnologias de Informação e Comunicação nos domicílios brasileiros.

Neste contexto bastante complexo, é preciso pensar na memória, naquilo que deve permanecer do ruído constante das informações em mídias digitais. As instituições que têm como missão a preservação da memória sempre foram desafiadas a lidar com o tratamento do que a mente humana foi capaz de produzir em qualquer tipo de suporte físico e ter clareza no que significa o conceito de memória social é o primeiro passo para trabalhar este tema. Segundo Dodebei (2011, p. 37), a ideia de memória social remete ao campo de estudos que “investiga as propriedades e os fenômenos que ocorrem com indivíduos ou com grupos sociais nas relações que estabelecem entre os fatos, imagens e acontecimentos”. A memória é então uma construção coletiva, que leva em consideração o grupo que a compartilha, que também se forma e deforma com o passar do tempo e diferentes

²Google confirms the leaked Search documents are real. The Verve
<https://www.theverge.com/2024/5/29/24167407/google-search-algorithm-documents-leak-confirmation?ref=hackernoon.com>

interpretações de um mesmo fato, assim, a memória social está também ligada à noção de patrimônio, de legado, daquilo que deve permanecer para as futuras gerações.

O Artigo 216 da Constituição Brasileira conceitua patrimônio cultural como sendo os bens “de natureza material e imaterial, tomados individualmente ou em conjunto, portadores de referência à identidade, à ação, à memória dos diferentes grupos formadores da sociedade brasileira” (BRASIL, 1988). A partir desse entendimento, pode-se pensar que o conteúdo produzido na Internet tem um potencial de impactar a memória social e, logo, pode ser tratado como um patrimônio, no sentido de valorizar essa produção digital como informação que contribui para a formação de nossa identidade. A circulação de informações na web possui uma linguagem particular, como por exemplo os memes que são imagens repetidas e reinterpretadas continuamente, com sentidos muitas vezes cifrados e limitados à comunidade que os compartilha. Adentrar nessas comunidades pode ser uma importante fonte de pesquisa para compreender movimentos sociais de agora e garantir a memória para a posteridade, portanto, entender como se constitui a memória de nosso tempo é um passo importante para o estudo da Biblioteconomia e da Documentação.

Conforme Le Goff (1990, p. 427), os tempos da memória são marcados pela transição entre a oralidade e a escrita, o autor a divide então em cinco segmentos: a memória étnica nas sociedades sem escrita, ditas “selvagens”; o desenvolvimento da memória, da oralidade à escrita, da Pré-história à Antiguidade; a memória medieval, em equilíbrio entre o oral e o escrito; os progressos da memória escrita, do século XVI aos nossos dias; e os desenvolvimentos atuais da memória. Sobre os últimos, o autor os caracteriza como uma verdadeira revolução da memória, sendo a memória eletrônica o mais espetacular dos elementos relacionados aos desenvolvimentos atuais da memória. Le Goff conclui que a memória das máquinas “se impõe pela sua grande estabilidade, semelhante ao tipo de memória que representa o livro, mas combinada, no entanto, com uma facilidade de evocação até então desconhecida” (1990, p. 468). A potencial facilidade de evocação da informação digital contrasta com o que se observa atualmente quando buscamos websites – muitas vezes com menos de dois anos de sua publicação - e falhamos em conseguir acessá-los.

Para melhor entender como a Internet transformou a relação humana com a informação e a construção da memória social, é preciso voltar ao início e entender como a Internet surgiu na vida das pessoas, como se desenvolveram os primeiros websites, e como é possível realizar o arquivamento desses websites. O presente capítulo irá apresentar um panorama desses tópicos e aprofundar-se nos procedimentos para viabilizar o arquivamento da web, pois este se converte em uma medida importante para a construção da memória social de nosso tempo.

2.1 Breve Histórico

Antes de debater as estratégias possíveis para o arquivamento da informação disponível na Internet, é importante primeiramente apresentar um breve histórico de como surgiu a Internet, seu impacto, e quais foram os desdobramentos da introdução dessa tecnologia no cotidiano das pessoas. Segundo Castells (2003), as origens da Internet estão ligadas à criação da Arpanet, uma rede de computadores montada pela Advanced Research Projects Agency (ARPA) em setembro de 1969. De viés técnico-militar, a ARPA surgiu em 1958 e representou a tentativa do Departamento de Defesa dos Estados Unidos de angariar pesquisadores universitários em projetos tecnológicos voltados ao enfrentamento da ameaça que a União Soviética representava na época ao Estado norte-americano. Ou seja, a Internet é um projeto fruto da Guerra Fria, momento histórico em que a telecomunicação ganhou status estratégico nos conflitos militares entre os EUA, que defendiam um modelo econômico voltado para a sociedade capitalista, e a União Soviética, que fomentava a ideia de comunidades socialistas em todo o globo.

O desenvolvimento contínuo da Internet ao longo dos anos de 1960 levou a mesma a ser usada também no ambiente acadêmico, para fins de pesquisas diversas, ou seja, o primeiro passo fora do objetivo militar inicial da Internet foi para dentro das Universidades, afinal, estas também eram bastante acionadas para o desenvolvimento de tecnologias que auxiliassem seus respectivos países nas disputas políticas da época. Desse modo, a tecnologia da Internet foi se desenvolvendo como uma grande rede de computadores dispersos por todo o planeta, capazes de trocar dados e mensagens utilizando um protocolo comum, unindo usuários particulares e entidades de pesquisa.

Entre as décadas de 1970 e 1990, cresceu a pesquisa dentro do projeto para desenvolvimento de códigos abertos e voltados ao compartilhamento de conhecimento acadêmico. Entretanto, foi apenas a partir de 1990 que a tecnologia de redes de computadores saiu totalmente do domínio militar e passou a domínio público. Primeiramente, o governo dos EUA delegou a administração da Internet à National Science Foundation, mas essa decisão durou pouco, apenas cinco anos, o que resultou, em 1995, no início da operação privada da Internet.

Conforme Castells (2003), é possível verificar que ao longo do desenvolvimento da Internet, o papel dos usuários não se limitou a apenas espectadores passivos. Historicamente, a todo momento são introduzidos novos usos da tecnologia, porém o autor ressalta que o intervalo entre o processo de aprendizagem pelo uso, e de produção pelo uso, é extraordinariamente abreviado, quanto mais se produz informação na Internet, mais se aprende com esse uso, e quem aprende e a aperfeiçoa é o próprio usuário, Castells ressalta que desde a origem, a tecnologia da Internet absorveu os valores de abertura e cooperação.

Neste ambiente de abertura da Internet à iniciativa privada, faltava algo que a levasse ao público geral. Esse “algo” veio após a criação e desenvolvimento da World Wide Web, mais conhecida pela sigla www. Foi o programador inglês Tim Berners-Lee, que trabalhava no CERN, o Laboratório Europeu para a Física de Partículas, em Genebra, o responsável por desenvolver a interface gráfica que permitiu a criação dos sites, sendo estes nada mais do que um sistema de documentos em hipermídia interligados e executados na Internet. Já em 1989, Berners-Lee propôs um sistema de informação interligado que foi fundamental para o desenvolvimento do primeiro servidor web, o primeiro navegador web e um protocolo de formatação de documentos, chamado Hypertext Markup Language (HTML). Em 1991, o programador publicou a linguagem de marcação, dois anos depois, o código-fonte do navegador foi liberado para uso público e, a partir de 1993, os sites para uso do público em geral começaram a surgir. A partir desse momento, a expansão da Internet não parou, surgiram os navegadores (browsers) como o Internet Explorer, da Microsoft, além de provedores de acesso e portais de serviços. A Internet passou a ser cada vez mais utilizada até chegar ao patamar de presença imprescindível na vida das pessoas.

É interessante resgatar o texto produzido por Berners-Lee em 1989 no qual ele propõe ao CERN o que mais tarde viria a ser a base para a criação da World

Wide Web, intitulado “Information Management: A Proposal” trata-se de uma série de reflexões e proposições acerca do problema de perda de informação sobre sistemas complexos. Ou seja, a ideia de que a Internet provocaria o acesso a uma quantidade muito grande de informações, e que seria necessária a criação de um mecanismo que permitisse a busca eficiente dessas informações, já está no cerne da criação da World Wide Web. O desafio da organização da informação naturalmente nos leva a pensar em Paul Otlet, desde a sua estação de estudos construída como uma roda, que permitiria aos usuários pesquisar, ler e escrever através de uma imensa base de dados mecânica armazenada em milhões de fichas 3x5, passando pelo Mundaneum, com sua vasta edificação documentária, até a publicação do “Traité de documentation”, o qual traz as reflexões do autor sobre estruturas de informação em rede:

Os livros conservam energia mental, o conteúdo dos livros passa a outros livros quando eles mesmos são destruídos e toda criação bibliológica, por mais original e poderosa que seja, implica redistribuição, combinação e novas amalgamações dos dados anteriores (Otlet, *apud* Santos 2008, p. 58).

A criação da Internet, que levou ao desenvolvimento da World Wide Web, é fruto de grande influência da necessidade humana de produzir, disseminar, organizar e preservar informações, temas basilares da Biblioteconomia e da Documentação. É de extrema importância também observar que a Internet e a web foram desenvolvidas em meio a uma atmosfera de culto à liberdade individual característica dos anos de 1960 e 1970, momento em que as universidades foram importantes espaços para manter redes comunitárias de inovação tecnológica, como salienta Castells:

(...) todos os desenvolvimentos tecnológicos decisivos que levaram à Internet tiveram lugar em torno de instituições governamentais e importantes universidades e centros de pesquisa. A Internet não teve origem no mundo dos negócios. Era uma tecnologia ousada demais, um projeto caro demais, e uma iniciativa arriscada demais para ser assumida por organizações voltadas para o lucro. (Castells, 2003, p. 28)

A Biblioteconomia e a Documentação mostram-se, desse modo, totalmente alinhadas com os princípios que nortearam a criação da Internet e da web, fica claro que o assunto arquivamento do conteúdo da Web deve ser tratado com muita

atenção por esse campo do conhecimento. Segundo Costa; Gomes; Silva, (2017), 80% das páginas da web não estão mais disponíveis em sua forma original após 1 ano de sua publicação. Esse dado, aliado ao fato de que a difusão de informações em meio digital cresce a cada ano, demonstra que é necessário pensar no arquivamento da web como um trabalho consistente e de responsabilidade dos especialistas em gestão de informações e preservação da memória.

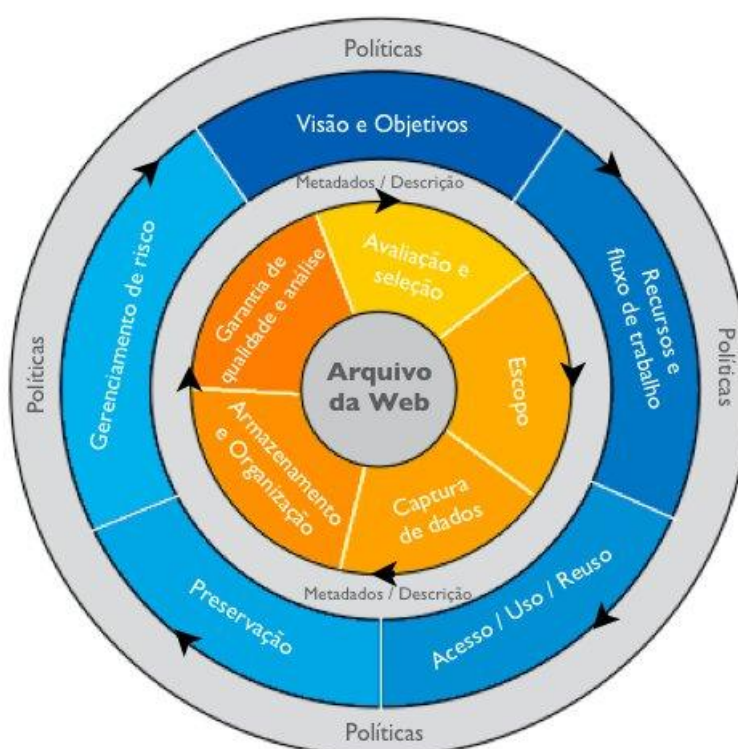
2.2 Como funciona o arquivamento da web

A Internet é um ambiente essencialmente dinâmico, a todo momento novas informações são criadas e compartilhadas especialmente no ambiente web e, para o público leigo, existe uma crença de que todo esse conteúdo publicado nas redes estará permanentemente acessível, porém, o fato é que ao mesmo tempo em que uma informação é criada, outras tantas estão sendo apagadas, além disso, informações antigas são constantemente sobrepostas por versões atualizadas, sem que o histórico permaneça. Este cenário dificulta a recuperação de informações importantes publicadas ao longo do tempo. Segundo Costa, Gomes e Silva (2017), dentre as referências online de artigos acadêmicos, 13% desaparecem após 27 meses da publicação, e 11% dos recursos de mídia social são perdidos após um ano. Esses dados apontam para o fato de que os motores de busca atualmente disponíveis para pesquisa na web não são suficientes para que aconteça efetivamente o arquivamento, pois segundo Rockembach e Pavão (2018), estes buscadores servem como indexadores da informação publicada, remetendo ao endereço do website e a responsabilidade de manter a informação disponível fica a cargo do dono do website.

Para que exista um arquivamento da web sistematizado, é necessário o desenvolvimento de uma política bem definida que aponte as atividades essenciais para que ele ocorra. Sobre o tema da política de arquivamento da web, cada instituição que a desenvolva irá estabelecer parâmetros próprios, no entanto, os requisitos técnicos mínimos para que este ocorra já estão em debate no meio especializado há algum tempo. O meio acadêmico produziu algumas publicações relevantes sobre o tema, podemos destacar o trabalho intitulado “The web archiving life cycle model” (Bragg; Hanna, 2013), o qual definiu as etapas de atividades essenciais para que exista um fluxo de arquivamento da web, ele pode ser melhor

visualizado no diagrama a seguir. A partir dos tópicos abordados no diagrama, é possível elencar como atividades essenciais a seleção do conteúdo a ser arquivado, a definição de metadados para a seleção, a definição da ferramenta de captura desses dados, o controle de qualidade da informação capturada, a definição do local de armazenamento da informação e a elaboração de um plano para garantir o acesso ao conteúdo arquivado.

Figura 1: Modelo de ciclo de vida do arquivamento da web



Fonte: Donovan, Hukill, Peterson, 2013 apud Rockembach (2017).

2.2.1 Seleção

O processo de definição de como será feita a seleção do que será arquivado é possivelmente o mais complexo e relevante no arquivamento da web, afinal a web é muito vasta e seria impossível tentar realizar o arquivamento total de seu conteúdo. Cabe à instituição que pretende fazer este trabalho definir um escopo específico de trabalho. Além disso, a seleção também deverá levar em consideração quais serão os usuários finais, ou seja, quem é o público em potencial que irá acessar o acervo.

Segundo Carreta (*apud* Brayner, 2016; Rockembach, Pavão, 2018), o processo de seleção para arquivamento pode ser classificado em quatro categorias:

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

2.2.2 Definição de metadados

Metadados são usados para realizar uma documentação descritiva e técnica de um documento, pois são informações estruturadas que descrevem, localizam, gerenciam, recuperam e acessam recursos de informação (Corrêa, 2010) e são geralmente chamados de “dados sobre dados”. São, portanto, os metadados que auxiliam o acesso ao contexto de cada documento, sua proveniência e seu uso. A preservação da web pressupõe que existam padrões definidos de metadados para rastrear e garantir o acesso a esses documentos.

2.2.3 Captura de dados

Do ponto de vista da técnica, as ferramentas mais comuns de arquivamento da web usam os chamados web crawlers para automatizar o processo de coleta de páginas. Em linhas gerais, os web crawlers são capazes de visualizar páginas da web da mesma forma que os usuários as visualizam por meio dos navegadores, portanto, os web crawlers representam um método relativamente simples de coleta remota do conteúdo (Heaton, 2002). Exemplos de web crawlers usados para arquivamento da web incluem:

- a) Heritrix - rastreador da web projetado para arquivamento da web. Foi escrito pelo Internet Archive e está disponível sob uma licença de software livre e escrito em Java;

- b) HTTrack - é um web crawler livre e de código aberto e também navegador offline, desenvolvido por Xavier Roche, ele é licenciado sob a GNU General Public License. Permite o download de sites da World Wide Web da Internet para um computador local. Por padrão, o HTTrack organiza o site baixado pela sua estrutura original;
- c) Wget: uma ferramenta de linha de comando criada pelo GNU Project usada para baixar arquivos de vários servidores na Internet.

2.2.4 Controle de qualidade

Para que seja possível garantir a informação fidedigna oriunda de um website, faz-se necessário que haja um controle da qualidade da informação, isso perpassa a atenção ao design do site, links, possibilidades de reprodução de todos os textos, imagens e reprodutores de multimídia (som e vídeo). Para garantir a qualidade da captura de um web site, temas como acessibilidade a tecnologias do passado são um dificultador relevante, as tecnologias utilizadas quando da criação do site podem rapidamente se mostrar desatualizadas, o que gera problemas de compatibilidade com os navegadores mais recentes. Um exemplo são os websites baseados na linguagem de programação javascript, pois esta utiliza scripts executados do lado do usuário, o que pode carregar dados sem alteração do Uniform Resource Identifier (URI), em português, Identificador Uniforme de Recurso.

2.2.5 Armazenamento

Os critérios para o armazenamento do conteúdo capturado da web baseiam-se especialmente na tecnologia que será empregada para tal. O Internet Archive desenvolveu o formato ARC_IA já em 1996, combinando diferentes fontes da web em um arquivo agregado através de “web crawlers” (LIBRARY OF CONGRESS, 2024a). Posteriormente, a ISO 28500 definiu como extensão padrão para armazenamento de arquivos web o formato WebARChive (Warc), que foi desenvolvido pelo International Internet Preservation Consortium (IIPC) em 2009, esse método de arquivamento combina vários recursos digitais em um único arquivo, assim como informações relacionadas (LIBRARY OF CONGRESS, 2024b).

2.2.6 Acesso ao conteúdo

Todo o trabalho de arquivamento da web só fará sentido se for também disponibilizado para consultas futuras. O desenvolvimento de uma política de arquivamento da web deve também garantir que o acesso seja feito de modo simples pelos usuários, mas sem deixar de lado as garantias de autenticidade e clareza do contexto em que o website estava inserido no momento da captura. A política de acesso ao conteúdo arquivado deverá levar em consideração também questões legais relacionadas ao acervo, tais como a Lei Geral de Proteção de Dados, a proteção de direitos autorais, entre outras.

Ao longo dos anos, as instituições especializadas em arquivamento da web foram agentes importantes para o desenvolvimento dos parâmetros necessários para o desenvolvimento dessa atividade, ou seja, desde a primeira iniciativa de que se tem registro: o Internet Archive, criado em 1996, as demais instituições que foram surgindo com o intuito de salvaguardar o conteúdo da web foram cruciais para que o debate acerca do tema ganhasse cada vez mais relevância, além disso, essas iniciativas apresentaram importantes mecanismos, softwares, políticas, entre outras contribuições que ampliaram o debate sobre o arquivamento web, não o limitando ao campo acadêmico. Por exercerem este importante papel, é de extrema relevância que as principais iniciativas de arquivamento da web sejam analisadas com mais atenção no presente trabalho, a fim de que seus procedimentos fiquem mais claros e detalhados.

2.3 Iniciativas relevantes para o arquivamento da web

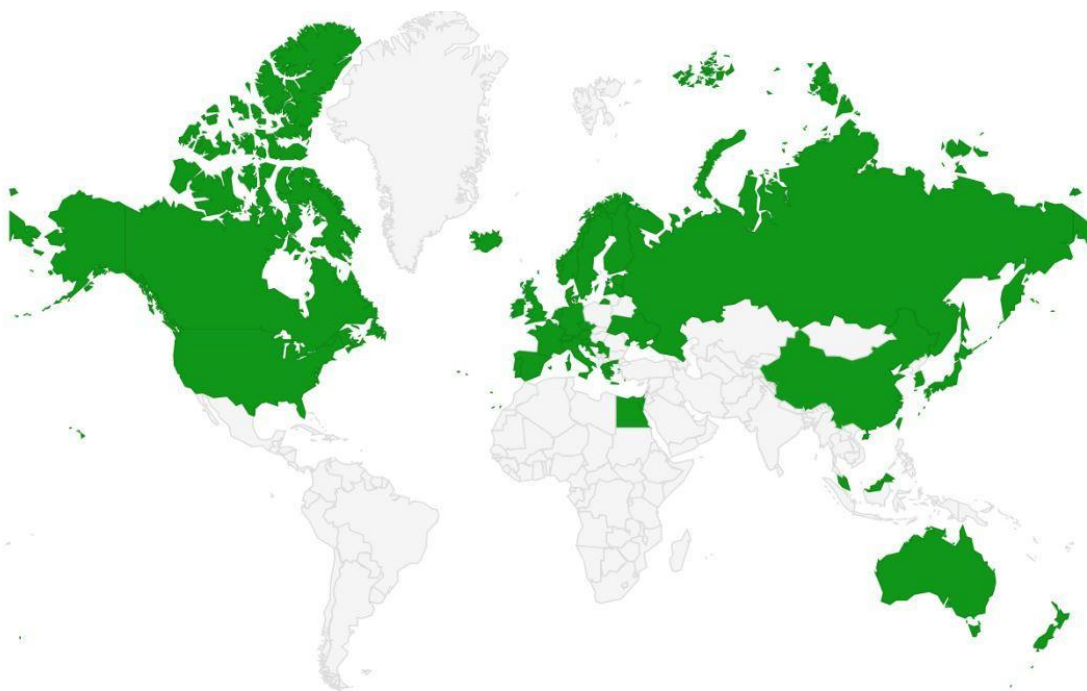
Diante da profusão de informações disponíveis na web, é natural que também existam iniciativas que visem à sua preservação. Segundo os autores Rockembach e Pavão (2018) há atualmente diferentes tipos de iniciativas que desenvolvem políticas próprias para a seleção dos conteúdos web a serem arquivados. Há aquelas que visam arquivar a web mundial como um todo, como é o caso do Internet Archive, iniciativa que será mais profundamente analisada a seguir no presente trabalho. Há também iniciativas que ocupam-se de grandes quantidades de websites, por exemplo, domínios nacionais ou de grandes regiões, essa política faz-se presente em trabalhos realizados por arquivos e bibliotecas

nacionais de diversos países. Por fim, há também casos de arquivamentos em menor escala, que procuram preservar sites institucionais ou temáticos e eventos específicos.

Desde o início da operação comercial da Internet, houve uma série de iniciativas que buscaram propor o arquivamento das páginas e mais páginas que eram criadas para compartilhamento de informações on-line. A seguir, serão apresentadas três das mais relevantes iniciativas já realizadas, a fim de que se construa um panorama de possibilidades para solucionar a questão do arquivamento web. Os projetos foram selecionados a partir de sua relevância e prevalência na bibliografia especializada consultada para a realização do presente trabalho. Atualmente, a organização de arquivamento da Internet com maior relevância é o Internet Archive, o qual emprega esforços com o intuito de arquivar toda a Internet. Bibliotecas e arquivos nacionais, consórcios e vários tipos de organizações também estão envolvidos no arquivamento de conteúdo culturalmente importante. Programas e serviços comerciais de arquivamento também estão disponíveis para organizações e empresas que necessitam arquivar seus próprios conteúdos para propósitos patrimoniais, regulatórios ou legais.

O mapa a seguir apresenta os locais onde estas iniciativas estão baseadas. Ele é fruto de uma pesquisa publicada em Portugal em 2011 e sua última atualização data de 2021 (Gomes; Miranda; Costa, 2011). É importante ressaltar que apesar do mapa abaixo não possuir nenhuma iniciativa demarcada na América Latina, o Chile faz sim parte do Consórcio Internacional de Preservação da Internet desde 2014, sendo o único país da região a ter atualmente um projeto em andamento para a preservação de conteúdos web locais, a inclusão do país latino-americano na lista de países participantes do Consórcio Internacional de Preservação da Internet pode ser verificada no site da Instituição, que apresenta um link para o “Archivo de la Web Chilena”, coleção sob a responsabilidade da Biblioteca Nacional do Chile.

Figura 2: Mapa de iniciativas de arquivamento da Web em todo o mundo



Fonte: GOMES, MIRANDA, COSTA (2011) e Wikipedia.org

Interessante frisar que o Chile, durante o mandato de Salvador Allende, desenvolveu um projeto intitulado “Project Cybersyn”,³ o qual valia-se de teorias do design informacional com o intuito de colocar em prática um planejamento econômico controlado e em tempo real. O projeto desenvolveu-se entre 1971 até 1973 e era composto por uma rede de máquinas de teletipo (máquina de escrever eletromecânica para transmissão de dados) que realizava a intercomunicação das fábricas do país a um centro de computação em Santiago, o qual controlava as máquinas empregando os princípios da cibernética. O principal arquiteto do sistema foi o cientista britânico Stafford Beer. Ao lado de uma equipe de programadores chilenos e consultores britânicos, Beer criou o software do projeto, batizado de Cyberstride, o qual utilizava filtros e controle bayesiano, ou seja, era baseado em métodos estatísticos para classificar documentos por categorias. O projeto teve fim após o golpe de Estado de 11 de setembro de 1973, quando militares chilenos, ajudados pela CIA e pelo governo dos Estados Unidos, derrubaram o regime democrático constitucional do Chile e de seu presidente, Salvador Allende.

³ PROJECT CYBERSYN. In: WIKIPEDIA. 2024. Disponível em: https://en.wikipedia.org/w/index.php?title=Project_Cybersyn&oldid=1237926741. Acesso em: 10 ago. 2024.

Chama a atenção que o continente latino-americano, que tem um país de dimensões continentais como o Brasil, além do Chile, nação capaz de colocar em prática um projeto do nível do Cybersyn ainda nos anos de 1970, esteja ausente da lista de locais engajados no arquivamento da web, seja ao não ter larga participação no Consórcio Internacional de Preservação da Internet, ou ao não desenvolver iniciativas locais voltadas para a atividade. Este é um dado que demonstra a fragilidade e falta de amparo na América Latina no que se refere à preservação da memória no ambiente da web, o que pode ser notado ao analisarmos as iniciativas mais relevantes de preservação da web no mundo.

2.3.1 Internet Archive

O Internet Archive foi fundado por Brewster Kahle, um cientista da computação, bibliotecário digital e empresário, em maio de 1996. Segundo seu fundador, a motivação inicial para a criação do projeto foi a percepção da efemeridade da informação na era digital, assim surgiu a ideia de criar um arquivo digital que pudesse armazenar e preservar o conteúdo da web. O Internet Archive é uma organização sem fins lucrativos que desde a sua fundação dedica-se à preservação e disponibilização de conteúdo digital para o público em geral. Atualmente, é uma das iniciativas mais importantes no campo da preservação da mídia digital em todo o mundo, disponibilizando ao público não só um arquivo permanente da web, mas também de outras formas de mídia digital, o acervo disponibiliza atualmente:

- a) 835 bilhões de páginas da web;
- b) 44 milhões de livros e textos;
- c) 15 milhões de gravações de áudio (incluindo 255.000 shows ao vivo);
- d) 10,6 milhões de vídeos (incluindo 2,6 milhões de programas de notícias de televisão);
- e) 4,8 milhões de imagens;
- f) 1 milhão de programas de software. (Internet Archive, 2024)

Como exposto acima, a coleção voltada para a preservação do conteúdo digital possui atualmente mais de 800 bilhões de prints de páginas da web, que são

disponibilizados aos usuários através de uma ferramenta chamada Wayback Machine. O funcionamento da Wayback Machine é baseado em um processo de arquivamento conhecido como "crawling" (rastreamento). Por meio de bots automatizados, a Wayback Machine visita regularmente sites da web e faz cópias de suas páginas, armazenando essas cópias em seu banco de dados. Quando os usuários acessam a Wayback Machine, eles podem inserir a URL de um site e visualizar versões anteriores dessas páginas.

Por tratar-se de uma iniciativa que lida com informações em larga escala, o Internet Archive possui alguns desafios importantes que estão diretamente relacionados ao fato da composição do corpus de seu acervo ser realizada de um modo virtual. Pode-se salientar especialmente questões legais e de direitos autorais que ocorrem com a digitalização e distribuição de conteúdo protegido por direitos autorais, legalmente não é sempre possível ao Internet Archive se certificar de que o conteúdo capturado nas redes não está sob qualquer restrição de patente ou salvaguardado por leis de reprodução, lembrando que o Internet Archive captura conteúdo do mundo todo, assim, existe sempre o risco de algo que a ferramenta capture esteja sob proteção legal, o que leva a pedidos de conteúdos serem retirados do ar.

Atualmente, o Internet Archive está respondendo a um processo nos Estados Unidos após um grupo de editoras acusar o site de uso indevido de obras sob a lei de direitos autorais ao realizar o empréstimo digital controlado de seus livros físicos que foram digitalizados para empréstimo virtual.⁴ O processo obrigou o Internet Archive a eliminar da sua plataforma mais de 500 mil livros em junho de 2024. Felizmente, o conteúdo do WayBack Machine – ou seja, o acervo da web – não foi afetado por este processo, o qual ainda está em fase de apelação.

Outro importante desafio do Internet Archive é garantir a preservação de seu conteúdo a longo prazo, pois manter a integridade e acessibilidade do acervo digital representa um desafio contínuo, uma vez que existe uma constante evolução tecnológica no que tange aos formatos de distribuição de informação na Internet.

⁴ PUBLISHNEWS. **Internet Archive retira cerca de 500 mil livros do seu arquivo virtual após processos de editoras nos EUA**, 2024. Disponível em: <https://www.publishnews.com.br/materias/2024/07/01/internet-archive-retira-cerca-de-500-mil-livros-do-seu-arquivo-virtual-apos-processos-de-editoras-nos-eua>. Acesso em: 23 jul. 2024.

Este tema é especialmente sensível em uma organização sem fins lucrativos, como é o caso do Internet Archive, o qual depende de doações e financiamento externo para operar. Logo, garantir recursos adequados para sustentar suas operações de longo prazo é uma preocupação constante, assim como acompanhar a tecnologia que é constantemente transformada na Internet pelos principais players do mercado.

Há ainda outra questão sensível na maneira como o Internet Archive opera, pois este tem um esforço constante para ser uma plataforma neutra e imparcial, porém, com o seu modelo de armazenamento em massa e o amplo acesso disponibilizado a qualquer usuário ao seu conteúdo digital, é natural que este enfrente pressões para remover ou restringir certos tipos de conteúdo controverso. Mesmo tendo uma equipe de moderadores do conteúdo, existe um grande desafio para o Internet Archive no que tange a esse tópico, o qual está diretamente ligado a questões complexas como censura e silenciamento de minorias.

Apesar de todos os desafios que se impõem ao Internet Archive, este é atualmente visto como a maior fonte de informação sobre o passado da web no mundo. Entretanto, os autores Thelwall; Vaughan, (2004) destacaram que existe um desequilíbrio entre diferentes países no Internet Archive, embora de modo não intencional, há evidência de que países pobres são sub-representados nas coletas do Internet Archive, o que demonstra ser um problema para a construção de um arquivo histórico eficaz da web.

2.3.2 Consórcio Internacional de Preservação da Internet

O Consórcio Internacional de Preservação da Internet (International Internet Preservation Consortium - IIPC) é uma iniciativa colaborativa que agrega instituições de todo o mundo dedicadas à preservação do conteúdo da web. Fundado em 2003 a partir de debates acerca do tema iniciados em 2001, o IIPC foi formalmente licenciado pela Biblioteca Nacional da França com 12 instituições participantes. Na época, os membros concordaram conjuntamente em financiar e participar em projetos e debates para cumprir as metas sugeridas em grupos de trabalho. O acordo inicial esteve em vigor por três anos, e a adesão era limitada a instituições fundadoras. O IIPC agora está aberto a bibliotecas, arquivos, museus e instituições de patrimônio cultural em todos os lugares e recebe pedidos de adesão. Os

membros do IIPC atualmente são organizações de mais de 35 países, incluindo bibliotecas e arquivos nacionais, universitários e regionais.

O IIPC opera como uma rede colaborativa de instituições comprometidas com a preservação do conteúdo da web. Seus objetivos principais incluem promover a colaboração internacional na preservação digital, desenvolver e compartilhar melhores práticas e padrões, e facilitar o acesso público ao conteúdo arquivado da web. O IIPC também desempenha um papel ativo na definição de estratégias de preservação, advocacia e sensibilização para questões relacionadas à preservação digital. Algumas das principais maneiras pelas quais o IIPC trabalha para preservar o patrimônio digital incluem a captura de websites e páginas da web, utilizando tecnologias de rastreamento e arquivamento para capturar o conteúdo online. Isso inclui a captura de páginas estáticas, bem como conteúdo dinâmico, como vídeos, áudios e interações de usuário. O IIPC também trabalha no desenvolvimento e aprimoramento de ferramentas e tecnologias para a captura, preservação e acesso a conteúdo da web, como software de arquivamento, sistemas de gerenciamento de coleções digitais e ferramentas de análise e visualização de dados (IIPC, 2024).

Um dos principais projetos desenvolvidos pelo IIPC é o Open Wayback, uma versão de código aberto da Wayback Machine do Internet Archive, desenvolvida em parceria com o Internet Archive, essa aplicação foi lançada em setembro de 2005, com o intuito de permitir a distribuição pública do aplicativo e aumentar sua capacidade de manutenção e extensibilidade. O IIPC disponibiliza também em seu site um GitHub, uma plataforma de hospedagem de código-fonte e arquivos com controle de versão que permite a programadores ou qualquer usuário cadastrado fazer contribuições ao projeto, além de garantir que as instituições que utilizam a Open WayBack Machine troquem informações e tenham acesso ao código para adaptações.

O IIPC também divulga uma lista parcial, com algumas das instituições que atualmente adotam o Open Wayback, ou que o utilizam de modo customizado. Ao se observar a lista, é perceptível o grande alcance dessa ferramenta nos Estados Unidos e na Europa, a tabela contém instituições como a Biblioteca Nacional da França e a Biblioteca do Congresso nos EUA como usuárias do OpenWayback. Assim como no mapa apresentado anteriormente, que continha as iniciativas de arquivamento da web no mundo, a presença mais relevante está novamente em países do Norte Global.

Quadro 4: Instituições adeptas do OpenWayback

Instituição	Tipo de Acesso	URL
National and University Library of Iceland	Público	http://vefsafn.is
The British Library	Público	http://www.webarchive.org.uk/
Stanford Web Archive Portal	Público	https://swap.stanford.edu/
The Library of Congress	Quase todo público, parte restrito on-site	Collections accessed via http://www.loc.gov/websites/collections/
Bibliotheca Alexandrina	Público	http://archive.bibalex.org
York University Digital Library	Público	http://digital.library.yorku.ca/wayback/
Bibliothèque Nationale de France	On-site	http://www.bnf.fr/fr/collections_et_services/livre_presse_medias/a_archives_internet.html
University of North Texas Libraries	Parte público, parte on-site ou com credenciais da universidade	Indirectly via http://digital.library.unt.edu/search/?fq=dc_type%3Awebsite
National Library of the Czech Republic	Público ou on-site	http://webarchiv.cz
National Library of Australia	Público	http://webarchive.nla.gov.au/
National Library of Austria	<u>Público (apenas para URLs http://www.onb.ac.at) e on-site</u>	https://webarchiv.onb.ac.at
National Széchényi Library, Hungary	Público	http://webadmin.oszk.hu/owb/wayback/

Fonte: <https://github.com/iipc/openwayback/wiki/OpenWayback-Users>. Adaptação própria.

2.3.3 O Arquivo.pt

O Arquivo da Web Portuguesa (Arquivo.pt) é uma iniciativa desenvolvida em Portugal com o objetivo de preservar o patrimônio digital do país. Lançado oficialmente em 2008 pela Fundação para a Ciência e Tecnologia (FCT) de Portugal, o projeto teve as suas origens já em 2001, como uma iniciativa para preservar o conteúdo da web relacionado com Portugal, inicialmente foi idealizado pelo projeto científico TUMBA! (sigla para Temos Um Motor de Busca Alternativo), de um grupo da Faculdade de Ciências da Universidade de Lisboa, posteriormente, já em 2007, o arquivo da Web Portuguesa passou o seu desenvolvimento para a FCCN, a Unidade de Computação Científica da FCT, que tem como propósito contribuir para o desenvolvimento da Ciência, Tecnologia e Conhecimento em Portugal. Desde então, o Arquivo.pt cresceu e se tornou uma das principais fontes de informação histórica online para pesquisadores, académicos e o público em geral. O projeto foi motivado pela necessidade de preservar o patrimônio digital do país, garantindo o acesso futuro a conteúdos importantes da web.

Segundo o site da instituição, Arquivo.pt tem páginas disponíveis para consulta desde o ano de 1996. Até 2007, o conteúdo foi adquirido principalmente a partir do Internet Archive, após essa data, o Arquivo.pt passou a fazer a sua própria seleção para arquivamento da web, por isso, o acervo após 2007 é totalmente focado na web portuguesa. Os serviços de pesquisa disponibilizados pelo Arquivo.pt incluem pesquisa de texto completo, pesquisa de imagens, listagem de histórico de versões, pesquisa avançada e interfaces de programação de aplicações (em inglês application programming interfaces – API). De acordo com Gomes (2022) a estrutura de hardware do Arquivo.pt é mantida em um data center próprio e gerida por profissionais dedicados, que além de garantir a aplicação da política de arquivamento da web da instituição, também publicam artigos técnicos e científicos relacionados ao tema sob o regime de dados abertos. Toda essa infraestrutura é a mantenedora de 145 coleções, as quais contêm mais de 13 milhões de arquivos da web advindos de mais de 28 milhões de websites.

O principal objetivo do Arquivo.pt é ser o responsável pelo arquivamento de toda publicação da web relativa a Portugal, para atingir esse objetivo audacioso, o

time de tecnologia usa diferentes tipos de crawlers para captura da informação na web. Essas capturas formam diferentes tipos de coleções conforme a seguir:

- a) rastreamentos diários: uma seleção de publicações online nacionais é coletada todos os dias usando uma combinação de rastreadores convencionais e baseados em navegador;
- b) crawls mensais: uma seleção de publicações online nacionais e internacionais, recolhida todos os meses através de crawlers baseados em navegador;
- c) crawls trimestrais: listas de todos os domínios registados no domínio “.pt” são recolhidas a cada 3 meses através de crawler convencional devido ao grande volume de dados envolvidos;
- d) rastreamentos de alta qualidade: uma seleção de sites que foram arquivados e curados iterativamente, com a mais alta qualidade possível, usando a melhor combinação de tecnologias disponíveis;
- e) rastreamentos Save Page Now: páginas web arquivadas pelos usuários em alta qualidade usando o serviço “savepagenow”;
- f) rastreamentos de páginas completas: coleções de recursos web arquivados através da invocação colaborativa do serviço Página Completa pelos usuários ou através de sua execução automática em lote para corrigir listas selecionadas de páginas web;
- g) coleções doadas: coleções doadas ou adquiridas de conteúdos históricos da web integradas no Arquivo.pt (ex.: coleção Geocities). O conteúdo da web é recebido em formatos variados e convertido no formato WARC padrão;
- h) rastreamentos especiais: seleções de páginas sobre um determinado tópico, rastreado com frequência e metodologias variadas. (Gomes, 2022, p. 56–57, tradução e adaptação próprias).

A robusta infraestrutura do Arquivo.pt faz dessa iniciativa uma fonte relevante de informação da web não apenas para o seu país, mas também para todos aqueles interessados no conteúdo de língua portuguesa, uma vez que as relações entre falantes lusófonos estão muito mais estreitas no ambiente da web. Assim, a memória social digital dos falantes de português pode se beneficiar das coleções desenvolvidas no Arquivo.pt, exemplo disso é a coleção desenvolvida durante a epidemia de COVID-19 da instituição, iniciada a partir de fevereiro de 2020, após o

International Internet Preservation Consortium (IIPC), propor aos associados a realização de uma coleção dedicada ao tema. O Arquivo.pt contribuiu com 1.237 endereços selecionados, principalmente em língua portuguesa, tendo realizado 3 inclusões na coleção do IIPC, a primeira a 23 de março, a segunda a 15 de junho e a terceira no final de agosto de 2020, foi acrescentando ao seu acervo conteúdos internacionais que podem auxiliar pesquisas de todo o mundo. Segundo o Arquivo.pt, foram utilizados diversos tipos de software para recolher as páginas. Por exemplo, para as recolhas diárias sobre 106 sites utilizou-se o Heritrix. Para a captura de 108 páginas do Twitter optou-se pelo Brozzler e para os vídeos a captura manual com o Webrecorder e Browsertrix. Essa captura de sites contemplou páginas de diversos países de língua portuguesa, não apenas Portugal, mas também Brasil, Angola e Moçambique, na tabela a seguir, é possível verificar exemplos de sites capturados na iniciativa que agora fazem parte da coleção

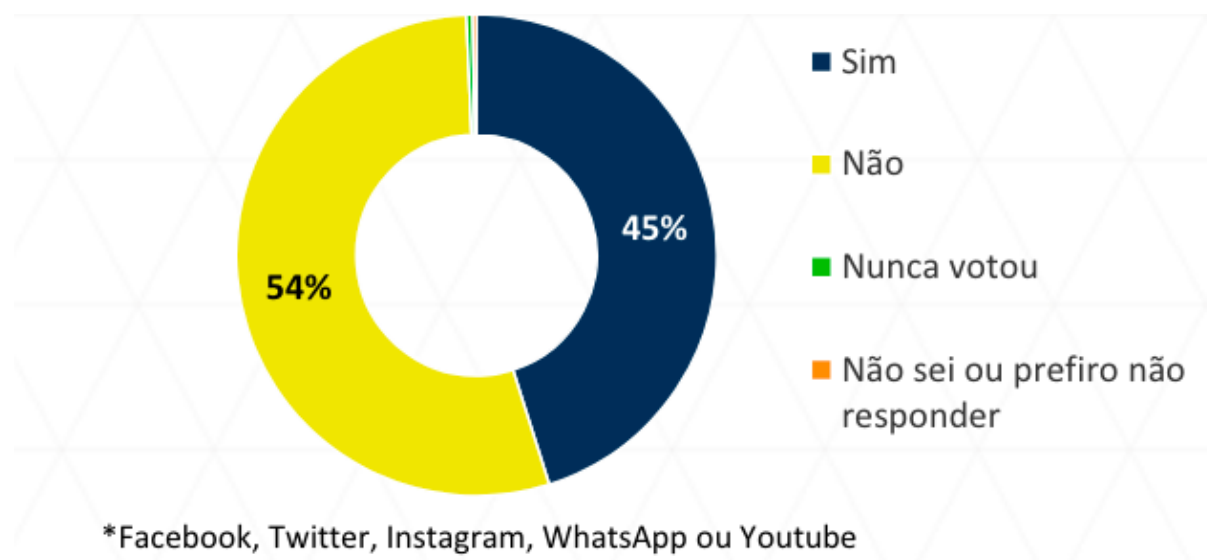
Figura 3: Contributo do Arquivo.pt para a 1ª recolha internacional (1.235 endereços)
– 20 de fevereiro

Title (of specific article)	Language	Description or Keywords	Scope 1=full seed host or directory 2=Crawl one page only 3=seed page plus 1 click all links on seed page	Additional notes	Position of the result or Relevance
News - Primeiro diário ca	Portuguese	Cape Vert , news	2		1
Página principal do Porta	Portuguese	Brazil, news	2		3
Navio de Singapura atrac	Portuguese	Brazil, news	2		9
Coronavírus: voos são vi	Portuguese	Brazil, news	2		1
China construirá hospital	Portuguese	Brazil, news	2		2
Coronavírus faz primeira	Portuguese	Brazil, news	2		10
Coronavírus assusta e m	Portuguese	Brazil, news	2		4
Coronavírus já matou 26	Portuguese	Brazil, news	2		5
Sintomas, transmissão e	Portuguese	Brazil, news	2		8
Coronavírus: nova metod	Portuguese	Brazil, news	2		6
Estudantes cabo-verdian	Portuguese	Cape Vert	2		10
Cagle Cartoons, Inc. - Se	English	social aspects, comics	3		2
Coronavírus arma biológi	Portuguese	social aspects, religion	2		8
Coronavírus transfere dis	Portuguese	Angola, news	2		4
Coronavírus assusta bols	Portuguese	Angola, news	2		5
Coronavírus. Balanço sol	Portuguese	Angola, news	2		2
Angola cria posto contra	Portuguese	Angola, news	2		6
Chinês sem coronavírus	Portuguese	Angola, news	2		3
País regista primeiro cas	Portuguese	Angola, news	2		1
Jornal Economia & Finan	Portuguese	Angola, news	2		2
Jornal dos Desportos - C	Portuguese	Angola, news	2		1
O Coronavírus e seus efe	Portuguese	Angola, news	2		1
Novo Jornal Notícias de	Portuguese	news,	3		1

Fonte: Arquivo.pt, Coleção especial acerca da Covid-19. Destaque próprio. Disponível em <https://sobre.arquivo.pt/pt/colecao-especial-acerca-do-covid-19/>

Os três exemplos de iniciativas aqui apresentados demonstram como o arquivamento da web é uma atividade rica em oportunidades para os países que se propõem a realizá-la, além disso, os frutos que se podem a partir do desenvolvimento de iniciativas voltadas para a preservação do conteúdo da web são inúmeros, basta pensar na relevância das redes sociais nos últimos resultados das eleições no Brasil, segundo uma pesquisa de opinião do Instituto DataSenado quase metade dos entrevistados (45%) afirmaram ter decidido o voto levando em consideração informações vistas em alguma rede social. A pesquisa foi publicada em 2019 e entrevistou 2,4 mil cidadãos com acesso à Internet, em todas as unidades federativas do país.

Gráfico 3: Ao decidir seu voto durante um período de eleições, você já levou em consideração informações vistas em alguma rede social*?



Fonte: Senado Federal, Instituto de Pesquisa DataSenado

<https://www2.senado.gov.br/bdsf/handle/id/603465>

Os números dessa pesquisa parecem apontar para a conclusão de que o Brasil tem um número alto de usuários de Internet e de redes sociais e que é alto o impacto dessas redes na opinião pública. Porém, a análise do presente trabalho sobre as iniciativas de preservação da web demonstra pouca expressividade do Brasil na preservação do conhecimento produzido e publicado na web, há muita

informação circulando e gerando debates importantes, porém, sem preservação nenhum conhecimento será relevante a longo prazo, o que demonstra um risco para o debate público em nosso país. Para melhor entender este cenário, será preciso avaliar com mais profundidade como o tema do arquivamento da web surgiu no país, desde a introdução da Internet comercial no Brasil, até a consolidação do tema no ambiente institucional e acadêmico.

2.4 Internet e a Web no Brasil

A introdução da Internet no Brasil deu-se em 18 de julho de 1989, operada pelo Instituto Brasileiro de Análises Sociais e Econômicas (IBASE) no Rio de Janeiro. Desde então, a Internet desenvolveu-se no país, assim como as Tecnologias de Informação e Comunicação (TICs), em especial a web. Foi a partir do apoio do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) que se formou a Rede Nacional de Ensino e Pesquisa (RNP), que conectou instituições acadêmicas e de pesquisa no território nacional. Ao longo dos anos de 1990, as instituições de ensino superior do país foram as primeiras a usufruir da nova tecnologia, mas já em 1995, o Brasil pode se conectar à web (Rockembach; Pavão, 2024). No âmbito governamental, o Ministério da Comunicação e da Ciência e Tecnologia anunciou em 1995 a criação do Comitê Gestor da Internet no Brasil (CGI.br), o qual foi devidamente estruturado em 2003, após a publicação do Decreto Federal n. 4.829. Foi também criado o Núcleo de Informação e Coordenação do Ponto BR (NIC.br), responsável pela implementação das decisões do CGI.br no que tange à coordenação da Internet no Brasil. O NIC.br é também responsável pelo registro e manutenção de nomes de domínios “.br”.

Os autores Rockembach e Pavão destacam o nome de Demi Getschko como um dos principais responsáveis pela disseminação da Internet no Brasil. Nascido na Itália, mas de nacionalidade brasileira, Getschko atuou como engenheiro eletricista e ocupou o cargo de diretor-presidente do Núcleo de Informação e a Coordenação do Ponto BR (NIC.br) a partir de 2005. Sua carreira inclui passagens pelo Centro de Computação Eletrônica (CCE) da USP e pelo Centro de Processamento de Dados da Fapesp. Nesse período, foi coordenador de operações da RNP e participou do esforço da implantação de redes no Brasil. Até maio de 2009, Getschko também

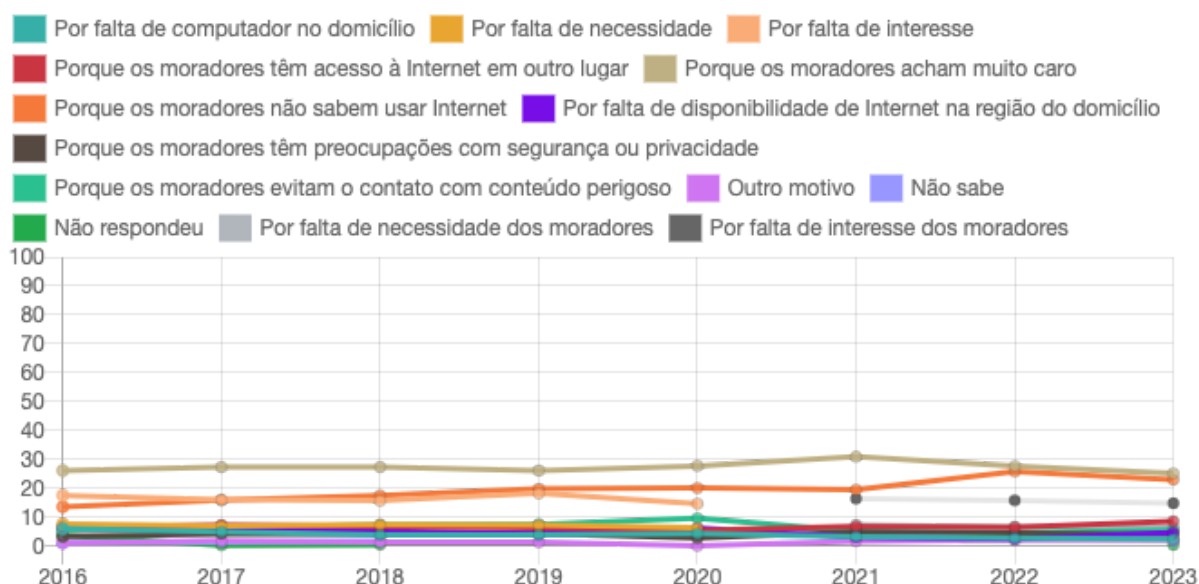
atuou como membro da diretoria da Internet Corporation for Assigned Names and Numbers (ICANN) pela ccNSO (Country Code Names Support Organization).

Outro importante nome para a Internet no Brasil é o do físico nuclear Oscar Sala. Nascido em Milão, Sala graduou-se em física em 1943 pela Universidade de São Paulo, no então Departamento de Física da Faculdade de Filosofia Ciências e Letras, em sua pesquisa dedicou-se inicialmente ao estudo da radiação cósmica. Em 1946, recebeu uma bolsa de estudos e foi aos Estados Unidos, onde participou no desenvolvimento de aceleradores eletrostáticos usados no estudo de física nuclear, o primeiro equipamento a usar pulsos de raios para o estudo de reações nucleares com nêutrons rápidos. Ao longo dos anos, teve destaque como cientista e professor do departamento de Física da Universidade de São Paulo, até que, em 1988, Sala criou o projeto Rede ANSP (An Academic Network In São Paulo), um projeto que buscou atender à solicitação de interconexão das redes das três universidades estaduais paulistas, USP, Unesp e Unicamp, e do Instituto de Pesquisas Tecnológicas à Fapesp. Ao lado de Demi Getschko, um dos responsáveis pela primeira conexão TCP/IP brasileira, em 1991, entre a Fapesp e a Energy Sciences Network (ESNet), nos Estados Unidos, por meio do Fermilab (Fermi National Accelerator Laboratory) Sala fez parte do projeto que passou a adotar o protocolo TCP/IP, tornando a Rede ANSP o único ponto de conexão acadêmica ou comercial à Internet entre 1992 e 1994.

A relevância de nomes como Demi Getschko e de Oscar Sala para a Internet em âmbito mundial é um índice de como o Brasil tem destaque não apenas como mercado, mas também como *player* importante para a tecnologia global. Desde a sua introdução, a Internet vê sua relevância crescendo entre todas as camadas sociais, porém, como outras tecnologias introduzidas em nossa sociedade, a Internet também carrega consigo marcas das desigualdades sociais e de acesso a bens culturais tão comuns na realidade brasileira. O Núcleo de Informação e Coordenação do Ponto BR (NIC.br), braço executivo do Comitê Gestor da Internet no Brasil (CGI.br) publicou um estudo revelador sobre o estado da Internet no Brasil. Conduzida pelo Centro Regional de Estudos para o Desenvolvimento da Sociedade da Informação (Cetic.br/NIC.br), a publicação intitulada "Conectividade Significativa: propostas para medição e o retrato da população no Brasil" foi lançada em abril de 2024 e traz um dados relevantes para melhor entendimento de como a população

utiliza a Internet. No gráfico abaixo, é possível verificar qual o principal motivo para que os domicílios brasileiros não tenham acesso à Internet:

Gráfico 4: Domicílios sem Acesso à Internet, por Principal Motivo para a Falta de Internet:



Fonte: CGI.br/NIC.br, Centro Regional de Estudos para o Desenvolvimento da Sociedade da Informação (Cetic.br), Pesquisa sobre o Uso das Tecnologias de Informação e Comunicação nos domicílios brasileiros.

Segundo os dados obtidos na pesquisa, os domicílios brasileiros não possuem acesso à Internet principalmente devido ao alto custo do pacote de dados para conexão, e em segundo lugar porque os moradores não sabem usar a Internet. Esses dois dados, que são as principais justificativas para que um domicílio não tenha acesso à Internet desde 2018, contrastam com outro dado importante revelado na pesquisa: 84% das pessoas no Brasil são usuárias de Internet, o que equivale a mais de 156 milhões de indivíduos. Obviamente, depois da introdução em larga escala dos smartphones, o uso da Internet no próprio domicílio não significa que o usuário está apartado da tecnologia, porém, denota que ainda existem barreiras sociais para que se usufrua da Internet em sua totalidade. O acesso universal à Internet perpassa outras condições que não apenas a conectividade, é necessário também que a população de usuários tenha letramento

digital, segurança e confiabilidade no conteúdo acessado, proteção de dados sensíveis e sigilosos, além de, é claro, a certeza de que o conteúdo acessado e compartilhado hoje será salvaguardado para as próximas gerações. É nesse momento que o arquivamento da web revela um dos seus mais importantes aspectos.

3. Arquivamento da web brasileira

A ideia de que o conteúdo informacional nato-digital é importante para a constituição do patrimônio brasileiro é uma discussão relativamente recente no país. As experiências dos usuários nas ferramentas de publicação na web cresceram em uma medida que não foi acompanhada pelas instituições de memória no Brasil. Porém, não significa que o debate não tenha existido. Neste sentido, destaca-se no ano de 2009, a criação do Fórum de Cultura Digital, uma iniciativa do Ministério da Cultura (Minc) em cooperação com a Rede Nacional de Pesquisa (RNP). Esta importante iniciativa do governo federal teve suas discussões iniciadas ainda na gestão do ministro Gilberto Gil, mas só foi implementado quando Juca Ferreira, assumiu a pasta em 2008 é importante aqui destacar que até então as questões relacionadas ao digital tinham sido encaradas como instâncias políticas ligadas à Ciência, à Tecnologia e à Comunicação. Segundo Dantas (2014) o Fórum desempenhou um papel relevante na discussão do Marco Civil da Internet, produziu encontros presenciais como o Seminário Internacional do Fórum da Cultura Digital Brasileira (novembro de 2009) e o II Fórum da Cultura Digital Brasileira (novembro de 2010), realizados na Cinemateca de São Paulo.

O Fórum era composto por 239 grupos em julho de 2011, os quais debatiam temas variados, tais como comunicação eletrônica, cibercultura, música na web, mídia livre entre outros. Até julho de 2012, quando foi transformado numa “plataforma pública de blogs”, o Fórum chegou a ter cerca de 400 grupos, 500 fóruns e mais de 7000 participantes. Um dos eixos temáticos abordados no Fórum de Cultura Digital era o da Memória Digital, conforme Dantas (2014) os debates desse eixo tinham como foco a digitalização de acervos, dos aspectos técnicos da digitalização até os mecanismos necessários para que o acesso ao patrimônio digital fosse fomentado. A preocupação com a digitalização dos acervos é também o ponto central do documento de base para o Eixo Memória Digital, lançado em novembro de 2009. O que é possível notar ao se debruçar sobre a iniciativa do Fórum de Cultura Digital é que na época ainda não havia no Brasil uma discussão madura sobre a web e sua preservação, o debate limitava-se aos processos de digitalização de acervos físicos e sua manutenção.

Ao longo dos anos, o tema do arquivamento web foi ganhando força no debate público, especialmente acadêmico, porém, é possível perceber que o tópico

não teve o mesmo desenvolvimento em ações e políticas claras por parte das principais instituições de memória do país. Segundo Melo e Rockembach (2021) há no país uma série de leis que apontam a responsabilidade do Estado no que diz respeito à gestão, acesso e preservação de seus documentos:

(...) existem outras dezenas de Decretos, Leis, Medidas Provisórias, Resoluções, Portarias e Instruções Normativas que deliberam a respeito do tratamento de documentos no âmbito da administração pública. Trata-se, portanto, de compreender essas legislações e discutir o direito ao acesso às informações de custódia do Estado e o quanto esse direito deve ser assegurado pelos governos. (Melo; Rockembach, 2021)

Cabe às instituições de memória fazer juz à legislação vigente e desenvolver uma série de políticas pertinentes para o arquivamento da web, não apenas no nível institucional, mas também que sirvam de apoio para coleções da web brasileira. Melo (2020) fez um trabalho de pesquisa acerca do Arquivamento dos websites do Governo Federal Brasileiro, o qual trouxe resultados promissores para a preservação da memória institucional brasileira, segundo o autor, uma pesquisa realizada pelo Núcleo de Pesquisa em Arquivamento da Web e Preservação Digital (Nuaweb, da Universidade Federal do Rio Grande do Sul) teve como objetivo investigar as possibilidades de de arquivamento de sites do Governo Federal brasileiro tendo como objeto de análise 23 sites governamentais. A pesquisa verificou os recursos oferecidos por esses sites; arquivou os sites selecionados, utilizando o Heritrix como rastreador de páginas; reconstituiu sites arquivados usando o software WABAC; e comparou os recursos disponíveis nas versões ao vivo e arquivadas dos sites selecionados. A conclusão da pesquisa foi de que os sites do Governo Federal Brasileiro são arquiváveis sem perda de informações relevantes.

De acordo com Rochemback e Pavão (2024), a necessidade de elaborar políticas de seleção com foco específico, seja ele institucional, temático ou por domínio, acompanhada pela gestão do ciclo de vida do arquivamento da web e pela adoção de tecnologias de código aberto, permitirá não apenas a preservação da memória digital brasileira, mas também contribuirá para a comunidade internacional de arquivamento da web. Para melhor entender como o arquivamento da web brasileira pode ser desenvolvido, o presente trabalho já explorou três iniciativas internacionais relevantes para o arquivamento da web, demonstrando assim os

requisitos mínimos para que o arquivamento da web seja colocado em prática. Agora é importante explicar também a diferença entre o arquivamento de coleções na web e o arquivamento web no âmbito institucional e governamental. Tal diferenciação tem como objetivo dar um panorama de como o Brasil tem lidado com esses tipos de arquivamento, e quais as possibilidades levando em consideração como o Brasil já atua para o desenvolvimento de ambos os tipos de trabalho de arquivamento da web.

Apontar essas diferenças, entre informações para a composição de coleções e informações institucionais e governamentais é importante para o presente trabalho, pois a atuação das instituições de memória nesses diferentes tipos de informação pode e deve ser diferente. Há bastante relevância na produção de informações na web em meios que não são chancelados por instituições ou pelo próprio governo do país, tome-se como exemplo as redes sociais, que atualmente são usadas por profissionais de renome e políticos importantes, como Deputados e Senadores, até mesmo o Presidente da República faz uso assíduo de redes sociais, porém, a rede social não necessariamente faz parte da comunicação oficial de uma instituição ou do governo de uma localidade. Por isso, a distinção se faz necessária, há a necessidade de se preservar a memória da web brasileira no que tange ao conteúdo publicado nas chamadas páginas oficiais, no entanto, as coleções levam em consideração também a informação produzida de modo mais informal, em microblogs como o X (antigo Twitter), redes sociais, entre outros.

O autor Bucci (2019) lembra da dimensão cultural da informação ao tratar do Código de Ética e Deontologia do Bibliotecário Brasileiro, a web está evidentemente implicada na cultura do nosso tempo, por isso é cada vez mais protagonista da construção de uma memória social. Levar em consideração a produção de informação em um meio popular como a web pode significar ampliar o debate público sobre assuntos relevantes, mas essa ampliação deve obviamente vir acompanhada de um viés crítico próprio do profissional de memória capacitado a entender a relevância de uma informação, seja dentro de uma coleção, seja a partir da publicação oficial de instituições ou governos.

3.1 Desenvolvimento de coleções da web

O desenvolvimento de coleções consiste no trabalho de selecionar todo material que seja relevante para uma determinada temática, para Vergueiro (1993) o desenvolvimento de coleções na Biblioteconomia tem origem na explosão bibliográfica ocorrida no século XX que acarretou a necessidade da biblioteca cada vez mais praticar uma seleção de acervo mais criteriosa que beneficie os seus usuários, segundo o autor, o controle do crescimento ou desenvolvimento das coleções em bibliotecas deu-se para que estas não ficassem abarrotadas de materiais com informações redundantes. A Biblioteconomia sempre pensou e debateu a seleção de materiais, afinal a busca da informação correta ao usuário está no cerne da disciplina, mas o desenvolvimento de coleções é um estudo que pode ser observado com mais ênfase a partir dos anos de 1980.

Vergueiro (1993) apresenta um panorama da teoria acerca do desenvolvimento de coleções, primeiramente ao citar Hendrik Edelman e a sugestão do autor de que existe uma hierarquia entre os termos desenvolvimento de coleções, seleção e aquisição: no primeiro nível, estaria o desenvolvimento de coleções entendido de maneira ativa, como uma função de planejamento; no segundo nível, estaria a seleção como função direta do desenvolvimento de coleções, ou seja, o processo de tomada de decisões relacionadas com a implementação dos objetivos anteriormente estabelecidos; e no terceiro nível desta hierarquia, estaria a aquisição, entendida como a implementação das decisões de seleção.

Mais recentemente, Achilles (2014) fez análise das teorias sobre o tema, apontando para os estudos do autor Edward G. Evans, os quais seguem uma linha denominada sistêmica, a qual atribui ao desenvolvimento de coleções um caráter de atividade rotineira na biblioteca. Achilles também destaca o trabalho de Simone Weitzel, autora da analogia do guarda-chuva, na qual o desenvolvimento de coleções é similar a um guarda-chuva composto por varetas, as varetas seriam as seis etapas do processo de composição de uma coleção: estudo da comunidade; política de seleção; seleção; aquisição; desbaste, incluindo descarte; e avaliação da coleção (Weitzel, 2013). Segundo Achilles (2014), a analogia de Weitzel retoma os conceitos já definidos anteriormente pelos autores Edward G. Evans e Waldomiro Vergueiro, porém, em 2013 a autora ampliou o número de “varetas do guarda-chuva” para nove itens, sendo eles: estudo da comunidade; política de

seleção; seleção; aquisição; política de aquisição; desbaste, incluindo o descarte; política de desbastamento; avaliação da coleção; e política de avaliação.

Chama a atenção como na teoria acerca do desenvolvimento de coleções surgida no final do século XX o papel de destaque que há na participação ativa do usuário para ajudar a orientar as instituições na escolha ou descarte de itens de uma coleção, o protagonismo do público que busca a informação parece um movimento natural da visão pós-modernista da sociedade, onde as novas tecnologias levam cada indivíduo a um status de, ao mesmo tempo, formador e consumidor de informações (ou conteúdos, no jargão atual). Na pós-modernidade, o usuário produz informação sem parar e sem perceber, o que remete ao livro “No Enxame”, do filósofo Byung-Chul Han (2021), que procura desenhar um retrato do nosso tempo usando a imagem do enxame, a qual tenta reproduzir o barulho perturbador das redes, barulho este que não visa a coletividade, mas sim uma confusão de informações sem viés ético. A informação não serve mais como catalisador de ações afirmativas, mas seu volume imenso leva a um sufocamento. A resposta do nosso tempo é a imobilidade, o acúmulo leva à inanição. Diante desse cenário, é inegável que o impacto das novas tecnologias irá mudar os rumos que as instituições de memória terão e, ao trazer os conceitos do desenvolvimento de coleções para o ambiente da web, fica claro como diante do barulho constante da web, o desenvolvimento de coleções é uma ferramenta pertinente para selecionar o que de mais importante foi publicado na web sobre um mesmo assunto, tema, pessoa, entre outros.

O intuito de uma coleção web para além da seleção de materiais relevantes sobre um tópico é também a preservação do ambiente na Internet inscrito em seu próprio tempo, uma vez que uma coleção da web é capaz de preservar não apenas um determinado conteúdo, mas também a navegação e acessibilidade naquele dado momento, a usabilidade de um determinado site, ou seja, o contexto em que aquela publicação da web estava inserida mantendo preservada, por exemplo, a arte original do site com as mesmas cores, fontes e quaisquer outros elementos acessórios. A importância disso reside no fato de que com a constante renovação de endereços eletrônicos – os quais são facilmente editáveis sem que haja um controle da parte do usuário das mudanças de layout, conteúdo e inserção de novas informações – sem a coleção, não é possível ter um acompanhamento da evolução de um tópico no ambiente da Internet, como defendem Rochemback e Pavão

(2024), o arquivamento da web deve ser representativo da diversidade da experiência on-line para que no futuro tenhamos uma boa visão de como era navegar na Internet naquele contexto específico e, por consequência, ter um retrato de determinado tópico e seus desdobramentos em um ambiente informacional virtual. Em suma, o desenvolvimento de uma coleção da web é uma importante ferramenta para analisar de modo mais abrangente o percurso daquele tema no ambiente das redes, ambiente este que cada vez mais tem respaldo na vida das comunidades e relevância como documento histórico.

A coleta de páginas da web relacionadas a um determinado acontecimento prevê o arquivamento de páginas que contenham notícias de sites, artigos de opinião, postagens em redes sociais, gravações de áudio e vídeo, fotografias e outros conteúdos multimídia. Essa diversidade de formatos é o que possibilita um registro amplo e diverso desse evento e seu impacto nos indivíduos envolvidos. Por esse motivo, é importante que o responsável pelo desenvolvimento de uma coleção web tenha cuidado ao elencar vozes dissonantes sobre um mesmo assunto, trazendo amplitude a um eventual debate e buscando não limitar o debate ou impregnar a coleção de vieses excludentes.

Uma coleção de arquivos da web brasileira pode fornecer um recurso único para pesquisadores, não apenas os dedicados ao tema da Internet ou às Humanidades Digitais, mas qualquer um que busque dados de circulação no ambiente virtual, em qualquer área ou nível de especialização. Como já citado no presente trabalho, a participação do Brasil na Internet e na web é bastante expressiva, o que leva à necessidade de aplicarmos políticas contundentes de arquivamento da web, não apenas limitado ao conteúdo que as instituições de memória já disponibilizam em sites ou redes sociais, mas o arquivamento do que é produzido no dia a dia, em diferentes veículos e para diferentes públicos. Outro aspecto relevante do desenvolvimento de coleções da web brasileira é a possibilidade de um determinado tema ter seu conteúdo na web salvaguardado ao longo dos anos, isso permitirá que um pesquisador possa acompanhar o desenvolvimento deste ao longo dos anos e dos diferentes veículos de comunicação da web.

Fica evidente que o debate acerca do desenvolvimento de uma coleção web tem enfoque no usuário final desse material, uma coleção de páginas da web bem selecionada e articulada com os usuários será de extrema importância para que a

participação no Brasil se mostre relevante em um assunto global, por exemplo, quando o tema é a Covid 19 e como o nacionalmente foram produzidas publicações e material de apoio no contexto da pandemia. Do mesmo modo, a coleção web poderá aproximar o usuário de uma instituição de memória. O autor Tony Horava, em 2010, já apontava como a relação de uma Biblioteca com o usuário advindo do ambiente virtual teria muito a ganhar com a incorporação dos recursos da web no âmbito do desenvolvimento de coleções:

(..) nossa presença na web nos permite novas capacidades de conectar e ser relevantes para a população que servimos. Ao focar no acesso remoto, na gestão ativa (localmente e em colaboração com grupos externos), em uma maior conscientização sobre os resultados de aprendizagem e na agenda do pesquisador, e em uma abordagem cuidadosamente equilibrada e sustentável para as atividades de aquisição, podemos aumentar o valor e o uso de nossas coleções. (Horava, 2010, tradução própria)

A capacidade de mapear o ambiente em constante mudança da web, com sua principal característica que é a efemeridade, e ter a capacidade de selecionar conteúdos relevantes e diversificados é uma importante tarefa de instituições de memória como bibliotecas, arquivos e museus. Diante do que se convencionou chamar de “infodemia”⁵, ou seja, epidemia de informações, o arquivamento de coleções da web brasileira se mostra uma ferramenta muito importante para evitar a proliferação de informações falsas, fora de contexto ou pura desinformação, além, é claro, do risco de se perder dados disseminados via web diariamente. Vint Cerf, matemático, informático e um dos idealizadores da Internet, alertou em 2015 para uma possível “era das trevas” digital, que seria o apagamento de todos os registros da Internet. Para escapar desse destino, o desenvolvimento de coleções da web é um dos caminhos, pois esse é um trabalho importante diante do dinamismo da web. As diferentes coleções possíveis de se formar diante da produção em altíssima

⁵ A Organização Pan-Americana da Saúde define o termo infodemia como referente a um grande aumento no volume de informações associadas a um assunto específico, que podem se multiplicar exponencialmente em pouco tempo devido a um evento específico. Nessa situação, surgem rumores e desinformação, além da manipulação de informações com intenção duvidosa. Na era da informação, esse fenômeno é amplificado pelas redes sociais e se alastra mais rapidamente, como um vírus. ORGANIZAÇÃO PAN-AMERICANA DA SAÚDE. Folha informativa: Infodemia. Brasília: OPAS, 2020. Disponível em: https://iris.paho.org/bitstream/handle/10665.2/52054/Factsheet-Infodemic_por.pdf?sequence=16. Acesso em: 23 maio 2024.

escala da Internet são um caminho para que a memória do final do século XX e do XXI não se percam para as gerações futuras.

3.2 Arquivamento web institucional e governamental

A preservação da web pode ser feita por qualquer instituição que entenda que esta é uma atividade relevante para a organização, não apenas no nível individual, as instituições podem vir a preservar a web visando garantir a consulta futura de conteúdos relacionados a temas de interesse do público. Segundo a Unesco (2003), é essencial que as instituições nacionais de referência tenham um papel de liderança, estabelecendo políticas e sistemas de coleta e gerenciamento de objetos digitais da web ou liderando redes colaborativas para a adoção de modelos comuns de seleção e preservação. No Brasil, segundo a política nacional de arquivos públicos e privados, disposta pela Lei n. 8.159, de 8 de janeiro de 1991, os proprietários de websites têm autonomia para definir seus procedimentos de preservação e acesso à documentação digital. A lei prevê que as pessoas físicas e jurídicas de direito privado, detentoras de arquivos, podem integrar o Sistema Nacional de Arquivos (Sinar) mediante acordo com o Conselho Nacional de Arquivos (Conarq), assim como conjuntos de documentos arquivísticos dessa natureza serem declarados de interesse público e social. Porém, antes de pensarmos como se dá a aplicação da Lei n. 8.159 no âmbito dos websites, é importante rever o caminho que levou até a sua publicação, mais especificamente, como a preocupação com o arquivamento de informações institucionais ou governamentais se iniciou.

Segundo o site Glaber (2018), a primeira instituição brasileira voltada para o arquivamento de documentos oficiais, e que deu origem ao Arquivo Nacional, foi criada em 2 de janeiro de 1838, nomeada como Arquivo Público do Império, era subordinada à Secretaria de Estado dos Negócios do Império, e fazia a guarda de documentos relacionados às áreas de ensino e saúde pública, saneamento, transporte, correio, agricultura, indústria e mineração. Após a Proclamação da República, em 1889, o então Arquivo Público do Império passou a se chamar Arquivo Público Nacional. Em outubro de 1891, a Lei n. 23 reorganizou os serviços da administração federal, criando o Ministério da Justiça e Negócios Interiores, assim, o arquivo passou a ser subordinado à recém-criada pasta a qual cuidava de

assuntos relacionados à Instrução Pública, aos Correios e Telégrafos e do Interior. O decreto n. 1.580, de 31 de outubro de 1893, determinou um novo regulamento para o órgão, o qual estabelecia os procedimentos de trabalho, os funcionários e o tipo de documentos a ser recolhidos em cada seção. Foi também criado o Museu Histórico para abrigar bens da família imperial.

O Ministério da Justiça e Negócios Interiores foi reorganizado pelo Decreto n. 9.196, de 9 de dezembro de 1911, e o Decreto n. 9.197 aprovou um novo regulamento para o órgão, que passou a se chamar Arquivo Nacional. A estrutura administrativa estabelecida em 1911 manteve-se até o Decreto n. 15.596, de 2 de agosto de 1922, que criou o Museu Histórico Nacional, que determinou a transferência para o novo órgão dos objetos integrantes do Museu Histórico do Arquivo Nacional. No ano seguinte, o arquivo foi reformado pelo Decreto n. 16.036 e passou a contar com as seções Administrativa, Histórica, Legislativa e Judiciária, além da Secretaria, à qual estavam subordinadas a Biblioteca, a Mapoteca e as oficinas Gráfica e de Encadernação. O decreto propôs ainda a criação de um Curso Técnico, em conjunto com a Biblioteca Nacional e o Museu Histórico Nacional, com o objetivo de formar funcionários aptos a trabalhar nessas instituições. O órgão manteve essa configuração administrativa até o final da Primeira República.

No período subsequente, o Arquivo Nacional sofreu as consequências de uma época de instabilidade institucional derivada da sucessão de governos provisórios e da ditadura da Era Vargas. O cenário só mudou a partir da gestão de José Honório Rodrigues, que assumiu o Arquivo Nacional em 1958 e em 1959 publicou um documento no qual detalha as dificuldades e a realidade obsoleta do Arquivo Nacional na época. O Decreto n. 44.862, de 21 de novembro de 1958, aprovou o novo regimento do Arquivo, o definindo como uma repartição nacional, além disso, apresenta a política de arquivos, que estabelece suas atribuições e objetivos, defende e amplia a coleta selecionada em todo território nacional expandindo a documentação para também filmes, discos e fotografias. Segundo Melo (2021), um novo regimento para o Arquivo Nacional só seria aprovado em 1975, que estabeleceu a seguinte estrutura para a instituição: Divisão de Pré-Arquivo, Divisão de Documentação Escrita, Divisão de Documentação Audiovisual, Divisão de Pesquisas e Atividades Técnicas, Divisão de Publicações, Divisão de Administração e Coordenadoria de Cursos de Arquivologia. Já nos anos de 1980, o Arquivo Nacional iniciou um processo de modernização, o qual foi

realizado a partir do convênio entre o Ministério da Justiça e a Fundação Getúlio Vargas e culminou no Decreto nº 88.771, de 27 de setembro de 1983, o qual tornou o Arquivo Nacional órgão autônomo da administração direta, do Ministério da Justiça.

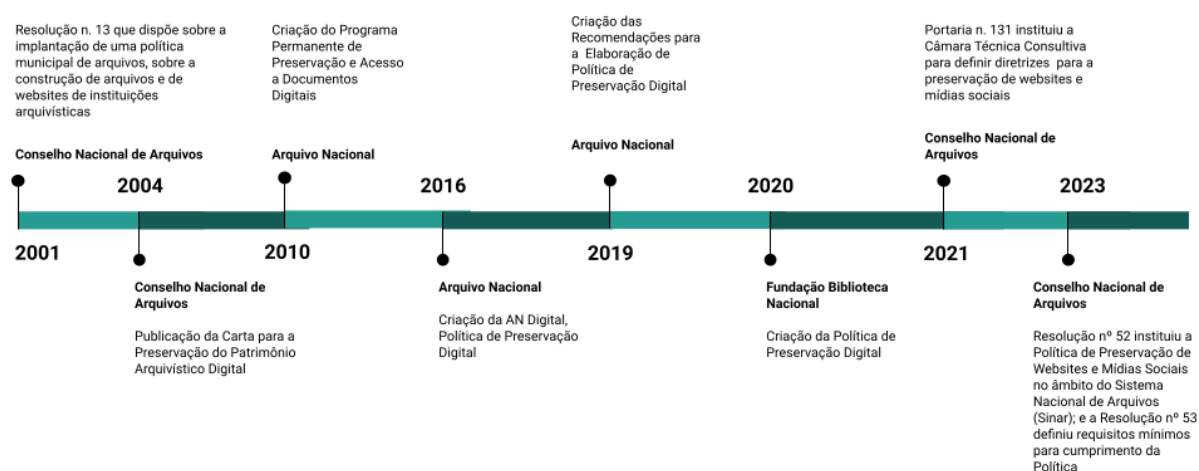
A breve retrospectiva apresentada anteriormente sobre a trajetória do Arquivo Nacional até a Lei n. 8.159, de 8 de janeiro de 1991 demonstra não apenas a importância da instituição para determinação do ferramental necessário para salvaguardar a memória institucional e governamental do Brasil, mas também é possível notar que ao longo dos anos o aperfeiçoamento dos métodos de trabalho foram capazes de processar grandes volumes documentais de natureza pública, o que antes se limitava a uma instituição sem plena agência dentro de uma subdivisão de uma Secretaria, ganhou sua autonomia, essencial para uma gestão mais eficiente do trabalho. Com essa autonomia, cresceram as subdivisões de trabalho que acompanharam a complexidade cada vez maior de tipos documentais e aplicações desses em pesquisas e decisões importantes para a democracia brasileira. Voltando para o texto da Lei de 1991, cujo texto diz que:

Consideram-se arquivos, para os fins desta Lei, os conjuntos de documentos produzidos e recebidos por órgãos públicos, instituições de caráter público e entidades privadas, em decorrência do exercício de atividades específicas, bem como por pessoa física, qualquer que seja o suporte da informação ou a natureza dos documentos. (BRASIL, 1991)

É evidenciado que naquele momento existia a percepção de que caberia à cada instituição a responsabilidade frente ao arquivamento de sua documentação produzida. Como o desenvolvimento da web no Brasil e a consequente adesão de cada vez mais instituições oficiais e órgãos públicos à publicação em não apenas sites oficiais, mas também redes sociais, tais como Facebook, Instagram e X (antigo Twitter) a aplicação dessa determinação (de que caberia à entidade o arquivamento de seus documentos) passa a ficar mais intrincada, afinal, apesar do conteúdo postado ser oficial, os donos dos veículos são as big techs, como por exemplo a Meta, de propriedade do norte-americano Mark Zuckerberg. Nota-se após a publicação da lei que paulatinamente houve uma evolução da ideia inicial de que somente as empresas proprietárias de websites façam o arquivamento da web. Ao longo dos anos, e da evolução da web, ficou evidente que haveria a necessidade de

outras instituições dedicarem tempo e recursos para garantir o acesso e a preservação do conteúdo da web periodicamente. A mudança de perspectiva no que tange ao arquivamento da web pode ser observada ao longo das decisões governamentais destacadas na linha do tempo a seguir:

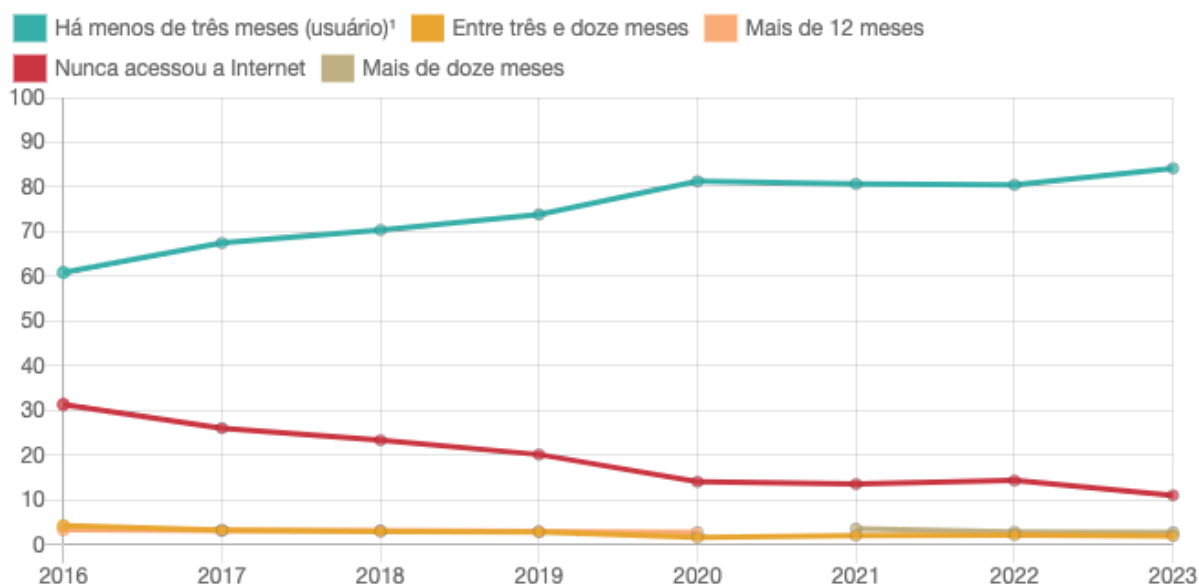
Figura 4: Principais Iniciativas Oficiais de Arquivamento Digital no Brasil



Fonte: Elaborado pela autora.

O que é possível notar ao longo dos anos é que houve primeiramente uma preocupação em trabalhar acervos digitais nas instituições, por exemplo, documentos originais em meio analógico que foram digitalizados. A atenção ao acervo nato-digital que circula na web é uma preocupação mais recente em nosso país, o que contrasta com o uso e circulação de informações na Internet brasileira, pois somos um país com larga participação na web. O gráfico a seguir demonstra uma linha crescente de usuários da Internet, desde o ano de 2016 até 2023, a população que declarou ter acessado a Internet há menos de um mês saltou da marca de 60 milhões para mais de 80 milhões de brasileiros, é também importante notar que o número de usuários que nunca acessou a Internet está em tendência de queda: saiu do patamar de 30 milhões para quase 10 milhões de brasileiros. São números impressionantes que demonstram a onipresença da tecnologia no país e sua importância no dia a dia das pessoas.

Gráfico 5: Indivíduos, por Último Acesso à Internet



Fonte: CGI.br/NIC.br, Centro Regional de Estudos para o Desenvolvimento da Sociedade da Informação (Cetic.br), Pesquisa sobre o Uso das Tecnologias de Informação e Comunicação nos domicílios brasileiros.

Pensando que o acesso universal à Internet pelos brasileiros está cada vez mais próximo, fica evidente que a publicação de conteúdos relevantes para o país é também expressiva. Segundo o site do Núcleo de Informação e Coordenação do Ponto BR, em 2024 o domínio “.br” atingiu a marca de 5,3 milhões de nomes registrados. Operado pelo NIC.br, o “.br” disponibiliza mais de 130 categorias de domínios que abarcam variadas áreas de atuação, iniciativas e serviços. Há categorias gerais (como “ong.br”, “art.br” e “eco.br”); para profissionais liberais (“bio.br”, “adm.br”, “mus.br”, “med.br”, “eng.br”, entre outros); pessoas físicas (“blog.br”, “vlog.br”, “wiki.br”, por exemplo); pessoas jurídicas (como “agr.br”, “far.br”, “tv.br” e “tur.br”); instituições de ensino superior (edu.br); e as que identificam cidades (“rio.br”, “manaus.br”, “cuiaba.br”, “floripa.br”, “foz.br” etc). Segundo Rockembach e Pavão (2024), o domínio de um site não é o único indicativo para certificar que o seu conteúdo esteja relacionado com o país, mas ao analisarmos o uso crescente deste domínio já é possível termos pistas do quão profícua é a

participação do Brasil na web. Pensando em como a quantidade de endereços ativos na web é tão grande, nota-se a necessidade de implementação de políticas de arquivamento da web com o objetivo de preservar nosso rico patrimônio digital.

As instituições de memória e governamentais têm papel fundamental na elaboração e promoção dessas políticas de arquivamento da web, uma vez que são referência para outras iniciativas sendo particulares, educacionais, empresariais, entre outras. Por isso, diretrizes nacionais devem ser publicadas para que as instituições oficiais de um país possam ter uma referência sólida para executar o trabalho de arquivamento da web. Essas diretrizes serão parecidas com as do arquivamento de qualquer outro tipo de informação em meio analógico, ou seja, deverão contemplar tópicos como responsabilidades, planejamento em caso de catástrofes, planos de contingência, financiamento, propriedade, direitos de propriedade intelectual e direitos de autor; e também os tópicos próprios da especificidade do conteúdo da web, tais como utilização da tecnologia, formato, armazenamento e backup, refrescagem e emulação, acesso e descrição de arquivo (Rockembach; Pavão, 2024).

3.3 Políticas de arquivamento web no Brasil

A política de arquivamento no Brasil é um tema que tem gerado debates não apenas no nível das instituições arquivísticas, mas também entre os estudiosos da área. Segundo Jardim (2015), desde 2011, ano que marcou a transferência do Arquivo Nacional para o Ministério da Justiça e a I Conferência Nacional de Arquivos (I CNARQ), cujo tema foi “Por uma política nacional de arquivos” é notável como este assunto tornou-se cada vez mais presente em artigos, dissertações e teses, ampliando e aperfeiçoando as bases teóricas e metodológicas para tratamento do tema. Uma política de arquivamento e preservação visa estabelecer as decisões que uma instituição deve tomar a fim de garantir a composição e conservação de seu acervo, esse tipo de diretriz envolve indicar quais suportes tecnológicos serão usados para uso e manutenção do acervo, além de decisões políticas, institucionais e administrativas, tudo isso levando em consideração estudos técnico-científico que respaldam as decisões.

Para auxiliar o desenvolvimento de qualquer política de arquivamento, seja ela institucional, nacional ou de iniciativas privadas, é possível se orientar tendo

como respaldo um ordenamento jurídico ou também um ordenamento técnico. As normas jurídicas propiciam sustentabilidade legal e, no Brasil, o Conselho Nacional de Arquivos é a principal entidade responsável pela formulação e definição dos eixos de atuação da política arquivística nacional, este ordenamento jurídico é evidenciado na Lei nº 8.159 de 8 de janeiro de 1991 que estabelece ser dever do Poder Público a gestão documental e a proteção especial a documentos de arquivos, como instrumento de apoio à administração, à cultura, ao desenvolvimento científico e como elementos de prova e informação. Ainda segundo Jardim, a aplicação de tais leis se depara com muitos entraves para que se consolide uma política nacional de arquivos, segundo o autor:

As declarações de políticas arquivísticas emitidas por serviços e instituições apontam para diferentes formas de visualizar os processos, ações e sujeitos da gestão arquivística. Não raro um dispositivo legal, projeto ou programa são caracterizados como uma política arquivística. No entanto, o fato de se tomar uma decisão política de implementar um projeto de gestão de documentos não confere a esse projeto a condição de uma política arquivística. (Jardim, 2015, p. 3).

A crítica de Jardim (2015) acerca da aplicação de uma política nacional de arquivos pode também ser observada em um documento publicado pela Escola Nacional de Administração Pública em 2022 intitulado “Política Nacional de Arquivos Públicos e Privados: uma análise dos avanços da gestão documental e arquivística brasileira”. O texto indica quatro principais entraves para a consolidação da Política Nacional de Arquivos Públicos e Privados, sendo eles:

- a ausência de poder de influência do grupo social interessado para motivar os agentes políticos a viabilizar a política;
- a percepção sobre a não essencialidade de uma política pública para o campo da arquivologia na agenda da gestão pública;
- a insuficiência de comprometimento do Estado com o uso adequado da informação pela administração pública; e
- a ausência de reconhecimento da sociedade civil do direito ao acesso à informação governamental e baixo envolvimento no usufruto deste. (Adão, 2017; Jardim, 2008 *apud* Bianchi; Reis, 2022, p. 15)

No que tange ao arquivamento da web brasileira, as dificuldades que já são apontadas para a aplicação de uma política sólida de arquivamento ganham uma problemática extra, pois as instituições brasileiras passaram muito tempo dedicadas ao arquivamento de documentos digitais, porém, sem se atentar às particularidades da documentação da web, ou seja, a atenção estava voltada aos arquivos digitalizados, mas deixando de lado aqueles nato-digitais produzidos por meio das tecnologias da web, tais como aplicativos de trocas de mensagens, redes sociais, publicações em sites e blogs, entre outros. A mudança parece estar começando agora, pois em 2023, foi finalmente publicada pelo Conselho Nacional de Arquivos (Conarq) a Resolução nº 52, instituindo a Política de Preservação de Websites e Mídias Sociais no âmbito do Sistema Nacional de Arquivos (Sinar) e, mais especificamente, a Resolução nº 53, a qual define os requisitos mínimos de preservação digital para websites e mídias sociais a serem adotados pelos integrantes do Sistema Nacional de Arquivos (Sinar).

Além das normas jurídicas, que poderiam apoiar e incentivar o desenvolvimento de uma política nacional de arquivamento da web brasileira, também há a possibilidade de se recorrer ao desenvolvimento de normas técnicas específicas para a área. Esse fato requer a compreensão de suas distinções e aproximações. Segundo Silva (2016), a aproximação entre normas jurídicas e normas técnicas é conhecida e experimentada por áreas como a saúde, meio-ambiente e segurança, em alguns casos, as normas técnicas podem ter peso regulamentário e até algumas normas jurídicas podem se reportar às normas técnicas ou normas técnicas que se estabelecem por meio de imperativo jurídico.

Um exemplo de normativa que trata de documentação da web vem da International Organization for Standardization, que publicou em 2013 a normativa específica para arquivamento da web intitulada “ISO/TR 14873:2013 Information and documentation — Statistics and quality issues for web archiving”. No documento estão descritos os métodos e a finalidade do arquivamento da web, além de indicadores estatísticos, de qualidade, usos e benefícios do arquivamento da web. Trata-se de um importante instrumento para guiar qualquer instituição que almeje aplicar uma política de arquivamento da web com indicadores seguros de desempenho. No entanto, a versão brasileira da ISO, o Comitê Brasileiro de Informação e Documentação (NBR/CB-014), ligado à Associação Brasileira de Normas Técnicas (ABNT) ainda não realizou a versão nacional de tal documento. A

última atualização da norma, datada de 2016, não traz em seu escopo a norma específica para o arquivamento da web.

Diante de uma certa lacuna, tanto no campo jurídico de antes das resoluções de 2023, quanto no campo normativo, o Brasil se fez presente no arquivamento da web com o auxílio de algumas iniciativas de outros países que abordaram os arquivos da web brasileira. Rockembach (2017) em artigo dedicado a delimitar conceitualmente e teoricamente o tema arquivamento da web, cita, por exemplo, uma iniciativa relacionada à web brasileira denominada Latin American Web Archiving Project, hospedada no site da Universidade do Texas em Austin, com ênfase em documentos governamentais e expressões políticas. A coleta de materiais para arquivamento começou em 2005, contudo, nem todas as coleções continuam a ser atualizadas.

Ao explorar a página do Latin American Web Archiving Project, o autor identificou quatro coleções: Latin American Government Documents Archive (LAGDA); México 2010; Archive of Venezuelan Political Discourse (ARVEPODIS); e Archive of Political Parties and Elections in Latin America (APPELA), destas, o Brasil aparece na última coleção, APPELA, a qual traz um recorte de publicações de sites dos partidos brasileiros. Entre os sites de partidos brasileiros, estão presentes na coleção os seguintes: Partido Comunista do Brasil (PCdoB), Partido Democrático Trabalhista (PDT), Partido da Frente Liberal (PFL), Partido da Social Democracia Brasileira (PSDB), Partido do Movimento Democrático Brasileiro (PMDB), Partido dos Trabalhadores (PT), Partido Liberal (PL), Partido Popular Socialista (PPS), Partido Progressista (PP), Partido Socialista Brasileiro (PSB), Partido Trabalhista Brasileiro (PTB), além do Tribunal Superior Eleitoral (TSE). Esta é uma coleção interessante, que consegue dar aos usuários uma contribuição inestimável sobre a política nacional e as estratégias de comunicação dos partidos políticos brasileiros ao longo dos anos. Infelizmente, das publicações salvas, a coleção referente ao TSE possui apenas um arquivo, o que demonstra que essa iniciativa não teve continuidade, possivelmente não houve interesse da instituição em seguir capturando dados relativos aos partidos políticos brasileiros e a notícias relacionadas às eleições do país ao longo dos anos, o que resta é apenas uma pequena amostra, com bastante potencial mas não tão diversa em conteúdo.

Figura 5: Print Screen do Site www.vermelho.org.br (PCdoB) armazenado no Internet Archive



Fonte: Latin American Network Information Center, 2024.

O fato de uma parte significativa da web brasileira estar contemplada em projetos internacionais é um fato preocupante, que denota a emergência de não só construir uma política nacional de arquivamento da web, como é a intenção das Resoluções n. 52 e n. 53 do Conselho Nacional de Arquivos, mas é preciso que haja projetos que coloquem em prática efetivamente tais decisões pelas instituições de memória brasileiras. A tese de doutorado de Dantas (2014) é um outro exemplo de como seria possível proceder ao arquivamento da web brasileira caso fossem implementadas políticas específicas para tal. Na tese, a autora abordou questões relativas ao arquivamento da web e também identificou iniciativas internacionais de arquivamento da web, constatando não haver coleções de páginas web em instituições brasileiras. Mas o trabalho da autora trouxe uma importante contribuição na construção de uma coleção relativa a ferramentas de busca no Brasil lançando mão de pesquisa no site Internet Archive, a autora teve êxito em localizar algumas coleções relacionadas ao conteúdo produzido no Brasil, mas de forma muito dispersa, sem uma linha definida ou políticas de seleção e arquivamento estabelecidas, por exemplo, coleções de revistas dos anos de 1984, 1985 e 1986.

Já a coleção “Brazilian Web Engines (1997- 2013)”, foi desenvolvida em um projeto de pesquisa na Universidade Federal do Estado do Rio de Janeiro (Unirio), que procurou armazenar snapshots de motores de busca brasileiros dos anos de 1990, como Cadê, UOL, TodoBr e Radix. O trabalho da autora é um exemplo prático da potencialidade de se ter uma política de arquivamento da web brasileira colocada em prática para o desenvolvimento de uma coleção, mesmo que esta use de um acervo internacional como o Internet Archive.

4 Considerações finais: o documento web e a memória social digital

A análise do desenvolvimento da web no Brasil leva à reflexão sobre o que seria uma memória social digital e como os documentos web são entendidos a partir de sua produção e posterior alcance de status de patrimônio digital. Todo esse processo que se inicia na produção a partir de qualquer usuário da Internet, sendo via redes sociais ou produtores de conteúdo em sites (levando-se em conta nessa reflexão apenas documentos nato-digitais, e não aqueles de acervo que já tiveram seu valor patrimonial determinado e passaram de um suporte físico para o digital), desse modo, faz-se necessário abordar os critérios para seleção, tratamento e disponibilização de bens culturais no ambiente web. A autora Vera Dodebei apresenta uma reflexão sobre o ciclo de vida de objetos digitais na Internet que está pautada a partir de três conceitos:

- “fluidez/efêmero” – aquilo que dura pouco, como no exemplo dos objetos concretos, um perfume;
- “estável/ resíduo” – aquilo que sobra, por exemplo, o pó de café depois que produzimos a bebida; e
- “reuso/recicláveis” – que representa dejetos que se transformam em outros objetos, como, por exemplo, uma camiseta feita com o material descartado de garrafas PET. (Dodebei, 2021, p. 1)

Os conceitos explorados pela autora denotam a natureza da memória na Internet, que é criada coletivamente e gera a massa de dados que é denominada big data, ou seja, esse emaranhado de bytes que atualmente são produzidos em uma escala impossível de assimilação pela mente humana, mas que, no entanto, impacta nossas vidas. Sendo essa massa de dados produzida em escala vertiginosa, e por diferentes agentes produtores, há uma grande mudança no que tange à pluralidade de saberes possíveis, uma vez que a probabilidade de uma comunidade antes à margem poder encontrar canais de comunicação é muito maior nesse tipo de ambiente. Entretanto, chama a atenção nos três conceitos delimitados pela autora especialmente a ideia de informações fluídas, ou seja, que podem ser descartadas, e as estáveis, as quais devem ser tratadas como patrimônio e por isso protegidas.

A partir dessa reflexão, é possível iniciar um caminho analítico sobre o conteúdo da web entendido como exemplo de objeto documental relevante para

memória digital social, mas não sem antes também incluir na análise a evolução que ocorreu nas instituições de informação, ou seja, bibliotecas, museus, centros de memória, arquivos públicos ou centros culturais. Tais instituições apontaram o que deveria ser considerado como patrimônio cultural de determinados grupos, porém, com o impacto profundo causado pela cada vez maior produção de conteúdo digital, tais instituições passaram a ter que delimitar procedimentos para lidar com um tipo de informação com um ciclo informacional próprio. Segundo Lídia Eugenia Cavalcante, os documentos nato-digitais são inerentemente efêmeros, além de serem produzidos de modo caótico e sem preocupação com a preservação, isso pode gerar um alto impacto no futuro da memória de instituições que ainda não definiram como lidar com esse tipo de produção documental (Cavalcante, 2017).

Antes de refletir sobre o papel da web como documento formativo da memória social digital, é importante estabelecer conceitos de memória social e o papel das instituições de preservação na delimitação do que é relevante e imprescindível nesta construção. O conceito de memória social é utilizado por diversos teóricos, não sendo limitado a um único campo de estudo. A autora Camila Dantas, em tese acerca da memória digital brasileira, destaca autores essenciais para entender o tópico e como ele se relaciona ao estudo da área de Documentação: para Dantas (2014) o autor Maurice Halbwachs é imprescindível para melhor entender o campo de estudo da memória social, pois para este autor a memória deve ser entendida como um fenômeno coletivo uma vez que os indivíduos não recordam sozinhos, a partir disso surge o conceito de quadro social, termo oriundo dos estudos literários, mas que segundo Maurice Halbwachs remete a um sistema de representações que, conforme Dantas, podem ser sistemas de lógica, de cronologia ou topografia, que antecipam a lembrança fornecendo assim um sistema global de localização do passado no presente (Namer, 1994, p. 325 *apud* Dantas, 2015, p. 62).

A tese de Dantas (2014) também destaca a importância do autor Michel Foucault para o debate sobre o tema, especialmente em sua abordagem sobre o arquivo como metáfora para relações de poder, a uma lei geral, como descrito em *Arqueologia do Saber*:

É preciso desligar a história da imagem com que ela se deleitou durante muito tempo e pela qual encontrava sua justificativa antropológica: a de uma memória milenar e coletiva que se servia de documentos materiais para

reencontrar o frescor de suas lembranças; ela é o trabalho e a utilização de uma materialidade documental (livros, textos, narrações, registros, atas, edifícios, instituições, regulamentos, técnicas, objetos, costumes etc.) que apresenta sempre e em toda a parte, em qualquer sociedade, formas de permanências, quer espontâneas, quer organizadas. O documento não é o feliz instrumento de uma história que seria em si mesma, e de pleno direito, memória; a história é, para uma sociedade, uma certa maneira de dar status e elaboração à massa documental de que ela não se separa. (Foucault, 2008, p. 7–8)

Foucault (2008) dá dimensão ao entendimento de uma memória construída coletivamente, pois o autor não a define como a acumulação de todos os textos guardados sobre algo, mas sim como a formação de um discurso produzido na história sendo este fenômeno algo contínuo, uma vez que a história é uma constante elaboração dos fatos. A partir dessa reflexão, é importante salientar que o caráter de efemeridade inerente aos documentos produzidos diretamente em meio digital pode afetar significativamente o futuro da memória de instituições, movimentos sociais, partidos políticos, associações e de comunidades, entre outros. A ideia de memória digital perpassa a renovação dos processos de identificação, descrição e comunicação de conteúdos informacionais. A maneira desordenada com que a informação digital é tratada denota que as instituições ainda não perceberam o valor da memória digital e como esta deve ser entendida como patrimônio cultural.

Retomando os conceitos enunciados por Dodebei (2021) a memória digital pode ser definida como um tipo de informação maleável, sendo ela efêmera em sua materialidade e reciclada como prática, ao pensarmos na constante reedição e readaptação de conceitos que infesta o debate em redes sociais, citando a aqui a cultura do meme, termo que surgiu na obra do biólogo Richard Dawkins, em 1976, em livro intitulado “The Selfish Gene”, o autor define meme como uma unidade de replicação cultural, funcionando para a cultura de forma análoga à que genes funcionam para características biológicas (Dawkins, 2006). Desse modo fragmentário e multiplicador, a cultura digital vai construindo uma cultura de difícil manutenção e classificação.

A dificuldade de manutenção e classificação do objeto documental produzido em meio digital reflete a escassez de parâmetros definitivos para a manutenção da

documentação da web e, conseqüentemente, a compreensão do percurso formativo da memória digital. Outro fator de enorme relevância quando se debate a cultura digital é o monopólio de megacorporações no que se refere à tecnologia empregada para a produção, distribuição e manutenção de informações digitais. As empresas particulares desenvolvem novas tecnologias constantemente sem que sejam obrigadas a prestar contas à sociedade sobre como é o real funcionamento dessas ferramentas, essas soluções muitas vezes não conseguem captar a confusa interação de valores, missões e tradições no coração das instituições que tradicionalmente se responsabilizam pela manutenção cultural (como museus, bibliotecas, arquivos). Desse modo, fica evidente que, ao analisarmos a memória digital, é de suma importância que essas instituições públicas, as quais não são guiadas pelo objetivo de gerar lucros, sejam protagonistas na discussão acerca da memória digital.

Quando se tem no horizonte inovações como o recente Chat GPT, o qual em linhas gerais lança mão de uma produção publicada e disponibilizada gratuitamente na Internet, mas que já passou por algum tipo de edição e controle de qualidade, nota-se que a memória digital corre sérios riscos de ser usurpada por corporações que usam-na como mero conjunto de dados fora de contexto que não leva em consideração a sociedade, os valores e os direitos dos cidadãos que primeiramente compartilharam suas reflexões. O valor documental de objetos da sociedade que geram sua memória digital passa pelos processos de identificação, descrição e comunicação de seus conteúdos informacionais. Como reflete Dodebei, é urgente preservar a memória digital em um contexto social que denota “o paradoxo da manutenção de bens patrimoniais quando o default social é o consumo e o esquecimento” (Dodebei, 2021). Refletir sobre a memória digital é refletir sobre quais movimentos, valores e comportamentos estão sendo consumidos e esquecidos, em nome de que tipo de ideologia e quais grupos sociais estão sendo constantemente apagados e em nome de quê. Importante pensar que o conteúdo produzido na web sob a responsabilidade de empresas que visam exclusivamente o lucro acaba transcendendo esta vertente utilitária. As redes sociais e demais endereços web acabam desvirtuados do papel de simples ferramentas tecnológicas trabalhando em prol de uma big tech e passam a abarcar toda a complexidade da produção humana, com seus inúmeros erros e acertos, lógicas e contradições.

Quando o aparato tecnológico da Internet foi entregue à iniciativa privada em meados dos anos de 1990, não havia ainda uma noção do quanto da produção na web iria impactar a esfera pública e guiar a cultura e consequentemente a memória de nosso tempo. Agora, vivemos uma realidade em que a falta de regulamentação específica deixa a sociedade carente de instituições que tratem a profusão de informação relevante on-line com as ferramentas necessárias, ou seja, instituições como as Bibliotecas que operem como repositórios de informação digital e consigam salvaguardar o que existe de mais importante na web, permitindo assim que o debate da esfera pública digital esteja disponível para esta e as futuras gerações. As instituições de memória já possuem métodos, critérios e guias acumulados durante anos de experiência lidando com a informação em meio analógico. O desafio atualmente é lidar com a chamada “plataformização” da Internet, um fenômeno causado pelo monopólio de empresas privadas sobre dados e navegação de usuários e, consequentemente, sobre a produção mais relevante que circula na web.

No livro “Sinais e Maravilhas: da arte & cultura na era digital”, Teixeira Coelho reflete sobre a tecnologia de processamento dos supercomputadores e como este cenário poderá ser capaz de colocar a humanidade à margem do processo de conhecimento. Teixeira Coelho também cita a nossa “eCultura” cotidiana, aquela que se constitui a partir das redes sociais, o autor destaca que: “o sujeito como o ator ativo no processo de conhecimento também se extingue porque seus vínculos com o objeto estão faltando ou igualmente desaparecem” (Coelho, 2021, p. 175), o que o autor procura salientar aqui é como a dinâmica computacional parece estar afastando a humanidade do papel de sujeitos conscientes de si, colocando a todos como simples propagadores de informação vazia. Um antídoto para este cenário é tratar a web com mais racionalidade, nesse sentido, o arquivamento da web tem um papel importante ao dar dimensão palpável e mais próxima de uma racionalidade à informação criada em ambiente virtual desordenado.

Referências

ACHILLES, Daniele. Desenvolvimento de coleções: apontamentos teóricos sobre bibliotecas especializadas. *In*: DESENVOLVIMENTO DE COLEÇÕES: APONTAMENTOS TEÓRICOS SOBRE BIBLIOTECAS ESPECIALIZADAS, 2014. **7º Seminário Internacional de Bibliotecas Públicas e Comunitárias**. 2014.

Disponível em:

<https://docplayer.com.br/24962027-Desenvolvimento-de-colecoes-apontamentos-teoricos-sobre.html>. Acesso em: 21 jul. 2024.

ARQUIVO.PT. Coleção especial acerca da Covid-19 – sobre.arquivo.pt. *In*: 2 nov. 2020. Disponível em:

<https://sobre.arquivo.pt/pt/colecao-especial-acerca-do-covid-19/>. Acesso em: 21 jul. 2024.

ARQUIVO.PT. **Informações gerais**. 2024. Disponível em: <https://arquivo.pt/>. Acesso em: 21 jul. 2024.

BERNERS-LEE, Tim. Information Management: A Proposal. 1989.

BIANCHI, Lorenzo; REIS, Breno Salomon. Política Nacional de Arquivos Públicos e Privados: uma análise dos avanços da gestão documental e arquivística brasileira.

<https://repositorio.enap.gov.br/handle/1/6939>, 2022. Disponível em:

<http://repositorio.enap.gov.br/jspui/handle/1/7415>. Acesso em: 27 jul. 2024.

BRAGG, Molly; HANNA, Kristine. **The Web Archiving Life Cycle Model**. 2013.

Disponível em: <http://archive.org/details/WALCM>. Acesso em: 21 jul. 2024.

BRASIL. **Constituição da República Federativa do Brasil de 1988**. Brasília, 1988.

Disponível em: https://www.planalto.gov.br/ccivil_03/constituicao/constituicao.htm.

Acesso em: 23 jul. 2024.

BRASIL. **Decreto nº 1.580 de 31 de outubro de 1893**. Reforma o Arquivo Público Nacional. Rio de Janeiro, 31 out. 1893. Disponível em:

<https://bdan.an.gov.br/items/4eb48a06-97e8-4d7f-80f6-bba60873c07c>. Acesso em: 27 jul. 2024.

BRASIL. **Decreto nº 9.196 de 9 de dezembro de 1911**. Reorganiza a Secretaria de Estado da Justiça e Negócios Interiores. Rio de Janeiro, 9 dez. 1911a. Disponível

em:

<https://legislacao.presidencia.gov.br/atos/?tipo=DEC&numero=9196&ano=1911&ato=cbf0TWE1UeVRVT6c0>. Acesso em: 27 jul. 2024.

BRASIL. **Decreto nº 9.197 de 9 de dezembro de 1911**. Aprova o regulamento do Arquivo Nacional. Rio de Janeiro, 1911b. Disponível em: <https://bdan.an.gov.br/items/85e005d2-bc1b-4ad4-8371-f9ca4a13afc0>. Acesso em: 27 jul. 2024.

BRASIL. **Decreto nº 15.596 de 2 de agosto 1922**. Cria o Museu Histórico Nacional e aprova o seu regulamento. Rio de Janeiro, 1922. Disponível em: <https://legislacao.presidencia.gov.br/atos/?tipo=DEC&numero=15596&ano=1922&ato=8b10TUU5UMNpWT3e8>.

BRASIL. **Decreto nº 16.036 de 14 de maio de 1923**. Aprova o regulamento para o Arquivo Nacional. Rio de Janeiro, 1923. Disponível em: <https://bdan.an.gov.br/items/29f7b4bd-48aa-4f70-9207-305780ffc552/full>.

BRASIL. **Decreto nº 44.862 de 21 de novembro de 1958**. Aprova o Regimento do Arquivo Nacional, do Ministério da Justiça e Negócios Interiores. Rio de Janeiro, 1958. Disponível em: <https://legis.senado.leg.br/norma/467428>.

BRASIL. **Decreto nº 88.771, de 27 de setembro de 1983**. Inclui o Arquivo Nacional, do Ministério da Justiça, no regime de que trata o Decreto nº 86.212, de 15 de julho de 1981, e dá outras providências. Brasília, 1983. Disponível em: https://www.planalto.gov.br/ccivil_03/atos/decretos/1983/d88771.html. Acesso em: 27 jul. 2024.

BRASIL. **Lei nº 8.159, de 8 de janeiro de 1991**. Dispõe sobre a política nacional de arquivos públicos e privados e dá outras providências. Brasília, 8 jan. 1991. Disponível em: https://www.planalto.gov.br/ccivil_03/leis/l8159.htm. Acesso em: 25 jul. 2024.

BRASIL. **Lei nº 23 de 30 de outubro de 1891**. Reorganiza os serviços da administração federal. Rio de Janeiro, 1891. Disponível em: <https://www2.camara.leg.br/legin/fed/lei/1824-1899/lei-23-30-outubro-1891-507888-publicacaooriginal-1-pl.html>. Acesso em: 27 jul. 2024.

BRASIL. **Portaria Conarq nº 131 de 9 de novembro de 2021**. Definir diretrizes para a elaboração de estudos, proposições e soluções para a preservação de websites e mídias sociais. Brasília, 2021. Disponível em:

<https://in.gov.br/en/web/dou/-/portaria-conarq-n-131-de-9-de-novembro-de-2021-359118428>. Acesso em: 27 jul. 2024.

BRASIL. **Resolução nº 13, de 9 de fevereiro de 2001**. Dispõe sobre a implantação de uma política municipal de arquivos, sobre a construção de arquivos e de websites de instituições arquivísticas. Brasília, Disponível em:

<https://www.gov.br/conarq/pt-br/legislacao-arquivistica/resolucoes-do-conarq/resolucao-no-13-de-9-de-fevereiro-de-2001>. Acesso em: 27 jul. 2024.

BRASIL. **Resolução nº 52, de 25 de agosto de 2023**. Estabelece a Política de Preservação de Websites e Mídias Sociais no âmbito do Sistema Nacional de Arquivos (Sinar). Brasília, 2023a. Disponível em:

<https://www.gov.br/conarq/pt-br/legislacao-arquivistica/resolucoes-do-conarq/resolucao-no-52-de-25-de-agosto-de-2023>. Acesso em: 27 jul. 2024.

BRASIL. **Resolução nº 53, de 25 de agosto de 2023**. Define requisitos mínimos de preservação para websites e mídias sociais no âmbito do Sistema Nacional de Arquivos (Sinar). Brasília, 2023b. Disponível em:

<https://www.gov.br/conarq/pt-br/legislacao-arquivistica/resolucoes-do-conarq/resolucao-no-53-de-25-de-agosto-de-2023>. Acesso em: 27 jul. 2024.

BUCCI, Eugênio. **Existe democracia sem verdade factual? cultura política, imprensa e bibliotecas públicas em tempos de fake news**. São Paulo: Estação das Letras e Cores, 2019. (Coleção Interrogações).

CARRETTA, Antonio Paulo. Preservação Digital: considerações sobre arquivamento sobre arquivamento de conteúdo na web. **VI Congresso Internacional em Tecnologia e Organização da Informação**, v. 6, 2020. Disponível em:

<https://toi.eca.usp.br/vittoi/article/view/34>. Acesso em: 21 jul. 2024.

CASTELLS, Manuel. **A Galaxia da Internet: Reflexões sobre a Internet, os negócios e a sociedade**. tradução: Maria Luiza X. de A. Borges. Rio de Janeiro: Zahar, 2003.

CAVALCANTE, Lidia Eugenia. **A construção do patrimônio digital: dimensões da política cultural para preservação e acesso**. 2017. Disponível em:

<http://biblioteca.pinacoteca.org.br:9090/bases/biblioteca/322696.pdf>. Acesso em: 20 maio 2024.

COELHO, Teixeira. **Com o cérebro na mão: no século que gosta de si mesmo**. São Paulo: Itaú Cultural, 2015.

COELHO, Teixeira. **Sinais e maravilhas**. São Paulo, SP: Editora Iluminuras, 2021.

CONSELHO NACIONAL DE ARQUIVOS. **Requisitos mínimos de preservação para websites e redes sociais**. Niterói, RJ: Arquivo Nacional, 2023.

CORRÊA, Amarílis Montagnolli Gomes. **Preservação digital: autenticidade e integridade de documentos em bibliotecas digitais de teses e dissertações**. 2010. - Universidade de São Paulo, São Paulo, 2010. Disponível em: Acesso em: 6 maio 2024.

COSTA, Miguel; GOMES, Daniel; SILVA, Mário. The evolution of web archiving. **International Journal on Digital Libraries**, v. 18, 2017.

DANTAS, Camila Guimarães. **Criptografias da memória: um estudo teórico-prático sobre o arquivamento da web no Brasil**. 2014. - Universidade Federal do Estado do Rio de Janeiro, Rio de Janeiro, 2014.

DAWKINS, Richard. **The selfish gene**. 30th anniversary ed. Oxford ; New York: Oxford University Press, 2006.

DODEBEI, Vera Doyle. Memória e Patrimônio: perspectivas de acumulação/dissolução no ciberespaço. **Aurora. Revista de Arte, Mídia e Política**, n. 10, p. 36–36, 2011. Disponível em: <https://revistas.pucsp.br/index.php/aurora/article/view/4614>. Acesso em: 21 jul. 2024.

DODEBEI, Vera. Memórias mediadas: histórias de vida de objetos digitais. **Études de communication**, n. 57, p. 111–130, 2021. Disponível em: <http://journals.openedition.org/edc/12425>. Acesso em: 21 jul. 2024.

FOUCAULT, Michel. **A Arqueologia do Saber**. Editora Forense, 2008.

GABLER, Louise. **Arquivo Nacional (1889-1930)**. 2018. Disponível em: <https://mapa.arquivonacional.gov.br/index.php/dicionario-primeira-republica/539-arquivo-nacional>. Acesso em: 21 jul. 2024.

GALF, Renata. Brasil engatinha em preservar arquivos virtuais históricos. **Folha de São Paulo**, São Paulo, 21 jan. 2024. Poder. Disponível em: <https://www1.folha.uol.com.br/poder/2024/01/brasil-engatinha-para-preservar-registros-virtuais-e-caso-bolsonaro-expoe-fragilidades.shtml>. Acesso em: 20 jul. 2024.

GHOSH, Pallab. Google's Vint Cerf warns of "digital Dark Age". **BBC News**, 13 fev. 2015. Science & Environment. Disponível em: <https://www.bbc.com/news/science-environment-31450389>. Acesso em: 25 jul. 2024.

GOMES, Daniel. Web archives as research infrastructure for digital societies: the case study of Arquivo.pt. **Archeion**, v. 123, p. 46–85, 2022. Disponível em: <https://www.ejournals.eu/Archeion/2022/123/art/22601/>. Acesso em: 25 jul. 2024.

GOMES, Daniel; MIRANDA, João; COSTA, Miguel. A Survey on Web Archiving Initiatives. In: GRADMANN, Stefan *et al.* (org.). **Research and Advanced Technology for Digital Libraries**. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011. (Lecture Notes in Computer Science). v. 6966, p. 408–420. Disponível em: http://link.springer.com/10.1007/978-3-642-24469-8_41. Acesso em: 23 jul. 2024.

HAN, Byung-Chul. **No enxame: perspectivas do digital**. Editora Vozes, 2021.

HEATON, Jeff. **Programming spiders, bots, and aggregators in Java**. San Francisco: Sybex, 2002.

HORAVA, Tony. Challenges and Possibilities for Collection Management in a Digital Age. **Library Resources & Technical Services**, v. 54, n. 3, p. 142–152, 2010. Disponível em: <https://journals.ala.org/lrts/article/view/5556>. Acesso em: 21 jul. 2024.

INSTITUTO DE PESQUISA DATASENADO. Pesquisa DataSenado: redes sociais, notícias falsas e privacidade de dados na internet. 2019. Disponível em: <https://www2.senado.leg.br/bdsf/handle/id/603465>. Acesso em: 25 jul. 2024.

INTERNATIONAL INTERNET PRESERVATION CONSORTIUM. **About IIPC**. 2024. Disponível em: <https://netpreserve.org/about-us/>. Acesso em: 21 jul. 2024.

INTERNET ARCHIVE. **Internet Archive**. [S. l.], 2024. Disponível em: <https://archive.org/details/texts>. Acesso em: 21 jul. 2024.

ISO/TR 14873:2013. **ISO/TR 14873:2013(en), Information and documentation — Statistics and quality issues for web archiving**. 2013. Disponível em: <https://www.iso.org/obp/ui/en/#iso:std:iso:tr:14873:ed-1:v1:en>. Acesso em: 27 jul. 2024.

JARDIM, José Maria. De que falamos quando falamos em políticas arquivísticas? Percepções sobre políticas arquivísticas no Poder Executivo Federal. **Ciência da**

Informação, v. 42, n. 1, 2015. Disponível em:

<http://172.18.0.3/ciinf/article/view/1393>. Acesso em: 21 jul. 2024.

LATIN AMERICAN NETWORK INFORMATION CENTER. **Archive of Political Parties and Elections in Latin America, APPELA**. 2024. Disponível em:

<http://lanic.utexas.edu/project/archives/appela/browse/>. Acesso em: 27 jul. 2024.

LE GOFF, Jacques. **História e Memória**. Campinas SP: Editora da UNICAMP, 1990.

LIBRARY OF CONGRESS. **ARC_IA, Internet Archive ARC file format**. 2024a. web page. Disponível em:

<https://www.loc.gov/preservation/digital/formats/fdd/fdd000235.shtml>. Acesso em: 23 jul. 2024.

LIBRARY OF CONGRESS. **WARC, Web ARChive file format**. 2024b. web page.

Disponível em: <https://www.loc.gov/preservation/digital/formats/fdd/fdd000236.shtml>. Acesso em: 23 jul. 2024.

MELO, Jonas Ferrigolo. **Arquivamento dos websites do governo federal brasileiro: preservação do domínio GOV.BR**. 2020. - Universidade Federal do Rio Grande do Sul, Rio Grande do Sul, 2020. Disponível em:

<https://lume.ufrgs.br/handle/10183/210671>. Acesso em: 21 jul. 2024.

MELO, Ivina Flores. O protagonismo do Arquivo Nacional na configuração da Arquivologia no Brasil (1838-1991). **ÁGORA: Arquivologia em debate**, v. 31, n. 62, p. 1–17, 2021. Disponível em: <https://agora.emnuvens.com.br/ra/article/view/981>. Acesso em: 21 jul. 2024.

MELO, Jonas Ferrigolo; ROCKEMBACH, Moisés. Panorama das Ações de Arquivamento da Web Governamental no Brasil e no Mundo. *In*: XXI ENCONTRO NACIONAL DE PESQUISA E PÓS-GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO, 2021. **XXI Encontro Nacional de Pesquisa e Pós-Graduação em Ciência da Informação**. 2021. Disponível em:

<https://ancib.org/enancib/index.php/enancib/xxienancib/paper/view/451>. Acesso em: 25 jul. 2024.

MILANESI, Luís. **O que é biblioteca**. 10. ed. São Paulo: Brasiliense, 1998.

NIC.BR. **Núcleo de Informação e Coordenação do Ponto BR - Início**. Disponível em: <https://nic.br>. Acesso em: 27 jul. 2024.

NTOULAS, Alexandros; CHO, Junghoo; OLSTON, Christopher. What's new on the web?: the evolution of the web from a search engine perspective. *In*: WWW04: THE 2004 WORLD WIDE WEB CONFERENCE (IN CONJUNCTION WITH ACM CONFERENCE ON ELECTRONIC COMMERCE [EC'04]), 2004, New York NY USA. **Proceedings of the 13th international conference on World Wide Web**. New York NY USA: ACM, 2004. p. 1–12. Disponível em: <https://dl.acm.org/doi/10.1145/988672.988674>. Acesso em: 10 ago. 2024.

NÚCLEO DE INFORMAÇÃO E COORDENAÇÃO DO PONTO BR; VELLOSO, Pilar. **Conectividade Significativa: propostas para medição e o retrato da população no Brasil**. tradução: Ana Zuleika Pinheiro Machado. São Paulo, SP: Núcleo de Informação e Coordenação do Ponto BR, 2024.

ORTEGA, Cristina Dotta. Relações históricas entre Biblioteconomia, Documentação e Ciência da Informação. **Datagramazero**, v. 5, n. 5, 2004. Disponível em: <https://brapci.inf.br/#/v/5664>. Acesso em: 20 jul. 2024.

PROJECT CYBERSYN. *In*: WIKIPEDIA. 2024. Disponível em: https://en.wikipedia.org/w/index.php?title=Project_Cybersyn&oldid=1237926741. Acesso em: 10 ago. 2024.

PÚBLICO. **Quanto tempo vive um “link”?** 2011. Disponível em: <https://www.publico.pt/2011/09/08/p3/noticia/quanto-tempo-vive-um-link-1811776>. Acesso em: 20 jul. 2024.

PUBLISHNEWS. **Internet Archive retira cerca de 500 mil livros do seu arquivo virtual após processos de editoras nos EUA**. 2024. Disponível em: <https://www.publishnews.com.br/materias/2024/07/01/internet-archive-retira-cerca-d-e-500-mil-livros-do-seu-arquivo-virtual-apos-processos-de-editoras-nos-eua>. Acesso em: 23 jul. 2024.

ROCKEMBACH, Moisés. Arquivamento da web: estudos de caso internacionais e o caso brasileiro. **RDBCI Revista Digital de Biblioteconomia e Ciência da Informação**, v. 16, n. 1, p. 7–24, 2017. Disponível em: <https://periodicos.sbu.unicamp.br/ojs/index.php/rdbci/article/view/8648747>. Acesso em: 27 jul. 2024.

ROCKEMBACH, Moisés; PAVÃO, Caterina Groposo. **Arquivamento da web e preservação digital**. Pimenta Cultural, 2024.

ROCKEMBACH, Moises; PAVÃO, Caterina Marta Groposo. Políticas E Tecnologias De Preservação Digital No Arquivamento Da Web. **Revista Ibero-Americana de Ciência da Informação**, v. 11, n. 1, p. 168–182, 2018. Disponível em:

<https://periodicos.unb.br/index.php/RICI/article/view/8473>. Acesso em: 21 jul. 2024.

RODRIGUES, José Honório. A situação do Arquivo Nacional. 1959. Disponível em:

<https://bdan.an.gov.br/handle/123456789/22>. Acesso em: 21 jul. 2024.

SANTOS, Paola De Marco Lopes dos. Paul Otlet: um pioneiro da organização das redes mundiais de tratamento e difusão da informação registrada. **Ciência da Informação**, v. 36, n. 2, 2007. Disponível em:

<https://revista.ibict.br/ciinf/article/view/1176>. Acesso em: 21 jul. 2024.

SATO, Mia. **Google confirms the leaked Search documents are real - The Verge**. [S. l.], 2024. Disponível em:

<https://www.theverge.com/2024/5/29/24167407/google-search-algorithm-documents-leak-confirmation?ref=hackernoon.com>. Acesso em: 20 jul. 2024.

SILVA, Armando Malheiro da. **A informação: da compreensão do fenómeno e construção do objecto científico**. Porto: Afrontamento, 2006.

SILVA, Juliane Teixeira e. Normas ISO para Gestão de Documentos: uma introdução. **Archeion Online**, v. 4, n. 1, p. 04–21, 2016. Disponível em:

<https://www.periodicos.ufpb.br/ojs/index.php/archeion/article/view/32299>. Acesso em: 21 jul. 2024.

SMIT, Johanna Wilhelmina. O documento audiovisual ou a proximidade entre as 3 Marias. **Revista brasileira de biblioteconomia e documentação**, v. 26, n. 1–2, p. 81–85, 1993. Disponível em: <https://repositorio.usp.br/item/000866736>. Acesso em: 20 jul. 2024.

THELWALL, Mike; VAUGHAN, Liwen. A fair history of the Web? Examining country balance in the Internet Archive. **Library & Information Science Research**, v. 26, n. 2, p. 162–176, 2004. Disponível em:

<https://www.sciencedirect.com/science/article/pii/S0740818804000246>. Acesso em: 10 ago. 2024.

UNESCO, 2nd. **Carta de la UNESCO para la Preservación del Patrimonio Digital**. [S. l.], 2003. Disponível em:

https://unesdoc.unesco.org/ark:/48223/pf0000229034_spa. Acesso em: 27 jul. 2024.

VERGUEIRO, Waldomiro. Desenvolvimento de coleções: uma nova visão para o planejamento de recursos informacionais. **Ciência da informação**, v. 22, n. 1, p. 13–21, 1993. Disponível em: <https://repositorio.usp.br/item/000858910>. Acesso em: 21 jul. 2024.

WEITZEL, Simone da Rocha. **Elaboração de uma política de desenvolvimento de coleções em bibliotecas universitárias**. Editora Interciência, 2013.