

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL  
FACULDADE DE BIBLIOTECONOMIA E COMUNICAÇÃO  
DEPARTAMENTO DE CIÊNCIAS DA INFORMAÇÃO  
CURSO DE ARQUIVOLOGIA

Raquel Silva da Fonseca

**A CONSTRUÇÃO DA NARRATIVA CONTEMPORÂNEA: a *web* como documento  
permanente.**

Porto Alegre

2018

Raquel Silva da Fonseca

**A CONSTRUÇÃO DA NARRATIVA CONTEMPORÂNEA: a *web* como documento  
permanente.**

Trabalho de conclusão de curso de graduação apresentado a Faculdade de Biblioteconomia e Comunicação da Universidade Federal do Rio Grande do Sul como requisito parcial para a obtenção do título de Bacharela em Arquivologia.

Orientador: Prof. Dr.º Moisés Rockembach  
Co-orientadora: Profa. Dra. Jeniffer Cuty

Porto Alegre

2018

**UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL**

Reitor: Prof. Dr. Rui Vicente Oppermann

Vice-Reitor: Prof.<sup>a</sup> Dr.<sup>a</sup> Jane Fraga Tutikian

**FACULDADE DE BIBLIOTECONOMIA E COMUNICAÇÃO**

Diretora: Prof.<sup>a</sup> Dr.<sup>a</sup> Karla Maria Müller

Vice-Diretor: Prof. Dr. Ilza Maria Tourinho Girardi

**DEPARTAMENTO DE CIÊNCIAS DA INFORMAÇÃO**

Chefe: Prof. Dr. Jeniffer Alves Cuty

Chefe Substituto: Prof. Dr. Eliane Lourdes da Silva Moro

CIP - Catalogação na Publicação

Fonseca, Raquel Silva da

A construção da narrativa contemporânea: a web como documento permanente. / Raquel Silva da Fonseca. - 2018.

67 f.

Orientador: Moisés Rockembach.

Coorientadora: Jeniffer Cuty.

Trabalho de conclusão de curso (Graduação) -- Universidade Federal do Rio Grande do Sul, Faculdade de Biblioteconomia e Comunicação, Curso de Arquivologia, Porto Alegre, BR-RS, 2018.

1. Arquivamento da web. 2. Jornadas de Junho. 3. Narrativa. 4. Memória. 5. Ética da Memória. I. Rockembach, Moisés, orient. II. Cuty, Jeniffer, coorient. III. Título.

Elaborada pelo Sistema de Geração Automática de Ficha Catalográfica da UFRGS com os dados fornecidos pelo(a) autor(a).

**Departamento de Ciências da Informação**

Rua Ramiro Barcelos, 2705

Bairro Santana – Porto Alegre – RS

CEP: 90035-007

Fone: (51) 3308 5067

E-mail: dci@ufrgs.br

Raquel Silva da Fonseca

**A CONSTRUÇÃO DA NARRATIVA CONTEMPORÂNEA: a *web* como documento permanente.**

Trabalho de Conclusão de Curso (TCC) apresentado como requisito para obtenção do título de Bacharel(a) em Arquivologia da Universidade Federal do Rio Grande do Sul.

Aprovado em: \_\_\_\_ de \_\_\_\_\_ de \_\_\_\_.

**BANCA EXAMINADORA**

---

Prof. Dr.º Fabiano Couto Corrêa da Silva - UFRGS

---

Profa. Me.<sup>a</sup> Priscila Chagas Oliveira - UFPEL

---

Prof. Dr.º Moisés Rockembach - UFRGS (orientador)

---

Profa. Dra. Jeniffer Cuty – UFRGS (co-orientadora)

“Agradeço os que vieram, cumprimento os que virão”.  
Para Niutom Barros da Fonseca (*in memoriam*).

## AGRADECIMENTOS

Agradeço ao curso de graduação em Arquivologia da UFRGS, em especial meus orientadores Moisés Rockembach e Jeniffer Cuty, por terem caminhado comigo até este momento.

Agradeço ao curso de graduação em História da Pontifícia Universidade Católica do Rio Grande do Sul pela minha primeira formação, em especial ao professor Helder V. Gordim da Silveira, responsável por ensinamentos que jamais esqueci. Fica aqui meu reconhecimento da importância dessa primeira formação, e as professoras Maria Cristina dos Santos e Janete Abrão, pessoas que admiro ainda hoje.

Agradeço à toda minha família, um pouco grande demais para ser nominada pontualmente, mas que com certeza caminharam junto comigo, dando seu apoio sempre que possível. Em especial agradeço à minha mãe Eva Eleni da Silva e ao meu irmão Daniel S. Barros da Fonseca pelo apoio incondicional até aqui.

Um agradecimento especial a Wellington Narde Navarro da Costa e família, por construir comigo uma família não tradicional brasileira. Junto com o felino Tim Lopes e o espírito do Farelo, comungamos com gatos-soldados, gatas-Frida, lagartos, gambás, selvas, filhos-planta, etc, e caminhamos pelas ruas macias de Viamão. Wellington, quando eu pensei que a cura por amar demais estava mais que suficiente, você me mostrou que o companheirismo do dia a dia seria a melhor parte de nós.

Finalizo meus agradecimentos com as palavras de Kamau: “Agradeço aos que vieram, cumprimento os que virão. Chego forte mas não passo como chuvas de verão”. [Eu vou].

## EPÍGRAFE

**“Roteiro de Chaplin”, por Coruja BCI e participação de Rael.**

*Enquanto a chuva cai, deixa tua alma ser lavada,  
Ainda busca de conquista pra quem nunca teve nada  
A fita já foi dada, lágrimas não comove  
O mundo quer resposta, mas a pergunta move  
Atalho oferecido, ingresso ao precipício  
Abraça o caminho fácil pra ver que ele é o mais difícil  
Melhor voltar pro início do que tá no trilho errado  
Atalho traz manjares mas corta aprendizado  
Não vai entrar pra história quem vive como um covarde  
Se o mar vermelho não abrir, atravesse nadando  
Se acreditasse em brincadeiras de dizer verdade  
Veria verdades, que insistem em dizer brincando  
Gratidão, demonstra a postura de um vencedor  
Quem não se curvou a estátua de Nabucodonosor  
No barco, Cristo falou: Pedro, joga a rede  
Quem não ama, é igual a profeta de Baal, morre de sede  
Os tempos já são outros, filhos quer ensinar os pais  
Pra quem a guerra é clichê, careta é falar de paz  
A inveja é pros incapaz, que morre sem companhia  
Profetize o bem, palavras nunca voltam vazias.  
Entre o espinho e a flor / Entre o medo e o amor  
Entre o preço e o valor / O que é importante é conseguir seguir  
Entre o mundo incolor, / Entre a luta e a dor,  
Entre o frio e o calor / As circunstâncias me trouxeram aqui  
Por onde eu for, eu vou sem medo  
A fé e o amor andam comigo desde cedo  
Por onde eu for eu vou sem medo.*

## RESUMO

O presente trabalho de conclusão de curso tem como objetivo principal analisar a construção da memória das Jornadas de Junho de 2013 nos Arquivos da *Internet Archive*. Delimitamos como objeto de nossa análise o arquivamento *web* do site do jornal O Estado de São Paulo, um dos veículos de imprensa nacional que produziu ampla cobertura sobre o evento. O exame do nosso objeto levou esta pesquisa a realizar um histórico sobre a *Internet Archive*, bem como produzir uma breve revisão de literatura sobre as pesquisas nacionais e internacionais acerca da tecnologia de arquivamento da *web*. Para observar as problemáticas referentes à memória, recorremos ao estudo dos conceitos de informação, patrimônio e preservação digital. Propomos um estudo do objeto a partir dos conceitos de narrativa, memória e ética da memória de Walter Benjamin. Na análise da construção da narrativa do jornal, averiguamos quais ferramentas tecnológicas foram utilizadas pela e como elas influenciaram no arquivamento. Ademais, constatamos como os excessos e perdas de informação poderão influenciar narrativas futuras sobre o evento. Apesar dos debates nacionais serem escassos em torno do tema, esse trabalho ressalta a potencialidade da Arquivística para tratar dos problemas mais contemporâneos na Ciência da Informação.

**Palavras-chave:** Arquivamento da web; Jornadas de Junho; Narrativa; Memória; Ética da Memória.



## ABSTRACT

This monography has as main objective analyze the construction of the memory of the social moviment kwon as “Jornadas de Junho de 2013” in the archives os Internet Archive. To achieve this goal, we delimited as the object of our analysis the filed newspaper “O Estado de São Paulo”, one of the national press vehicles that produced wide coverage on the event. The examination of our object led us to a bibliographic review over the Internet Archive, as well as write about its story. To observe the problematics concerning memory, we reached to the concept studies of information, patrimony and digital preservation. We propouse the study of our object in the light of the concepts of narrative, memory and ethics of memory propoused by the author Walter Benjamin. In the analysis of the construction of the newspaper narrative, we investigated what technological tolls were used and how they influenced the archiving. In addition, we verified how the excesses and losses of information can influence future narratives about the event. Although national debates are rare around the theme, this work highlights the potential of Archives to deal with the most contemporary problems in Information Science.

**Keywords:** Web Archiving; “Jornadas de Junho”; Narrative; Memory; Ethics of Memory.

## LISTA DE ILUSTRAÇÕES

- Figura 1** - Elaboração própria. Visão geral do arquivamento da url: [www.estadao.com.br](http://www.estadao.com.br) 42
- Figura 2** - Elaboração própria. Na imagem acima, podemos visualizar melhor os erros de arquivamento do IA. Para o mês analisado (junho de 2013), apenas um dia apresentou problemas 44
- Figura 3** - Elaboração própria. Detalhamento do arquivamento. Reportagem sobre as manifestações em São Paulo publicada no dia 06 de junho de 2013 53
- Figura 4** - Blog Estadão Dados do dia 11 de junho de 2013 55
- Figura 5** - Detalhamento do arquivamento da matéria "Ao vivo: manifestantes chegam à Paulista" 59

## **LISTA DE ABREVIATURAS E SIGLAS**

**IA** – Internet Archive

**CERN** - Conseil Européen pour la Recherche Nucléaire

**NUAWEB** – Núcleo de Pesquisa em Arquivamento da Web e Preservação Digital.

## SUMÁRIO

<b>1. INTRODUÇÃO</b>	12
<b>2. METODOLOGIA</b>	15
<b>3. A REALIDADE DIGITAL E AS NOVAS FORMAS DE RELAÇÕES: <i>World Wide Web e a Internet Archive</i></b>	19
<b>4. INFORMAÇÃO E MEMÓRIA</b>	30
<b>4.1. O Conceito de Informação na Ciência da Informação</b>	31
<b>4.2. Patrimônio Digital e Memória</b>	34
<b>4.3. A memória como “memória ética”</b>	36
<b>5. ARQUIVAMENTO DA WEB – O caso do jornal O Estado de São Paulo</b>	42
<b>5.1. Estrutura do Arquivamento</b>	42
<b>5.2. O arquivamento da <i>web</i> e a construção da narrativa do jornal O Estado de São Paulo</b>	52
<b>5.3. Abundância do silêncio</b>	57
<b>6. CONCLUSÃO</b>	61
<b>7. REFERÊNCIAS BIBLIOGRÁFICAS</b>	64

## 1 INTRODUÇÃO

Minha existência foi pautada pela crítica extrema e pelo amor incondicional a História. Licenciada e bacharel em História pela PUCRS, nutri um profundo amor pela disciplina, e por todas as etapas históricas. Da História Antiga à História do Tempo Presente, tudo me agradava em termos de conteúdo. Mas o fascínio nunca me impediu de ter uma postura crítica e ácida com relação aos temas mais polêmicos, pelo contrário. Quanto mais eu amava o curso da disciplina, mais eu o criticava, mais exigia posturas políticas e sociais frente aos problemas sócio-humanos colocados em sala de aula, mais eu percebia que o universo da sala de aula não era capaz de responder as demandas da sociedade atual, por mais que eu ame a posição de aluna. Terminada a graduação em História, retornei à universidade para cursar Arquivologia na UFRGS no ano de 2012, encarado como etapa necessária para finalizar meus estudos e ingressar no ambiente profissional que eu escolhi estar, os Arquivos.

Dito isso, foram as circunstâncias da minha personalidade que me impulsionaram a buscar fora da Universidade as respostas para os problemas sociais que eu via de maneira latente. No entanto, a busca por respostas me trouxe ainda mais questionamentos. O movimento político e social na cidade de Porto Alegre que tenciona a questão do transporte público iniciou muito antes de 2013. A cada aumento do valor do preço da passagem havia reivindicação: grupos sociais tomavam as ruas, vereadores pautavam a crítica ao aumento na câmara da cidade, a imprensa analisava a necessidade do aumento, etc. Havia uma certa clareza discursiva dos grupos participantes: o transporte público carecia de qualidade por não ser uma prioridade política municipal, e os grupos donos das frotas de transporte não tinham opção em função da alta dos preços dos combustíveis. Até um determinado momento, a dimensão política do problema era clara: se tratava de uma demanda municipal, em função do crescimento urbano da cidade, que deveria proporcionar aos seus moradores condições básicas para o deslocamento ao trabalho no cotidiano.

O que aconteceu em junho de 2013, as chamadas “Jornadas de Junho”, carece de resposta sócio política adequada até os dias de hoje. As disputas políticas daquele ano ainda estão em campo de batalha em nossa rotina. Uma reivindicação claramente municipal se transformou em uma bandeira nacional de disputa de interesses diversos, abarcados pelo guarda-chuva da “luta anticorrupção”. Ainda não temos distanciamento histórico do ocorrido necessário para compreender a magnitude do momento, mas podemos tecer algumas

considerações sobre as formas de recordar. A partir disso, pensemos no seguinte problema de pesquisa: que memória das Jornadas de Junho estamos resguardando?

Nosso questionamento nos levou a buscar fontes sobre o período, para refletir sobre como estamos arquivando esse momento, a memória e a fonte arquivística que preservamos. Como o acontecimento delimitado faz parte da nossa história recente, buscamos fontes e arquivos não tradicionais para a arquivística brasileira e, portanto, o jornal se apresentou como primeira fonte para pesquisa. Devido ao impacto que as redes sociais tiveram durante as manifestações, sendo colocada por muitos como a principal forma de organização para o protesto, buscamos versões virtuais dos jornais, inicialmente com a intenção de confrontar o conteúdo físico do jornal com o conteúdo *online*.

Entretanto, ao nos depararmos com a plataforma da *Internet Archive* e sua magnitude, convergimos nossa pesquisa bibliográfica para entender essa ferramenta da melhor forma possível. A primeira e maior dificuldade foi o encontro da produção bibliográfica, majoritariamente escrita em inglês, com apenas alguns artigos em português escritos no último ano acerca do tema ‘arquivamento da *web*’. A *Internet Archive* (IA) é uma plataforma criada e gestada nos EUA, fundada em 1996 com a intenção de se tornar a nova Biblioteca de Alexandria. Iniciou seus trabalhos arquivando livros disponíveis *online*, como um projeto de digitalização das obras ainda não disponíveis na *web*, para garantir acesso universal às obras. A barreira linguística é desconsiderada pelos seus idealizadores, porém a organização conta com várias parcerias de digitalização de livros históricos com o *Google* e financia trabalhos em vários locais do mundo, no continente africano, europeu e asiático. É notória a falta de participação da América Latina nesses projetos, o que justifica a baixa presença dos livros escritos em português e espanhol.

A primeira década dos anos 2000 apresentou um crescimento vertiginoso da *web*, e os idealizadores do IA anteciparam a importância que esse meio de comunicação teria no nosso cotidiano. O arquivamento dos livros logo se tornou o arquivamento da própria *web*, das páginas e conteúdos criados exclusivamente em ambiente virtual. Já na década de 1990, o IA produziu arquivamentos temáticos em torno das eleições estadunidenses, em uma época que a quantidade de conteúdo *online* era pequena e que o arquivamento intencional de conteúdo, selecionado por humanos, ainda era possível. O volume de informação presente na *web* já era considerável no início dos anos 2000, mas a tecnologia interativa da *web 2.0* tornou a produção informacional da *web* em algo imensurável. O IA manteve sua ambição primeira, recorrendo ao arquivamento

automático através de robôs para garantir a coleta e a preservação do máximo de conteúdo que lhes é possível manter. Mesmo com um arquivamento automatizado e sem seleção prévia, muita informação é perdida para sempre dentro da *web*.

Com uma produção informacional de larga escala, mesmo tendo um objeto limitado nos deparamos com um grande volume de documentação. Nos pareceu imprescindível tratar da *Internet Archive* e sua história como uma parte fundamental de nossa pesquisa, uma vez que o tema não é muito explorado em âmbito nacional. Além disso, pretendemos tratar sobre a disciplina Arquivística e a Ciência da Informação como relacionadas através do conceito de informação. Pela *web* produzir um novo conceito de suporte, que a informação é apenas visível a partir de um aparato técnico e tecnológico capaz de ler o suporte, trazemos autores que discutem a primazia da informação sobre o suporte, como o objeto chave da nossa ciência. Ainda sobre a Arquivística nacional, pontuamos suas limitações frente ao tema, porém sem fazer da ausência a principal condição da nossa narrativa. A ausência aqui será tratada como um campo aberto para possibilidades que devem ser exploradas.

Como a iniciativa foi um traço que marcou a trajetória dessa pesquisa, recorreremos a um teórico pouco utilizado nas Ciências da Informação para dialogar com o objeto através do conceito de memória e narrativa. Walter Benjamin não aparece na literatura arquivística como um autor de primeira ordem, ele é apenas mencionado algumas vezes relacionados a outros autores que tratam sobre a questão da narrativa. A abundância do tema de pesquisa corre nos dois sentidos: no excesso e na ausência/perda de informação. Escolhemos Benjamin justamente pelo objeto ter essa característica. Através do conceito de narrativa e do conceito sobre ética da memória em Benjamin, organizado pela autora psicanalista e mestre em filosofia Manuela Sampaio de Mattos, poderemos analisar os dados pela sua riqueza e por quais de seus lapsos podem influenciar na construção da memória das manifestações.

Esse trabalho busca trazer uma espécie de renovação na escrita acadêmica da área de Arquivologia e uma pequena contribuição para a Ciência da Informação. Analisando um arquivo existente, a partir da documentação que já está arquivada, iremos refletir sobre a disciplina arquivística, sobre pontos chaves da Ciência da Informação e sobre como enfrentamos a questão da memória na contemporaneidade.

## 2 Metodologia

A presente pesquisa tem por objetivo analisar qual memória foi preservada das “Jornadas de Junho” dentro da *Internet Archive*, com o propósito de analisar como se deu o arquivamento de conteúdo *web* do jornal O Estado de São Paulo, bem como as características da documentação preservada, e como sua preservação influencia na preservação da memória desse período histórico.

Os objetivos específicos da pesquisa visam:

Verificar como se realiza o arquivamento do conteúdo *online* dentro da plataforma *Internet Archive*, bem como descrever sua história;

Analisar a narrativa existente no conteúdo *web* produzido pelo jornal O Estado de São Paulo no período delimitado;

Investigar qual memória foi construída a partir do arquivamento da *web* já existente, e de que forma esse arquivamento pode compreender uma memória ética benjaminiana.

A presente pesquisa é de caráter quanti-qualitativo descritiva, que visa analisar uma amostra dos arquivos *Internet Archive*, acessados de forma livre e sem necessidade de cadastro prévio através da plataforma de busca *Wayback Machine*. A delimitação de um mês do jornal O Estado de São Paulo justifica-se pela quantidade de documentos *web* que foram arquivados nesse período: 276 no total, com uma média de 9,2 documentos *web* por dia. Todavia, existem algumas discrepâncias no conteúdo arquivado como, por exemplo, um dia com apenas um documento *web* para ser acessado enquanto outras datas contam com 18 documentos para acesso.

Através da análise dessa fonte, aliamos duas fontes arquivísticas importantes, porém pouco trabalhadas em termos teóricos conceituais pela Arquivologia no país: o jornal e a *web* como documento de arquivo. Colocamos ambas as fontes como documentos de arquivo porque elas já existem em arquivos. No caso da *web*, a arquivística nacional precisa se apropriar dessa forma de arquivamento maciça que já é feito em larga escala nos países no norte do mundo<sup>1</sup>.

---

<sup>1</sup> Ao falar “países do norte do mundo”, nos referimos a distinção econômica que divide o mundo em países desenvolvidos e subdesenvolvidos, conhecidos como o “norte” e o “sul” do mundo, respectivamente.



No primeiro levantamento para a pesquisa, foi averiguado o processo “visível”<sup>2</sup> de arquivamento. Primeiro coletamos dados gerais sobre as capturas em todos os dias do mês de junho de 2013, formando uma tabela (Ver capítulo 5.1.) com as seguintes informações: data, total de *snapshots*<sup>3</sup>, *link* da última captura<sup>4</sup> e observações. Nesse primeiro levantamento, obtivemos 276 capturas no total, sendo o dia 20 de junho de 2013 a data com mais capturas no total (18) e o dia 30 de junho de 2013 com a menor taxa de captura (01). Mesmo com valores discrepantes, pudemos concluir uma média de 9,2 capturas por dia ao longo de todo o mês.

Ao trabalhar na construção das observações sobre cada capa de jornal<sup>5</sup>, percebemos que o formato do arquivamento provocava desdobramentos na pesquisa. Os *snapshots* não são equivalentes ao número de *hiperlinks*<sup>6</sup> disponíveis para cada data. Isso fez com que procurássemos entender como funcionou as capturas de conteúdo a partir de cada capa, verificando quais reportagens tiveram seus conteúdos capturados assim como a página principal. Cada reportagem da capa principal gerou um *hiperlink* específico, ou seja, para entender esse arquivamento como sendo de qualidade, é preciso considerar quais dos *hiperlinks* da capa foram resguardados bem como a primeira página do jornal. Dito isso, precisamos considerar que ao arquivar um documento de *web* já estamos lidando também com a potencial falta de informação, caso algum desses *hiperlinks* não estejam disponíveis no repositório do *Wayback Machine*.

Verificando a totalidade do conteúdo das datas, percebemos que alguns dos *hiperlinks* da capa do jornal redirecionavam para um arquivamento muito mais maciço daquele feito na capa principal. Existem *hiperlinks* de reportagens específicas do jornal que contam com várias capturas, algumas com mais de 50, em um período de tempo que oscila da data da reportagem até o ano de 2017. Ou seja, os *snapshots* de algumas reportagens continuaram sendo realizados por, pelo menos, 4 anos depois de sua publicação *online*. Ao considerar esse desdobramento

---

<sup>2</sup> “Visível” porque não temos acesso ao *wayback machine* como estrutura administrativa. As informações foram coletadas no site <https://archive.org/web/> de forma livre e gratuita, sem a necessidade de cadastro prévio.

<sup>3</sup> *Snapshots* é o termo em inglês utilizado para designar a captura da página *web*. Durante o texto, utilizaremos o termo em inglês e a respectiva palavra em português “captura” para nos referirmos ao material arquivado no site.

<sup>4</sup> Consideramos que o último link disponível das capturas realizadas em cada dia, individualmente, corresponde a edição “mais acabada” do jornal.

<sup>5</sup> Consideramos a página inicial do site [www.estadao.com.br](http://www.estadao.com.br) como a capa do jornal. Seria o equivalente a capa física do jornal.

<sup>6</sup> *Hiperlink* ou apenas *link* significa ligação. Dentro do contexto da linguagem de programação, *hiperlink* é uma referência de um documento em hipertexto, que liga um documento a outros. Fonte: <https://sites.google.com/site/sitesrecord/o-que-e-um-hiperlink>.

apenas das reportagens relacionadas a política nacional, acrescentamos 216 capturas para o *corpus* documental.

Com um *corpus* documental de quase 500 capturas<sup>7</sup>, optamos por fazer uma análise amostral dessas alterações e discrepâncias, analisando pontos distintos em algumas publicações do jornal dentro da delimitação estabelecida. Esses pontos foram escolhidos a partir da captura maciça de outros *hiperlinks* além da capa principal, tornando o jornal mais completo em termos de informação. As datas escolhidas foram: 06 de junho de 2013; 07 de junho de 2013; 11 de junho de 2013; e 17 de junho de 2013.

O terceiro capítulo deste trabalho visa explicitar o crescimento da *web* e como se formou a *Internet Archive* e o *Wayback Machine*. Através de uma revisão de literatura estadunidense e brasileira, iremos discorrer sobre a *web 2.0*, sobre as formas de arquivamento da *Internet Archive*, sobre os problemas e falhas, sobre as implicações em pesquisas e fontes, e sobre os debates atuais acerca do tema. Esse capítulo será a base para um bom entendimento da análise das fontes, uma vez que demonstra as potencialidades e problemas que podem ser encontrados no arquivamento.

O quarto capítulo trata do conceito de informação e memória na Ciência da Informação e na Arquivística. Os autores Capurro e Hjørland (2007) realizaram uma extensa revisão sobre os usos do conceito de ‘informação’, não apenas na CI, mas nas ciências humanas e exatas. Para eles, ‘informação’ é por vezes utilizada sem uma conceituação precisa, sendo determinada em última instância a partir das necessidades e questionamentos de uma pesquisa. Encarando a imprensa como fonte para o conhecimento sobre o passado, é preciso iniciar a reflexão sobre os sentidos possíveis da informação dentro do contexto nacional. A análise do jornal O Estado de São Paulo referente ao período das Jornadas de Junho nos mostrará qual narrativa e qual memória foram construídas para esse acontecimento. Segundo Benjamin (2012), o processo narrativo tem um claro cunho pedagógico acerca da memória que se quer resguardar. Todo o processo narrativo guarda uma intencionalidade, modifica-se ao longo do tempo, e seu caráter pedagógico serve como ferramenta de organização do social. O autor afirma que a imprensa é o instrumento de comunicação mais importante no alto capitalismo, tendo um impacto na narração e na memória tão fundamental quanto o romance. Benjamin afirma que a imprensa criou a forma de comunicação através da informação (BENJAMIN, 2012).

---

<sup>7</sup> É possível considerar que cada captura é um documento de arquivo diferente.

No quinto capítulo apresentaremos nossa análise do objeto delimitado, a partir das ferramentas teórico conceituais referidas nos capítulos anteriores. A fim de explicarmos nosso material empírico, utilizaremos imagens das páginas *web* do IA como instrumento pedagógico para narrar como o arquivamento é realizado. Iremos expor uma tabela com os dados quantitativos colhidos sobre as capturas do jornal O Estado de São Paulo. A seguir, trabalharemos com a construção da narrativa do jornal, e avaliaremos como algumas escolhas influenciaram no arquivamento dos dados. Além disso, refletiremos sobre alguns aspectos qualitativos, principalmente no que tange o processo de rememoração do evento.

### 3 A REALIDADE DIGITAL E AS NOVAS FORMAS DE RELAÇÕES: *World Wide Web e a Internet Archive*

Para compreender como ocorre o arquivamento da *web* é preciso entender como a *web* surgiu. As implicações sobre a *web* como documento estão relacionadas ao funcionamento da mesma, seguindo uma lógica organizativa diferente das organizações sociais físicas. O ambiente digital abriu um mundo de relações colaborativas onde a informação e o conhecimento são construídos coletivamente, gerando documentos dinâmicos e mutáveis no decorrer do tempo.

Neste capítulo, faremos um breve levantamento sobre o tema da *web*, seu crescimento como espaço de comunicação e produção de conhecimento para, a seguir, expor a fundação da *Internet Archive* e seu funcionamento, bem como o processo de arquivamento da *web* e suas implicações.

A *World Wide Web*, ou apenas *web*, foi criada e desenvolvida pelo engenheiro e cientista da computação Timothy John Berners-Lee, a partir de uma proposta para a manutenção geral de informação sobre experimentos realizados pela CERN<sup>8</sup>, a Organização Europeia para Pesquisa Nuclear<sup>9</sup>. A primeira página da *web* mundial foi lançada em dezembro de 1990, e é possível ser acessada ainda hoje<sup>10</sup>. Segundo Rockembach (2018), a primeira página da *web* e seus *hyperlinks* são preservados por um projeto desenvolvido pela própria CERN. A ideia central do grupo é garantir o crescimento da *web* aberta e colaborativa (ROCKEMBACH, 2018).

A *web* está prestes a fechar 30 anos de existência, um curto período histórico, porém com vasta produção documental, talvez a maior produção de informação já vista pela humanidade. A *web* produz documentos fluídos, em constante mudança, cria documentos digitais e fontes que despertam o interesse dos mais variados grupos. A publicação de livros em meio digital fez das bibliotecas as primeiras instituições a se preocupar com a *web*, suas formas de armazenamento e produção de metadados sobre o objeto guardado, além de delimitar formas de arquivamento a partir do conteúdo de interesse de uma instituição<sup>11</sup>. É a partir do interesse

<sup>8</sup> É possível verificar partes do projeto no site <http://info.cern.ch/Proposal.html>. Acessado em 08 de junho de 2018.

<sup>9</sup> <https://home.cern/>. Acessado em 08 de junho de 2018.

<sup>10</sup> <http://info.cern.ch/hypertext/WWW/TheProject.html>. Acessado em 08 de junho de 2018.

<sup>11</sup> A Biblioteca do Congresso dos EUA é um exemplo. A instituição participa do *International Internet Preservation Consortium* (IIPC), junto com outras instituições de bibliotecas e arquivos de todo o mundo. Apesar da reunião de iniciativas mundiais, é visível a concentração de países localizados ao norte da divisão socioeconômica do mundo, tendo apenas a Biblioteca Nacional do Chile como a única instituição localizada na

de bibliotecas nacionais que o arquivamento da *web* começa a ganhar maior sistematização (BRÜGGER, 2011).

O'Reilly afirma que, no ano de 2001, a *web* solidificou sua importância social e tecnológica, mesmo quando alguns prognósticos da época apontavam para a supervalorização e queda dessa ferramenta (O'REILLY, 2005). A *web 2.0* surgiu como conceito numa conferência entre O'Reilly e a *MediaLive Internacional*, a partir da discussão sobre a perda de importância dos domínios *dot com*<sup>12</sup>, ao mesmo tempo que outras formas de relações estavam ganhando importância através da *web* no início dos anos 2000 (O'REILLY, op. cit., p. 225). Ao lado de Dale Dougherty, O'Reilly e *MediaLivre Internacional* fundaram uma conferência chamada *Web 2.0*. Em sua primeira reunião, no ano de 2004, pontos importantes foram estabelecidos, marcando a forma como entendemos a *web* até os dias atuais. A *web 2.0* não é um conceito com demarcações rígidas, sendo melhor entendida como um núcleo de princípios e práticas com movimentação gravitacional, onde o primeiro princípio é da “*web* como plataforma” (Op. cit., p. 226).

O autor Alex Primo, ao comentar estas propostas de O'Reilly, afirma que a *web 2.0* “ênfatiza o desenvolvimento do que se chama ‘arquitetura de participação’”, onde a tecnologia incorpora características de interconexão e compartilhamento (PRIMO, 2007, p. 2). Além disso, o segundo princípio chave colocado por O'Reilly sobre a *web 2.0* é que “os serviços tornam-se melhores quanto mais pessoas o usarem” (PRIMO, 2007, p. 3). Primo ainda elenca as formas de crescimento da *web 2.0*, descrevendo a importância dos *blogs* como espaço de conversação, as interconexões das “micro-redes”, as novas formas de circulação da informação<sup>13</sup>, processos coletivos para organização e recuperação de documentos eletrônicos<sup>14</sup>, enfim, recursos que servem a “gestão coletiva do trabalho comum” (PRIMO, op. cit., p. 6).

---

região sul do mundo. O IIPC foi fundado em 2003 com 12 instituições nacionais, contanto hoje com mais de 50 instituições participantes.

<sup>12</sup> *Dot dom* ou *.com*, é uma companhia que trabalha na *internet*, notadamente durante a década de 1990, em que transformou *.com* num domínio popular da *web*. A queda do domínio *.com* referida pelo autor O'Reilly versa sobre as empresas que gestavam seus negócios através da *web*. No início dos anos 2000, houve uma queda nas vendas das empresas relacionadas a esse domínio, enquanto outras formas de relações sociais estavam se estabelecendo como fundamentais na época, notadamente os *blogs*. Para maiores informações ver: [https://en.wikipedia.org/wiki/Dot-com\\_company](https://en.wikipedia.org/wiki/Dot-com_company). Acessado em 08 de junho de 2018.

<sup>13</sup> O exemplo colocado por Primo são o *Real Simple Syndication*, ou *RSS*, que é um sistema de assinatura de conteúdos onde o internauta pode escolher quais informações quer receber automaticamente em um *software* agregador que ele escolha, numa espécie de *clipping* automatizado (PRIMO, 2007, p. 4).

<sup>14</sup> Aqui, Primo menciona o *social bookmarking*, que é um registro em que sites são destacados como “favoritos”. A seleção dos “favoritos” não é realizado de maneira passiva, os internautas criam metadados que identificam e qualificam o material selecionado, formando “etiquetas” (“*tags*”) que referenciam o material. O processo em

A IA é uma instituição sem fins lucrativos fundada em 1996 por Brewster Kahle, localizada em São Francisco, EUA, com a finalidade de construir uma biblioteca digital de *websites* e outros objetos culturais em formato digital. A missão da instituição é “promover acesso universal a todo o conhecimento”<sup>15</sup>. No ano de criação, a instituição arquivava a internet em si, que ainda estava em fase de crescimento de uso. A instituição compara o conteúdo da *web* com a efemeridade do jornal em formato impresso, sinalizando a importância do resgate do suporte ao mesmo tempo que representa um objeto com poucas iniciativas para preservação. Assim, a IA preconizou o único arquivamento da *web* em configuração transnacional<sup>16</sup>, protegendo uma parte de conteúdo *web* mundial. São 20 anos de história da internet, disponíveis através da plataforma *Wayback Machine* e acessíveis a todos<sup>17</sup>. O volume informacional trazido com a *web 2.0*, e a mudanças com relação à qualidade do conteúdo, faz com que o *Wayback Machine* seja o maior arquivo da *web* disponível hoje, arquivando diferentes formas de mídia, como áudios, vídeos, páginas *web*, fotografias, livros, conteúdos interativos, etc., somando, até 2015, 438 bilhões de páginas *web*, ocupando 23 petabytes de espaço de armazenamento (WEBER IN BRÜGGER, SCHOREDER et al., 2017, p. 84).

A estratégia básica de arquivamento do *Internet Archive* é via *snapshots*, ou “capturas”, em que todo o material recolhido pelos *crawlers*<sup>18</sup>, uma ferramenta robótica que arquivava material a cada 8 semanas. O IA começou com uma pequena coleção selecionada em torno de eventos políticos, arquivando as páginas dos candidatos à Presidência dos EUA no ano de 1996 (BRÜGGER, SCHOREDER et. al., 2017, p. 06). Devido ao volume de informação adicionada a proposta inicial do grupo jamais esquecida, não existe uma seleção temática nem uma organização formal para o conteúdo arquivado, não há qualquer organização arquivística ou biblioteconômica dos dados que auxilie na recuperação da informação. Esse é o maior problema

---

que o internauta pode colocar qualquer palavra que julgarem estar relacionadas com o material é chamado de “folksonomia” (PRIMO, 2007, p. 5).

<sup>15</sup> Ver <https://archive.org/about/>. Acessado em 08/06/2018. Tradução livre.

<sup>16</sup> Brügger foi quem cunhou o termo “transnacional” para se referir ao início do arquivamento da *web* proposto pela *Internet Archive*. Segundo o autor, o *Internet Archive* tinha como proposta mostrar o valor potencial que as *web pages* tinham numa época em que esse valor não era evidente. Por esse motivo, o escopo da *Internet Archive* sempre foi arquivar a *web* mundial, e não apenas o conteúdo produzido nos EUA (BRÜGGER IN CONSALVO, 2011, p. 31).

<sup>17</sup> A Internet Archive é um dos membros fundadores do IIPC, consórcio mencionado anteriormente.

<sup>18</sup> *Crawler* ou *web crawler* é um robô da internet que auxilia na indexação da *web*. Eles rastreiam uma página *web* em determinado momento até que todo o conteúdo da mesma seja indexado. A página *web* para ser rastreada e indexada por um *crawler* precisa ter em seu html códigos que permitam a indexação. É em função da variação de códigos html que alguns *hyperlinks* não são indexados pelo *crawler*. Para mais informações, acessar o site: <https://www.techopedia.com/definition/10008/web-crawler>. Acessado em 08 de junho de 2018.

atual da plataforma *Wayback Machine*: é necessário que se tenha a *url*<sup>19</sup> de uma página que deseja ser recuperada para buscar no banco de dados (BRÜGGER, SCHOREDOR et. al., 2017, p. 11).

Portanto, é necessário realizar uma busca consciente sobre o conteúdo, ter uma página *web* em mente e saber a *URL* correta da mesma para que ocorra uma recuperação efetiva no banco de dados. A autora Jinfang Niu, escreveu em 2012 um texto didático sobre o arquivamento da *web* e seus diferentes formatos, para os cursos de graduação relacionados com Biblioteconomia e Arquivologia nos EUA (NIU, 2012). A autora afirma que o *Internet Archive* trabalha com arquivamentos em massa, lidando com um grande volume de informações, mas que o repositório captura apenas as páginas que estão na superfície da *web* (NIU, 2012). A hierarquia da *web* relaciona-se com a importância que cada *web page* (ou site) tem na internet, baseando-se no número de acessos, alcance e interação que uma página tem com os internautas, empresas, entidades governamentais e não-governamentais, etc. A seleção de conteúdo arquivado no *Wayback Machine* é, geralmente, escolhido a partir dessa importância, ainda tendo certo espaço para a seleção humana. Essa seleção humana de conteúdo é, em grande medida, feita pelos internautas que acessam e conhecem a página do IA. Qualquer pessoa que conhece o *Wayback Machine* e gostaria de arquivar uma página que não consta em seus arquivos pode fazê-lo sem muitos problemas, bastando inserir a *url* que deseja arquivar na página inicial no setor “*save this page*”. Logo, a seleção de grande parte do conteúdo do *Internet Archive* não passa pelo crivo humano, sendo realizada pelos *crawlers*.

Segundo Niu,

Esse método fundamenta-se na coleta de conteúdo das páginas web através dos *crawlers*. *Crawlers* utilizam uma lista de *seed* para iniciar o download de conteúdo web, e acompanha os *hyperlinks* para descobrir novos conteúdos para download. As decisões da seleção são baseadas na compilação de uma lista *seed* e na configuração do parâmetro do *crawler*. (NIU, 2012, p. 04, tradução nossa)<sup>20</sup>

Os parâmetros de captura do *crawler* podem acarretar em limitações no arquivamento, não coletando algumas informações que podem estar numa página, como por exemplo, vídeos,

<sup>19</sup> *Url*, ou *Uniform Resource Locator*, é o endereço de uma página na internet, bem como o seu protocolo de acesso. Ver: <https://www.techopedia.com/definition/1352/uniform-resource-locator-url>. Acessado em 05 de outubro de 2018.

<sup>20</sup> “This method relies on *crawlers* to harvest content from web servers. *Crawlers* use a *seed list* to start downloading web content, and follow the *hyperlinks* to discover and download additional web content. Selection decisions are the basis for compiling a *seed list* and configuring *crawler* parameters.” (NIU, 2012, p. 04)

conteúdos dinâmicos e *streaming media*<sup>21</sup>. Por ser um processo automatizado, o *crawling* pode arquivar conteúdo repetido, produzindo arquivos duplicados. *Crawlers* como o *Heritrix*, usado pelo *Wayback Machine*, tem uma funcionalidade inteligente que reduz o problema, mas não o elimina. No entanto, Niu aponta que a captura em larga escala de sites frequentemente atualizados causa imprecisões temporais nos documentos arquivados (NIU, op. cit.).

Cada página *web* possui uma estrutura hierárquica interna, delimitada pela forma como os componentes e subcomponentes do site são organizados de maneira relacional (NIU, op. cit.). Considerando nosso objeto de pesquisa, cada capa do jornal apresenta os conteúdos abordados pelo órgão jornalístico em um dia, e o conteúdo relacionado com a capa são os *hiperlinks* para cada reportagem, com textos, imagens, vídeos e demais conteúdos interativos. A relação com a disposição espacial feita por O'Reilly anteriormente não se dá ao acaso: cada página *web* se relaciona com outras páginas *web*, formando uma constelação de informação e documentação imbricada e complexa. O conteúdo de um jornal não está totalmente disposto em um único documento, mas em uma série de diferentes documentos conectados entre si. Para entender um documento, é preciso ter uma visualização do total de informação que toda essa rede é capaz de oferecer. Segundo Niu, essa relação entre os *hiperlinks* pode ser comparada com a teoria da organização interna de documentos da arquivística tradicional, através do princípio da organicidade (NIU, op. cit.).

A captura em massa realizada pelo IA pode gerar um contexto histórico de produção de uma página *web*. A repetição de capturas demonstra como uma página alterou seu conteúdo ao longo do tempo, dando a perspectiva histórica de produção, mostrando quais informações foram lançadas primeiro, quais foram alteradas e como o produto final se apresentou (NIU, op. cit.). Os autores Niel Brügger e Ralph Schoreder, ao oferecerem um panorama da história da *web*, comentam o caso do Partido Conservador do Reino Unido que, em 2013, deletou discursos controversos de seu site e impediu o acesso dos mesmos conteúdos arquivados em sites do *Google* e no IA. Gerando uma discussão sobre a manutenção de conteúdo *web* como “secreto”, uma investigação sobre arquivamentos de conteúdo semelhante mostrou que as páginas deletadas foram arquivadas também por outra organização, sendo mantidas numa coleção especial da Biblioteca Britânica (BRÜGGER, SCHOREDER et. all., 2017). O arquivamento

---

<sup>21</sup> *Streaming media* é um formato de entrega de conteúdo com várias multimídias, usualmente áudio e vídeo, para determinados usuários da *web*. Esses usuários são selecionados por um serviço de informação de *streaming*, que recolhe informações básicas de internautas através dos protocolos: HTTP, TCP/IP e HTML. Para maiores informações visitar a página: <https://www.techopedia.com/definition/14586/streaming-media> Acessado em 05 de outubro de 2018.



mostrou como o site do partido foi alterado, quais discursos foram omitidos, além de atrair atenção negativa sobre o conteúdo do site (Ibidem, p. 01).

Este caso também mostra como se dá o marco legal de acesso e uso do conteúdo arquivado no IA. Segundo Niu, o marco legal está fundamentado no país em que o arquivo está hospedado (NIU, op. cit., p. 08), mesmo quando o documento arquivado foi produzido em outro país. O IA tem um mecanismo robótico de entrada e saída, em que é permitido a exclusão de conteúdo, caso seja requisitado (NIU, op. cit., p. 5). Foi assim que o Partido Conservador do Reino Unido pode retirar parte do seu conteúdo arquivado no IA, sem comprometer a totalidade do arquivo relacionado ao partido resguardado no repositório.

A sistematização em torno de um pensamento e diretrizes institucionais sobre arquivamento da *web* não tem estado na agenda das tarefas do Conselho Nacional de Arquivos do Brasil (CONARQ). Houve algumas iniciativas que tocam no assunto, porém um programa nacional de arquivamento da *web* ainda não foi lançado. O Arquivo Nacional tem um Programa de Preservação e Acesso a Documentos Arquivísticos Digitais – AN Digital, iniciado em 2010, com o objetivo de capacitar a instituição em infraestrutura organizacional e tecnológica para os documentos digitais recolhidos e preservados pela entidade. No segundo volume da publicação “Políticas de Preservação Digital”, editado em 2016, afirma-se que “em momento futuro, outros tipos mais complexos de documentos em formato digital, como multimídia e página web, deverão ser também contemplados” (ARQUIVO NACIONAL, 2016, p. 11 & ROCKEMBACH, 2018, p. 14). A arquivística brasileira trabalha num eterno contexto de crise, o que justifica algumas lacunas e a conservação de um pensamento acadêmico majoritariamente conservador em termos científicos. Até o momento da redação desse trabalho de conclusão de curso, não foram encontradas dissertações ou teses sobre o tema, apenas alguns artigos científicos que tratam do arquivamento da *web*. Artigos lançados em 2018 por Rockembach e Pavão são pioneiros sobre o tema no Brasil, oferecendo um panorama geral sobre o assunto (ROCKEMBACH, 2018; ROCKEMBACH, PAVÃO, 2018; ROCKEMBACH, 2017; FERREIRA, MARTINS, ROCKEMBACH, 2018). Encontramos o trabalho monográfico de Natali Reigert, sobre arquivamento e preservação da *web*, defendido no ano de 2017 (REIGERT, 2017).

Acreditamos que a falta de uma iniciativa nacional sobre o arquivamento da *web* brasileira não impede os Cientistas da Informação de produzirem pesquisas sobre o assunto. É possível dialogar em períodos de crise com a crise, e os hiatos devem nos impulsionar a

construir objetos de pesquisa que nos façam refletir sobre nosso ofício. Na ausência de um arquivamento da *web* de iniciativa nacional, aos moldes observados em outros países, buscamos por instituições que realizam arquivamentos transnacionais, com o objetivo de tomar conhecimento sobre o que da *web* brasileira está sendo guardada em outros arquivos.

A discussão em torno dos novos formatos documentais tem se centrado no documento digital, e textos que consideram o arquivamento da *web* como objeto de análise são escassos em língua portuguesa. Rockembach fez um levantamento de pesquisas nacionais e internacionais sobre o tema, recuperando artigos científicos da base de dados *Scopus* (2018). Segundo o autor, a qualidade da coleta é fundamental para um melhor arquivamento e recuperação da informação (2018), aponta que:

Conforme colocado por Hockx-Yu (2012), isto significaria tentar armazenar de forma idêntica ao que vemos ao acessar diretamente um site web, contudo, por uma série de fatores, como scripts dinâmicos, streaming de mídia, estruturas das redes sociais e conteúdo baseado em banco de dados, torna-se necessário garantir a qualidade a partir de quatro aspectos também enquadrados por Hockx-Yu (2012), que trazem mais ênfase ao conteúdo do que ao visual gráfico: I. Se o conteúdo pretendido foi coletado integralmente; II. Se o conteúdo intelectual, em oposição ao estilo e layout, pode ser reproduzido na ferramenta de acesso; III. Se a cópia coletada pode ser reproduzida, incluindo o comportamento presente no site ao vivo, como a capacidade de navegar interativamente entre links; IV. Se há a manutenção da aparência de um site. (ROCKEMBACH, 2018, p. 06-07)

Rockembach ainda comenta sobre as implicações legais acerca do arquivamento, afirmando que os usos da informação arquivada implicam não somente em problemas legais mas também éticos, que não estão explícitos numa página *web* (ROCKEMBACH, op. cit., p. 09). De acordo com o autor

Uma das alternativas legais e éticas a esta coleta e uso do arquivamento da *web*, mesmo não possuindo os direitos autorais sobre os sites, acontece por meio do fair use, ou uso justo, termo usado na legislação americana e que tem relação com a tradição da Commonlaw, aplicada ao uso de conteúdo sob direitos autorais, desde que em certas situações, como o uso pedagógico/educacional, ou como notícia ou ainda como pesquisa, por exemplo. (ROCKEMBACH, 2018, p. 09)

Buscamos ainda literatura científica estadunidense para complementar informações e proposições analíticas sobre o arquivamento da *web*. O livro *The web as History* (2017), editado ano passado pela UCL, é a primeira obra de fôlego sobre o tema (BRÜGGER, SCHOREDER

et. all., 2017). Seleccionamos três capítulos desse livro para orientar nossa pesquisa. A introdução do livro, escrita pelos organizadores da obra, Niel Brügger e Ralph Schoreder, oferece um panorama sobre a história da *web*. Os autores mencionam a importância de uma informação *online* em exemplos de como a *web* é usada como prova em processos jurídicos ou são relevantes em movimentos políticos (BRÜGGER, SCHOREDER, p. 01). Afirmam também a importância da *Internet Archive* para o arquivamento (ibid, p. 07), e questiona como os arquivos da *web* são constituídos, como eles são coletadas, como são organizados e como tornar essa informação disponível a pesquisadores (ibid, p. 9-10). O texto também oferece uma revisão bibliográfica importante para a constituição de uma historiografia digital no futuro (ibid, p.14).

O segundo capítulo do livro selecionado é escrito por Matthew S. Weber, intitulado “*The tumultuous history of news on the web*”. Weber tem o jornal como fonte principal de pesquisa, e oferece um panorama do desenvolvimento dos sites de notícias na *web* (WEBER In: BRÜGGER, SCHOREDER et. al., 2017). Devido a fluidez do conteúdo da *web*, Weber salienta que a informação jornalística pode ser facilmente suprimida, alterada ou completamente deletada (Op.cit., p. 85). Além disso, o autor comenta o impacto da *web 2.0* para a interatividade e produção de conteúdo jornalístico, afirmando que o jornal virtual atinge um período de “maturidade” em 2010, passando a produzir conteúdo exclusivo *online* (Ibidem, p. 89). O estudo do autor é relevante para considerarmos a evolução do site resgatado no arquivo, já que trabalha com páginas de jornais.

O último capítulo selecionado do livro é da autora Jane Winters, intitulado “*CODA: Web archives for humanities research – some reflections*”. Winters estabelece uma reflexão a partir do questionamento sobre como alguém irá pesquisar sobre a história do mundo desenvolvido no final do século XX e início do XXI sem utilizar a *web* arquivada como fonte (WINTERS In: BRÜGGER, SCHOREDER et. al., 2017). A autora trabalha com pesquisadores das ciências humanas e faz um apanhado de questionamentos que cientistas e estudantes do futuro podem fazer com relação a *web* enquanto fonte de pesquisa. A data de publicação de uma página, os problemas do processamento de arquivamento, se o processo de arquivamento de uma instituição é completamente transparente com relação a captura, como um pesquisador pode lidar com a incompletude de uma fonte digital (Op. cit., p.240-244). Winters oferece caminhos metodológicos da historiografia para lidar com esses problemas. Ela atesta que já podemos falar de uma “idade das trevas digital”, uma vez que já perdemos grande parte do conteúdo disponível na *web*. No entanto, a autora aponta que historiadores da História Medieval

e História Oral trazem ferramentas metodológicas importantes que podem ser utilizadas para solucionar esses problemas (ibid, p. 244).

Tendo em mente a necessidade de refletir sobre essa forma de arquivamento, procuramos um objeto de pesquisa que contemplasse essa demanda. Apesar de inexistir uma iniciativa nacional de arquivamento da *web*, páginas de *web* produzidas no Brasil têm sido arquivadas em diferentes repositórios internacionais. Os autores Schroeder e Brügger afirmam que existe um crescimento da *web* em língua espanhola, no Brasil e na Índia, fazendo com que os arquivistas da *web* contemplem essa produção e suas implicações (BRÜGGER, SCHOREDER et. al., 2017, p. 03). Em função do IA realizar um arquivamento transnacional em massa e sem uma seleção fechada de conteúdo, foi necessário escolher uma página *web* brasileira que alcançasse um dos requisitos básicos da coleta automatizada do IA. Para este fim, selecionamos um veículo de comunicação em massa, que produzisse conteúdo constante ao mesmo tempo que tivesse um número de acessos que colocasse a página em questão na superfície da *web*.

O jornal O Estado de São Paulo atingiu esses requisitos. Cada característica sobre o IA e o arquivamento da *web* podem gerar pesquisas científicas ricas para a arquivística, e é capaz de colocar a produção acadêmica brasileira em contato direto com as discussões mundiais sobre o tema. Apesar do Brasil não ter uma iniciativa própria de arquivamento do conteúdo *web* construído no país, nós produzimos um volume de material que é percebido pelos países desenvolvidos. A ausência do arquivo não será tomada nessa pesquisa como narrativa da crise ou de falha da disciplina, mas como uma forma de pensarmos sobre os arquivos, seu conteúdo e o como isso pode refletir nas nossas formas de recordar o passado.

A proteção dos documentos com valor de prova tem sido uma das pedras basulares da organização arquivística. Esse valor muitas vezes está relacionado com a produção institucional, o cotidiano de trabalho de um indivíduo cumprindo suas funções profissionais. Entretanto, essa característica se manifesta ao longo do tempo, bem como seu valor histórico social. Os Arquivos resguardam todo tipo de documento que representam esses aspectos, e quando a vigência das suas atribuições administrativas e legais expirou. O valor histórico e social de jornais e outras fontes impressas fez dessas produções peças presentes nas instituições arquivísticas. Complementando questões trazidas sobre o conceito de informação, Rondinelli (2013), em seu estudo sobre documentos arquivísticos frente a realidade digital, relaciona esse conceito com o conceito de documento. A autora resgata o autor Buckland (1991 Apud

RONDINELLI, 2013), para recolocar o conceito de documento sob a luz do conceito de informação, em que o autor afirma:

(...) considerar qualquer coisa informativa documento é consistente com as origens e uso inicial da palavra, a qual vem do verbo latino *docere*, ensinar ou informar, com o sufixo *mento* que denota meios. Assim, documento, originalmente, denota um meio de ensinar ou informar, seja uma lição, uma experiência ou um texto. (BUCKLAND, 1991, APUD RONDINELLI, 2013, p. 40)

Com o propósito de unir diversas percepções sobre o conceito de documento, Rondinelli (2013) se aprofunda em teorias arquivísticas que consideram documento todo e qualquer testemunho da atividade humana (p. 46). A autora também recorre a Antonia Heredia Herrera:

Documento, em um sentido muito amplo e genérico, é todo registro de informação independentemente de seu suporte físico. Abarca tudo o que pode transmitir o conhecimento humano: livros, revistas, fotografias, filmes, microfilmes, [...] mapas, [...] fitas gravadas, discos partituras, [...] selos, medalhas, quadros [...] e, de maneira geral, tudo que tenha um caráter representativo nas três dimensões e que esteja submetido à intervenção de uma inteligência ordenadora. (HERRERA, 1991, p. 121 Apud RONDINELLI, 2013, p. 13)

Frente aos conceitos amplos de documento e considerando que o conceito de informação se delimita a partir dos questionamentos feitos a ele (CAPURRO, HJORLAND, 2007, p. 187-188), entendemos nessa pesquisa o jornal como uma fonte primária, e não apenas bibliográfica. Compreendemos que os Arquivos são compostos em grande medida por documentos oriundos das instituições públicas, todavia parte da massa documental é formada por jornais e periódicos antigos, tomadas como fontes primárias por grande parte dos Cientistas Humanos. Analisaremos no capítulo 05 como o jornal em formato *web* se constitui, confeccionado nosso objeto de pesquisa, e não tomamos os textos do jornal como fonte bibliográfica informativa.

O valor histórico testemunhal dos documentos de arquivo são avaliados em relação direta com as noções de memória que os arquivistas, historiadores e profissionais do arquivo tem no momento da avaliação no processo de gestão documental. O Brasil é marcado pela desorganização e falta de iniciativas pontuais de organização de arquivos, o que implica em massas documentais diversas. O que tem acontecido no âmbito nacional é que quando uma geração peca em organizar um acervo, a próxima se vê na obrigação de resguardar muito mais do que apenas os documentos da história de uma instituição, com uma intenção subjetiva de um

dia compreender inclusive a nossa omissão frente a esses documentos e a nossa história. Numa profissão forjada na crise, seus profissionais passam a não mais teme-la, pois a crise institucional faz parte do seu cotidiano. Iniciativas de pesquisa e trabalho sobre o arquivamento da *web* tem surgido em grupos de pesquisa da Universidade Federal do Rio Grande do Sul. O Núcleo de Pesquisa em Arquivamento da Web e Preservação Digital (NUAWEB)<sup>22</sup>, criado em 2017, é pioneiro na investigação das características do arquivamento da *web* no país, além de orientar a comunidade científica nacional para a necessidade de políticas públicas que envolvam a questão.

No próximo capítulo, vamos explorar os conceitos de informação e memória, um pouco de suas construções para a Ciência da Informação, e como o conceito de narrativa e ética da memória em Walter Benjamin podem ser utilizadas como ferramentas de análise para pesquisa arquivística.

---

<sup>22</sup> Para maiores informações, acessar o site: <https://www.ufrgs.br/nuaweb/>. Acessado em 27 de novembro de 2018.

#### 4 Informação e Memória.

A Arquivística manteve por muito tempo suas preocupações centradas no conceito de “registro humano”, sendo este colocado como “remédio para a lembrança” (DURANTI, 1994, p. 50). O registro humano se dá a partir da união entre informação e um determinado suporte físico, dando ao Arquivista a obrigação de proteger fisicamente esse suporte, bem como a informação contida no mesmo.

Durante o século XX a informação passou a ser transmitida através de meios eletrônicos, modificando as formas de comunicação e interação entre as pessoas. O registro da informação passou a ser realizado por esse meio, divergindo da ideia do suporte tradicional. O *boom* informacional trazido pela Segunda Guerra Mundial fez com que Cientistas da Informação, bibliotecários e arquivistas considerassem diferentes suportes como parte do acervo preservado por esses profissionais. A mutabilidade da tecnologia, segundo Duranti, faz com que os arquivistas valorizem o documento em si e sua produção, para que o gerenciamento dos documentos eletrônicos não tome por base os recursos tecnológicos (DURANTI Apud RONDINELLI, 2005, p. 18). Sobre a importância dos registros documentais, Duranti afirma:

A razão pela qual eles puderam servir a tantas finalidades é que os materiais arquivísticos, ou registros documentais, representam um tipo de conhecimento único: gerados ou recebidos no curso das atividades pessoais ou institucionais, como seus instrumentos e subprodutos, os registros documentais são as provas primordiais para as suposições ou conclusões relativas a essas atividades e às situações que elas contribuíram para criar, eliminar, manter ou modificar. A partir dessas provas, as intenções, ações, transações e fatos podem ser comparados, analisados e avaliados, e seu sentido histórico pode ser estabelecido. (DURANTI, 1994, p. 50)

A arquivística contemporânea argumenta que o caráter testemunhal do registro humano é fundamental na era da tecnologia, passando assim a valorizar o conceito de informação arquivística, que está para além da ideia de registro humano. Silva et al., (1999) afirmam que armazenar, adquirir e recuperar informação são as três funções fundamentais dos serviços relacionados com o tratamento da informação, que são as bibliotecas, os centros de documentação e os arquivos. Os autores argumentam sobre a importância da linguística no estudo da informação, para a compreensão da formação das mensagens e símbolos que registramos, bem como para o entendimento da relação entre mensagem, símbolo, informação e memória. Sobre a recuperação da informação, os autores afirmam:

O tratamento da informação, no sentido técnico do termo, visa precisamente a criação de ‘memórias’, passíveis de serem utilizadas sempre que houver necessidade de recuperar dados (informação) nelas armazenados. Isto implica procedimentos de controlo da informação, de criação de meios de acesso às referidas memórias e de desenvolvimento de dispositivos susceptíveis de acionar os meios de acesso, com vista à recuperação da informação armazenada. Tais procedimentos são, naturalmente, objecto do trabalho dos profissionais que desempenham funções no mais diversos sistemas de informação. (SILVA et al., 1999, p. 27)

Segundo os autores, apesar da Arquivística ter a sua própria evolução enquanto ciência, seus fundamentos podem ser encontrados em ciências pré-existentes, as quais evoluíram e traçaram caminhos paralelos sem perder autonomia (SILVA et al., 1999). Definindo a informação como o objeto principal, porém impreciso, da Ciência Arquivística, os autores propõem uma aproximação necessária com a História e a Sociologia como ferramenta metodológica para delimitar o objeto de pesquisa (Op. cit., p. 36).

#### **4.1. O conceito de Informação na Ciência da Informação.**

Capurro é um dos autores clássicos no estudo do conceito de informação dentro da Ciência da Informação. Em texto escrito em conjunto com Hjørland, os autores registram a importância do conceito de informação dentro da disciplina, e como esse conceito é trabalhado em outras ciências (CAPURRO, HJORLAND, 2007). Segundo os autores, “informação” contém uma característica específica para a matéria, estando intimamente ligado com o conceito de comunicação do conhecimento, referindo-se ao processo de transformação do conhecimento, sua seleção e interpretação dentro de um determinado contexto (CAPURRO, HJORLAND, 2007). Ao estudar os usos históricos do conceito, os autores também apontam para o protagonismo dos bibliotecários sobre a apropriação do tema a partir de sua disciplina, que através da Biblioteconomia especializada e Documentação, fundaram espaços e centros profissionais com o nome de “ciência da informação” (Op. cit., p. 177). No entanto, Hjørland afirma que o uso do termo informação está ligado ao crescente interesse nas formas de utilizar as novas tecnologias computacionais para o processamento da mesma, voltando sua preocupação para as bibliotecas especializadas e de pesquisa, desenvolvimento de base de dados e as atividades ligadas à organização e à disseminação da literatura científica (Ibidem).



É dessa forma que o conceito de “recuperação de informação” se torna uma das bases da Ciência da Informação.

Entretanto, os autores apontam que em todas as áreas do conhecimento humano em que ‘informação’ é tomado como conceito, ele implica em uma atividade intelectual multi ou interdisciplinar, inclusive para as ciências naturais (Ibidem, p. 166). Os autores resgatam a formulação clássica proposta por Shannon, afirmando:

“Shannon correlaciona informação - isto é, o número de escolhas possíveis a fim de criar uma mensagem – e incerteza. Quanto maior a liberdade de escolha, maior a incerteza, isto é, a informação. O conceito de informação parece, como observa Weaver, decepcionante e bizarro – decepcionante porque não tem nada a ver com o significado e bizarro porque lida não com uma única mensagem, mas, em vez disso, com o caráter estatístico de todo um conjunto de mensagens, bizarro também porque nestes termos estatísticos as duas palavras informação e incerteza são parecidos.” (SHANNON e WEAVER, 1972, p. 27 Apud CAPURRO, HJORLAND, 2007, p. 163)

Dentro das perspectivas colocadas pelos autores, nos parece mais importante atentarmos para o conceito de informação nas ciências humanas e sociais. Na década de 1950, “informação” teve um papel central para a Psicologia, nos estudos sobre evolução cognitiva e dentro do paradigma do processo de evolução (Ibidem). Os autores comentam a decepção em torno do conceito, predominando uma abordagem funcionalista acerca do mesmo, colocando os processos cognitivos humanos como análogos ao processamento de informações praticado por computadores, não elaborando o conceito de forma explícita (Ibidem). Recapitulando Shannon, os autores fixam uma característica trazida por ele, o sentido da seleção da informação, que está em concordância com o sentido de comunicação de conhecimento (Ibidem). A questão da seleção de informação traz para a análise o receptor da informação, ou o usuário. Capurro e Hjørland falam sobre teorias que divergem acerca do papel do receptor na comunicação de conhecimento. A teoria semântica de Dretske faz uma distinção entre informação e significado, entendendo que a informação não requer um processo interpretativo para existir, embora esse ponto seja fundamental para aquisição de conhecimento (Ibidem). Ainda sobre Dretske, um ponto que nos parece fundamental para a pesquisa, quando o autor afirma que “não há informação falsa, mas há significado sem verdade” e, do mesmo modo, “informação é capaz de produzir conhecimento e uma vez que o conhecimento requer verdade, a informação também a requer” (DRETSKE Apud CAPURRO & HJORLAND, 2007, p. 170).

Capurro & Hjørlando dialogam ainda com a tecnologia associada ao acesso, que hoje acesso ao mundo tecnológico e virtual é compreendido como acesso à informação. Os autores recorrem a Romm quando comentam sobre as implicações éticas acerca do trabalho do cientista da informação, colocando que:

Romm (1997), mostra sérias implicações éticas que estão envolvidas na definição de alguma coisa como factual em oposição a pleno (p. 185) de significado. Na medida em que a informação é vista ou apresentada e legitimada em termos de seu suposto conteúdo factual, fica autorizada uma determinada visão de mundo – em vez de convidar ao debate sobre a construção e relevância desta visão. (Ibidem, p. 184-185)

Para os autores, em última instância, o que conta como “informação” ou “informativo” para determinada pesquisa ou ciência depende da questão a ser respondida, e que o trabalho em torno da ‘informação’ deve estar baseado nos problemas, questões e objetivos que a informação deverá satisfazer (Ibidem).

As mudanças tecnológicas têm produzido novas relações sociais e novas demandas para os Arquivistas. Até agora, no Brasil, a *web* tem sido pensada como uma potente ferramenta para a difusão dos arquivos e seus acervos, mas em poucos momentos a *web* é pensada como uma fonte documental passível de ser arquivada.

Para complementar a necessidade de conceituar a informação como peça fundamental da Arquivística enquanto ciência, existe ainda a necessidade de se pensar as novas configurações arquivísticas que a tecnologia tem imposto aos profissionais do arquivo. A autora Fernanda Monteiro, parafraseando a autora Heymann, reflete sobre a questão afirmando:

De acordo com a autora, há uma nova configuração dos lugares de arquivamento, “alterando a paisagem arquivística”. Considerando o fato de que os arquivos e suas funções ainda são desconhecidos por muitas pessoas em nossa sociedade, estamos em um momento no qual novos conjuntos documentais são requisitados e documentos nunca antes imaginados passam a ser arquivados. Esse cenário aponta para uma nova relação entre arquivos e sociedade, na qual haveria uma espécie de “mão dupla” em que cada grupo passa a investir na sua memória por meio da produção de documentos e de espaços de arquivamento, da mesma forma que os arquivos que já existem passam a ser mais atrativos para os diversos setores sociais. (MONTEIRO, 2014, p. 316)

Apesar de ser um conceito basilar na Ciência da Informação, “informação” não pode ter um único significado para a ciência. A multi ou interdisciplinaridade e as necessidades

apresentadas em determinada pesquisa são as únicas certezas no momento de construir o termo ‘informação’ como conceito. Devido às mudanças tecnológicas, os arquivistas modificaram sua prioridade do resguardo do suporte documental para a informação contida nele. Enquanto profissionais da informação, precisamos refletir sobre as novas configurações dos locais de arquivamento, como bem colocou a autora Heymann (Apud Monteiro, 2014), para que possamos ponderar também sobre nosso papel como produtores e protetores da memória.

#### **4.2. Patrimônio digital e memória.**

Dodebei (2006 & 2008) traz uma discussão em torno da *web* e do espaço digital como novos objetos da memória dentro da Ciência da Informação, além de conceituar o virtual como patrimônio do século XXI. A autora resgata conceitos de informação e memória, afirmando a necessidade de conjunção das CI com áreas do conhecimento antes não trabalhadas. O conceito de memória aparece como o ponto de intersecção entre diversos setores do conhecimento humano. A autora afirma que a *web* é um espaço mítico da memória em função da relação do centro (espaço fixo) e espaço (local onde o movimento se realiza) e como essa relação se dá a partir da circularidade do tempo, conduzindo a ideia de um só conceito de espaço/tempo (DODEBEI, 2006, p. 4). Com relação à preservação das memórias que circulam no espaço *web*, a autora referencia os trabalhos de Luis Fernando Sayão que atentam para o risco da amnésia digital (Op. Cit., p. 8).

Acerca do virtual como patrimônio, Dodebei afirma que patrimônio deve ser entendido enquanto objeto e enquanto um valor de informações sobre o objeto, seja ele de natureza material ou imaterial (DODEBEI, 2008). Relacionando o conceito de patrimônio com as novas relações suscitadas pela *web*, a autora propõe uma mobilidade do conceito de patrimônio:

1) o patrimônio existe como valor necessário à produção de subjetividades e à garantia da diversidade, ao mesmo tempo que é uma possibilidade de resistência à globalização cultural; 2) as condições de ser patrimônio na contemporaneidade são dadas pela tecnologia intelectual da simulação ou ampliação, quer dizer, pela atribuição constante de conteúdos informacionais ao núcleo do objeto simulado no ciberespaço; 3) a constituição do patrimônio digital como valor informacional possibilita a convivência entre suas condições de circunstancialidade e permanência, em constante tensão criadora; e 4) patrimônio, nos sentidos que a sociedade lhe confere, é um valor que não se sustenta no mundo virtual. (DODEBEI, 2008, 12).

Dodebei salienta que a tecnologia não pode ser considerada novidade em nossa realidade social. O novo trazido pela tecnologia atual são as formas como ela é utilizada para transmissão do saber. A autora propõe categorias de observação de natureza espaço-temporal, sendo a primeira categoria utilizada para estudar as formas de transmissão do saber entre gerações, podendo ser subdividida em três grupos: a oralidade, a escrita e a imagem (Ibidem). O relato mítico é o representante principal da oralidade, de cunho pedagógico e memorialístico, que serve para regulamentar a vida cotidiana. A autora recorre a Jean-Pierre Vernant para explicar os aspectos da oralidade, em que afirma:

O primeiro deles diz respeito à autoria. De acordo com Vernant, o relato mítico ‘não resulta da invenção individual nem da fantasia criadora, mas da transmissão da memória’. Assim, o mito só vive, se for contado, de geração em geração, na vida cotidiana; as condições de sua sobrevivência são, justamente, a memória, a oralidade e a tradição. O segundo aspecto se refere à sua integridade informacional ou unicidade. O mito não está fixado numa forma definitiva. Esta sempre varia em decorrência do agente – ‘contador de histórias’, aedo – ou do processo, isto é, a transmissão, que abandona algumas analogias e incorpora outras, tecendo-se, assim, um *continuum* formado por lembranças e esquecimentos, permanentemente atualizado na memória coletiva. (DODEBEI, 2008, p. 15)

Conforme visto, a oralidade mítica desempenha um papel dinâmico, dentro do tempo e espaço, estabelecendo uma zona de crescimento e mutação que variam de acordo com o narrador e com um propósito pedagógico final. A escrita, segunda tecnologia de transmissão, tem um caráter mais estático, com temporalidade linear, mudando a configuração da memória social. De acordo com Dodebei, “na tecnologia da escrita seu conceito [memória social] depende da possibilidade de crítica ligada a uma separação parcial do indivíduo e do saber, com ênfase no estatuto da verdade” (Ibidem, p. 17). O polo da escrita se concentra na fundamentação teórica e na interpretação, e a autora o relaciona com uma distinção criada por Pierre Lévy de que esse período histórico corresponde a era informática midiática.

A última forma de transmissão do saber, a infografia, pode ser considerada uma evolução da escrita, da mesma forma que apresenta uma ruptura na estabilidade da mesma. Signos imagéticos constituem a forma do relato atual, e abre espaço para a possibilidade de autoria múltipla, o que não era possível na transmissão escrita (Ibidem). Para a autora:

Tal preocupação, marcada pelo espírito de cientificidade em busca de consistência teórico-metodológica, foi desencadeada justamente pela inconsistência, a inconstância, a mutação e a obsolescências das informações lançadas na rede mundial de comunicação. A forma paradoxal da memória

social na atualidade pode ser descrita nestes termos: a um só tempo plural, como processo em permanente construção, passível de múltiplas interferências, e singular, como um único conjunto, ou forma do hipertexto. (DODEBEI, 2008, p. 19)

Segundo Dodebei, o que muda em relação à autoria, ao formato e à relação espaço-tempo entre os dois polos (oralidade e escrita) “é a velocidade das ações sociais mediatizadas pela dimensão da técnica” (Ibidem).

Patrimônio foi um conceito gerado com os Estados Nacionais do século XVIII, com a incumbência primeira de salvaguardar a memória nacional. A organização em torno dessa memória tinha claro cunho pedagógico: a necessidade de construir uma narrativa dita nacional em contraposição ao considerado diferente e estrangeiro. No presente, o conceito de patrimônio tem sido cunhado como forma de resistência para as identidades que tiveram seu espaço negado dentro da identidade nacional. Enquanto o conceito de memória retém uma clara subjetividade, o patrimônio confunde-se com o conceito de propriedade (GONÇALVES, 2003). Todavia, a colocação do espaço virtual como passível de ser patrimonializado indica que o valor informacional de um documento virtual pode ter apelo memorialístico forte o suficiente para determinado grupo social.

#### **4.3. A memória como “memória ética”**

Em um breve estudo sobre o conceito de memória na Ciência da Informação, as autoras Oliveira e Rodrigues (2011) assinalam pontos pertinentes para a nossa pesquisa. Analisando como o conceito de memória é utilizado em teses e dissertações em CI<sup>23</sup> no país, as autoras apontam que os trabalhos utilizam os termos ‘informação’, ‘documento’ e ‘memória’ como os principais conceitos de análise, e que o termo ‘memória’ aparece ligado a algum qualitativo, como ‘social’, ‘operária’, ‘clandestina’, ‘governamental’, ‘coletiva’, ‘oficial’, etc (OLIVEIRA, RODRIGUES, 2011, p. 320). Os trabalhos relacionam ‘memória’ e ‘informação’ associadas ao

---

<sup>23</sup> As autoras ressaltam que quase inexistiam Programas de Pós-Graduação em Ciência da Informação no Brasil quando realizaram a pesquisa e, portanto, consideraram enquanto pós-graduação trabalhos realizados em programas que estavam ligadas a uma das disciplinas componentes da CI, ou seja, a biblioteconomia, a museologia e a arquivologia. Dentro da pesquisa foram analisados trabalhos que foram defendidos em Programas de História Social, por exemplo, mas com ênfase em arquivologia e organização documental (DE OLIVEIRA & RODRIGUES, 2011, 319).

‘documento’, enquanto que os trabalhos que versam sobre a ‘preservação da memória’ são vinculados aos locais físicos socialmente instituídos e legitimados para sua custódia, como as bibliotecas, os museus, os arquivos e os centros de documentação (Ibidem, p. 325). As autoras concluem que o conceito de memória não foi ainda trabalhado de forma efetiva dentro da disciplina, sendo utilizado apenas como um conceito genérico que não é questionado ou reformulado dentro dos objetivos específicos da área (Ibidem, p. 327).

Abriu-se, então, uma necessidade para o presente trabalho, o de buscar conceitos de memória pouco utilizados pelas disciplinas da CI para aprofundar o diálogo interdisciplinar a fim de estabelecer uma nova abordagem teórico conceitual sobre memória para a arquivística. Diante disso, Walter Benjamin<sup>24</sup> se apresenta como um autor interessante para a análise e para nosso contexto histórico atual. Exponente do materialismo histórico dialético, Benjamin enfrentou o nazi fascismo europeu com uma escrita militante, fragmentária e potente, ligado diretamente ao extremo do social em que estava inserido naquele momento. Benjamin delimitou um caminho pouco seguido pelos marxistas até aquele momento: frente aos autoritarismos societários, ele buscou nas artes e na literatura sua fonte principal de pesquisa, criticando severamente a narrativa histórico linear baseada no progresso e na história dos vitoriosos, tecendo conceitos em torno das palavras “despertar”, “consciência”, “experiência” e “narratividade”.

Benjamin morreu sem deixar uma obra escrita por completo em torno de suas pesquisas, porém sua escrita fragmentária pode ser tomada por vários pontos em contextos diferentes mantendo suas criações cheias de significado. O primeiro texto que nos parece fundamental para a análise é “O Narrador”, escrito em 1936, onde o autor anuncia a morte de uma narrativa de cunho pedagógico, em que o narrar tinha por base fundamental ensinar através da experiência (BENJAMIN, 2012). Segundo o autor, “é a experiência de que a arte de narrar está

---

<sup>24</sup> “Walter Benjamin, nascido em Berlim, a 15 de julho de 1892, foi um dos mais notáveis intelectuais alemães do século XX. Graduado em Filosofia pela universidade de Freiburg im Breisgau, doutorou-se em 1919 com a tese O conceito de crítica de arte no Romantismo alemão. No início dos anos 1920, em meio à efervescência cultural e às turbulências políticas da República de Weimar, aproximou-se de Adorno e Horkheimer, então jovens pensadores empenhados na crítica da cultura e da razão capitalistas. Essa aproximação e o diálogo intenso que travou ao longo de toda a vida com esse grupo de teóricos, em especial com Adorno, não impediram que elaborasse uma reflexão original e particularíssima – nem sempre bem aceita pelo meio universitário germânico” (BENJAMIN, 2012). Benjamin era judeu e em função da perseguição nazista aos judeus na Europa, empreendeu uma fuga a pé com um grupo de amigos até o município fronteiriço de Port Bou. Temendo ser capturado pelos nazistas e com a fronteira da cidade fechada para a passagem dos “sem nacionalidade” (condição em que Benjamin se encontrava), suicidou-se com uma dose letal de morfina em 26 de setembro de 1940 (MATTOS, 2016, p. 13).

em vias de extinção. São cada vez mais raras as pessoas que sabem narrar devidamente” (BENJAMIN, 2012, p. 213), e complementa:

Com a guerra mundial começou a ornar-se manifesto um processo que segue desde então ininterrupto. Não se notou, ao final da guerra, que os combatentes voltavam mudos do campo de batalha; não mais ricos, e sim mais pobres em experiência comunicável? E o que se derramou dez anos depois, na enxurrada de livros sobre a guerra, nada tinha em comum com uma experiência transmitida de boca em boca. E não havia nada de anormal nisso. Porque nunca houve experiências mais radicalmente desmentidas que a experiência estratégica pela guerra de trincheiras, e experiência econômica pela inflação, a experiência do corpo pela batalha material e a experiência moral pelos governantes. Uma geração que ainda fora à escola num bonde puxado por cavalos encontrou-se desabrigada, numa paisagem em que nada permanecera inalterado, exceto as nuvens, e, debaixo delas, num campo de forças de torrentes e explosões destruidoras, o frágil e minúsculo corpo humano. (BENJAMIN, 2012, p. 214)

É a experiência do ‘boca a boca’ no qual recorrem os narradores, e as melhores narrativas escritas são aquelas que diferem menos das histórias contadas pelos inúmeros narradores anônimos (Ibidem). Percebemos aqui uma aproximação com a oralidade mítica trazida por Dodebei (2008). A riqueza da narrativa perdida estava na multiplicidade de seus autores, que mantêm amplo número de ensinamentos justamente por serem diversos. Além disso, para a narrativa manter seu caráter pedagógico através da experiência ela precisa de espaços não lineares de tempo para manter sua compreensão contemporânea. Benjamin (2012) afirma que os narradores mantêm relação direta com o senso prático, uma utilidade clara para o presente, que pode consistir em um ensinamento moral ou uma sugestão prática (BENJAMIN, 2012, p. 216). Sobre o ato de “dar conselhos”, ele afirma:

O conselho tecido na substância da vida vivida tem um nome: sabedoria. A arte de narrar aproxima-se de seu fim porque a sabedoria – o lado épico da verdade – está em extinção. Mas este é um processo que vem de longe. E nada seria mais tolo do que ver nele um ‘sintoma de decadência’, e muito menos de uma decadência ‘moderna’. Ele é muito mais um sintoma das forças produtivas seculares, históricas, que expulsam gradualmente a narrativa da esfera do discurso vivo, conferindo, ao mesmo tempo, uma nova beleza ao que está desaparecendo. (BENJAMIN, 2012, p. 217)

Benjamin (2012) trata de diferentes formas de comunicação humana no texto do ‘Narrador’. Entre elas, o romance moderno, que ganhou destaque com o advento da imprensa que auxilia, segundo o autor, na decadência da narrativa, pois o romance está apartado da oralidade e essencialmente atrelado ao livro. Para ele, o romance não procede da tradição oral

nem a alimenta, executando uma experiência alienada da realidade (Ibidem, p. 217). Ademais, o autor delinea o tempo de transformação de cada narrativa, sendo que a oralidade se alimenta da experiência. Dentre as formas de comunicação humana, ele cede um espaço para a informação<sup>25</sup>, afirmando que ela é tão estranha à narrativa quanto o romance (Ibidem, p. 218).

Segundo o autor:

Villemessant, o fundador do *Figaro*, caracterizou a essência da informação com uma fórmula famosa. ‘Para meus leitores’, costumava dizer, ‘o incêndio num sótão do Quartier Latin é mais importante que uma revolução em Madri’. Essa fórmula lapidar mostra claramente que o saber que vem de longe encontra hoje menos ouvintes que a informação que forneça um ponto de apoio para o que está próximo. (...) A informação, porém, aspira a uma verificabilidade imediata. Para tal, ela precisa ser, antes de mais nada, ‘compreensível em si e para si’. Muitas vezes não é mais exata que os relatos antigos. Mas enquanto esses relatos recorriam frequentemente ao miraculoso, é indispensável que a informação soe plausível. Nisso ela se revela incompatível com o espírito da narrativa. Se a arte da narrativa é rara, a difusão da informação tem uma participação decisiva nesse declínio. (BENJAMIN, 2012, p. 218-219)

Vimos até aqui pontos relevantes do texto “O Narrador”, delimitando como Benjamin (2012) entende as diferentes formas de comunicabilidade e como cada uma delas tem sua importância para o entendimento da realidade social. A tradição oral, a que perdura no tempo, tem um sentido pedagógico constante quando narrado no presente. O romance e a informação não resguardam o ensinamento através da experiência pois suas narrativas se afastam da realidade concreta das pessoas. A autora Manuela Sampaio de Mattos (2016) argumenta em sua obra sobre a ética da memória em Walter Benjamin que a temática do sonho e despertar é central no pensamento benjaminiano. Ao analisar a obra *Passagens*, coletânea de textos iniciais de Benjamin organizada de maneira póstuma, Mattos mostra que Benjamin afirmava que a realidade moderna se tornou um mundo de sonho (MATTOS, 2016), e que sonhamos coletivamente o sonho do progresso. A autora argumenta que “é possível visualizar que o culto permanente do capitalismo é exercido através de um ‘sonhar coletivo’ e ‘inconsciente’” em um duplo sentido: sonha porque está distraído e inconsciente porque acredita que o sonho mercadológico é único e pessoal (Op. cit., p. 32).

O despertar do sonho imposto pelo capitalismo só pode ser empreendido pelo historiador materialista, que deve empreender uma crítica ao conceito de progresso, que para

---

<sup>25</sup> Segundo Gagnebin, Benjamin refere-se a informação jornalística. (GAGNEBIN in BENJAMIN, 2012, p. 14).



manter o estado da realidade imutável fundamenta a ideia de catástrofe (Op. cit., p. 35). Complementando com dialética, Mattos (2016) recorre ao autor Buck-Moss que afirma que “a teoria é única em sua abordagem da sociedade moderna, pois leva a cultura de massa a sério, não meramente como uma fonte da fantasmagoria do mundo social, mas como uma fonte de energia coletiva capaz de superá-la” (BUCK-MOSS, apud MATTOS, 2016, p. 38-39). Considerando que o conceito de progresso delimita formas limitadas de existência e interpretação, Benjamin (2012) utiliza metáforas relacionadas com fragmentos (‘fiapos’, ‘farrapos’) para compor uma interpretação que considera o micro espaço social potente para interpretar o macro. Pires (2014) afirma que “o conhecimento é pensado por Benjamin de modo não linear; como uma paisagem urbana, a partir de lugares diferentes, fragmentariamente, nas reconfigurações da memória; não a partir de um lugar fixo, mas movendo-se em uma constelação de ideias.” (PIRES, 2014, p. 815). Sobre a “tarefa de escovar a história a contrapelo”, Pires afirma:

Ao colocar-se a tarefa de "escovar a história a contrapelo", o filósofo propõe-se desconstruir a historiografia e os métodos tradicionais de pesquisa a partir de um olhar atento sobre as transformações históricas da percepção humana; sobre as ruínas da modernidade e os estilhaços urbanos das metrópoles; sobre os atos de barbárie que se cometem em nome do progresso – os quais ele presenciou na iminência dos catastróficos acontecimentos europeus da Segunda Guerra. (PIRES, 2014, p. 815)

Mattos (2016) consolida sua tese da ética da memória benjaminiana percorrendo o conceito de “farpa messiânica” cunhado por Reyes Mate (MATTOS, 2016, p. 69). Para Mate, a “farpa messiânica” é o processo no qual o passado retorna ao presente em lampejos, traumatizando o agora (Idem). Em outras palavras, o trauma não narrado retorna constantemente ao presente como gatilhos de alguém que foi profundamente violado. O terror reaparece no agora como lampejos, porque no despertar ecoa o grito mudo dos vencidos. Reyes Mate afirma que para Benjamin “o limite e a ambição de sua filosofia política: ela propõe o direito à felicidade de quem não a tem, sabendo que só pode exigi-la” (MATE apud MATTOS, 2016, p. 70). A memória ética antevista por Mattos parte dessa memória involuntária, e da concepção de rememoração, a partir da imagem dialética que interrompe a imobilidade produzida pelo sonho do progresso (Ibidem).

Essas ferramentas conceituais auxiliarão na análise da memória digital que está em construção através das plataformas de arquivamento. O diálogo aberto entre Ciências da Informação e Ciências Sociais nos permitirá refletir, inclusive, sobre as disputas políticas que

acontecem no cenário contemporâneo do Brasil. No próximo capítulo, veremos as imobilidades informacionais na fonte arquivada, do mesmo modo em que é possível verificar a constante movimentação e modificação do documento *web*.

## 5 O ARQUIVAMENTO DA WEB – O caso do jornal O Estado de São Paulo.

Ao nos depararmos com um arquivamento da *web* referente ao jornal O Estado de São Paulo realizado em larga escala pela *Internet Archive* e disponibilizado de forma gratuita e livre, elaboramos uma forma de manejar o objeto de pesquisa. A delimitação para pesquisarmos um mês do jornal nos pareceu razoável para o escopo de um trabalho de conclusão de curso. Fizemos um levantamento inicial do nosso objeto dentro do mês de junho de 2013, somando 276 documentos arquivados, referentes aos links disponíveis para as capas do jornal em todos os dias do mês.

### 5.1. Estrutura do Arquivamento.

Nesse primeiro momento, observamos que o número de capturas não é idêntico ao número de links disponíveis para cada dia, sendo que o número de capturas é sempre maior do que o número de links disponíveis. Colocando a *url* no buscador do *Wayback Machine*, a tela de apresentação dos dados aparece em destaque como a seguir:

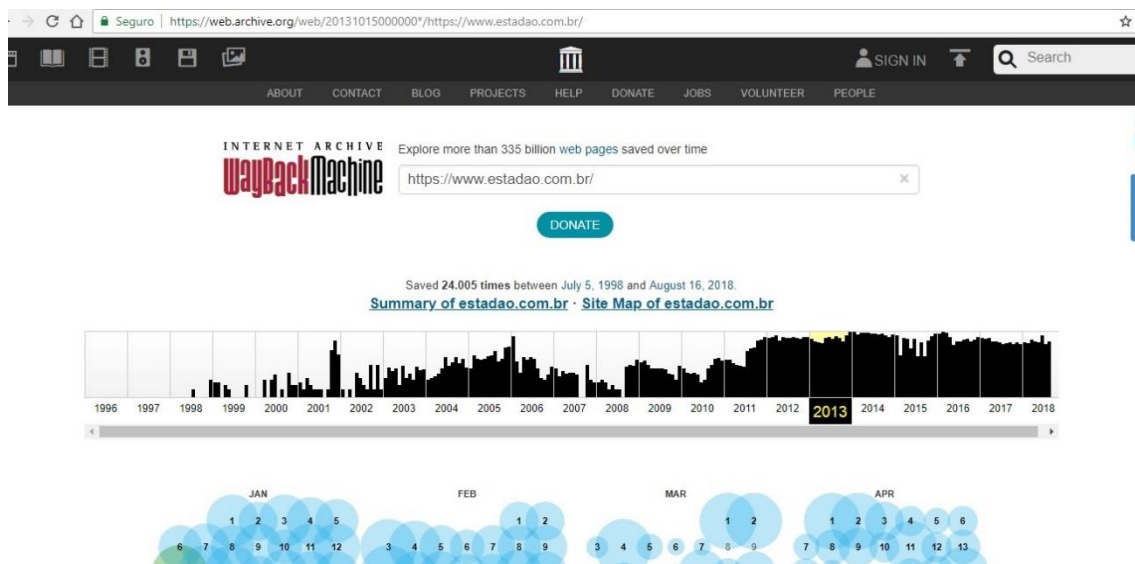


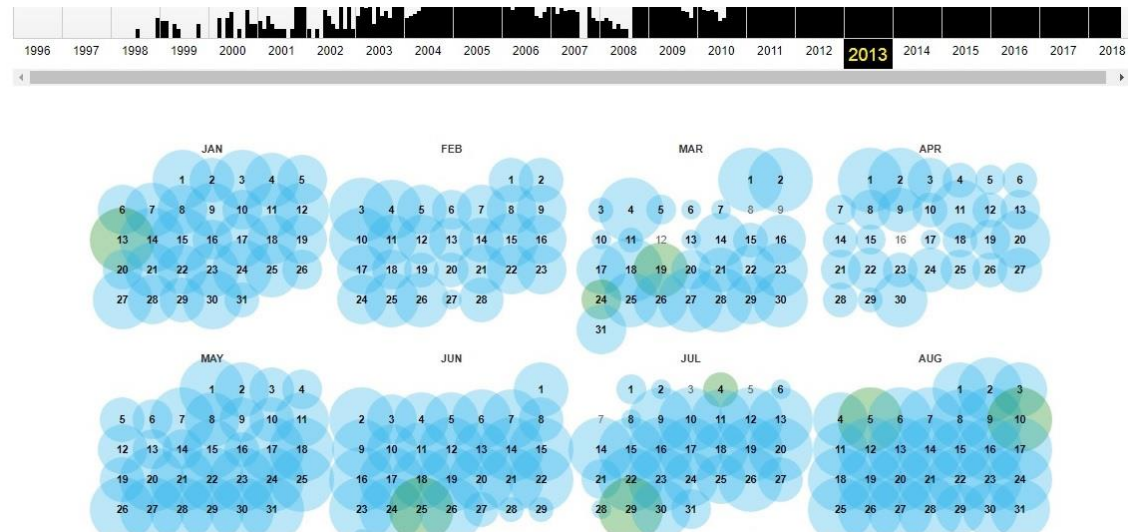
Figura 1: Elaboração própria. Visão geral do arquivamento da url: [www.estadao.com.br](http://www.estadao.com.br).

Como podemos observar, o *Wayback Machine* oferece um panorama geral de todos os arquivamentos realizados da *url* fornecida. Através da imagem, percebemos que existe uma estimativa de arquivamentos por anos, e a partir do ano selecionado para análise, uma estimativa para todos os dias de cada mês. O primeiro arquivamento feito do site O Estado de São Paulo

data de 1998, as colunas pretas representam os dados preservados, enquanto que os espaços em branco representam períodos em que o arquivamento não foi realizado. Os motivos para o não arquivamento podem ser diversos, mas considerando os autores Niu (2012) e Schroeder & Brügger (2017) citados no capítulo 01, nos períodos em que constam muitos lapsos com falta de dados sugere que o site do jornal não se encontrava na superfície da *web*, possivelmente em função do baixo nível de acessos e visualizações. É interessante assinalar que entre os anos de 2003 e 2006 o arquivamento se manteve constante, contemplando pelo menos todos os meses do ano com dados disponíveis, porém os anos de 2007 e 2008 retornam as falhas. Os anos de 2012 a 2014 se mantêm como os anos mais arquivados pelo *IA* disponibilizando dados para quase todos os dias do ano.

Os círculos azuis ao redor de cada dia representam a intensidade do arquivamento realizado, quanto maior o círculo maior o número de dados. Já os círculos verdes representam problemas no arquivamento: *hiperlinks* quebrados ou redirecionamentos para *url* diferentes do período do arquivamento. O tamanho dos círculos representa também a qualidade do arquivamento, ou seja, se parte da hierarquia dos *hiperlinks* relacionados a capa do jornal também foram preservados. As falhas (círculos verdes) decorrem muitas vezes desse arquivamento hierárquico: são os *hiperlinks* das reportagens apresentados na capa que apresentam erros, mantendo a capa idêntica ao período de publicação. Os redirecionamentos para *urls* diferentes do período do arquivamento indicam mudanças na estrutura do site realizadas pelo jornal, uma mudança no conteúdo de uma reportagem, uma troca de *url* por outra, uma troca de suporte multimídia para representar determinada reportagem, etc., mas que foi captado pelo *IA* em função do seu arquivamento em massa.

Para refinar a questão do número de capturas, e porque elas representam um número maior do que os dados disponíveis, passamos a segunda etapa de busca de dados da nossa pesquisa, em que analisamos pontualmente o último link disponível para cada dia (ou seja, a última capa do site por dia) para captar quais eram as principais discrepâncias entre elas a partir do conteúdo do jornal. Escolhemos a última capa de cada dia porque entendemos que este era o formato ‘mais acabado’ de notícias e informações que o site gostaria que seu leitor obtivesse. Entretanto, quando as manifestações começam a ganhar atenção especial do jornal, fizemos uma comparação entre o último e o primeiro *link* disponível, para percebermos a evolução da edição do site.



*Figura 2: Elaboração própria. Na imagem acima, podemos visualizar melhor os erros de arquivamento do IA. Para o mês analisado (junho de 2013), apenas um dia apresentou problemas.*

Produzimos a tabela a seguir para organizar os dados colhidos sobre as capturas.

DATA	NÚMERO TOTAL DE CAPTURAS	LINK DA ÚLTIMA CAPTURA	OBSERVAÇÕES
01 de junho de 2013	08	<a href="https://web.archive.org/web/20130601231859/https://www.estadao.com.br/">https://web.archive.org/web/20130601231859/https://www.estadao.com.br/</a>	Afirma: 08 snapshots, mas aparecem 04 links possíveis. Cada link representa o horário da captura.
02 de junho de 2013	13	<a href="https://web.archive.org/web/20130602232824/https://www.estadao.com.br/">https://web.archive.org/web/20130602232824/https://www.estadao.com.br/</a>	Aparecem 09 links possíveis. É possível verificar diferença no conteúdo da página principal entre o primeiro link e o último porque houve um jogo da seleção brasileira no final do dia. *após essa data não foram comparados os links para verificar o conteúdo, salvo exceções explícitas.
03 de junho de 2013	07	<a href="https://web.archive.org/web/20130603232054/https://www.estadao.com.br/">https://web.archive.org/web/20130603232054/https://www.estadao.com.br/</a>	Aparecem 3 links possíveis.
04 de junho	08	<a href="https://web.archive.org/web/20130604232859/https://www.estadao.com.br/">https://web.archive.org/web/20130604232859/https://www.estadao.com.br/</a>	Aparecem 07 links possíveis.

de 2013			
05 de junho de 2013	08	<a href="https://web.archive.org/web/20130605222557/https://www.estadao.com.br/">https://web.archive.org/web/20130605222557/https://www.estadao.com.br/</a>	Aparecem 06 links possíveis.
06 de junho de 2013	09	<a href="https://web.archive.org/web/20130606231631/https://www.estadao.com.br/">https://web.archive.org/web/20130606231631/https://www.estadao.com.br/</a>	Aparecem 07 links possíveis.
07 de junho de 2013	09	<a href="https://web.archive.org/web/20130607233554/https://www.estadao.com.br/">https://web.archive.org/web/20130607233554/https://www.estadao.com.br/</a>	Aparecem 07 links possíveis. Nesse último link, uma reportagem sobre a manifestação aparece na capa, setor “última hora”. Ao clicar no link da reportagem, wayback reporta um erro no link e redireciona a página. <sup>26</sup>
08 de junho de 2013	06	<a href="https://web.archive.org/web/20130608120947/https://www.estadao.com.br/">https://web.archive.org/web/20130608120947/https://www.estadao.com.br/</a>	Aparecem 05 links possíveis.

<sup>26</sup> Houve uma nova captura para o link direcionado, apresentando duas capturas desse redirecionamento. Acessar: [https://web.archive.org/web/20130611002856/http://blogs.estadao.com.br:80/estadao-urgente/acompanhe-a-manifestacao-contra-aumento-da-passagem-de-onibus-em-sp?doing\\_wp\\_cron=1370910518.9690771102905273437500](https://web.archive.org/web/20130611002856/http://blogs.estadao.com.br:80/estadao-urgente/acompanhe-a-manifestacao-contra-aumento-da-passagem-de-onibus-em-sp?doing_wp_cron=1370910518.9690771102905273437500). A data da última captura é do dia 11 de junho de 2013, o que pode indicar que alterações na página foram feitas até esse dia. **Alteração importante! (visto dia 17/08/2018).**

<b>09 de junho de 2013</b>	09	<a href="https://web.archive.org/web/20130609131808/https://www.estadao.com.br/">https://web.archive.org/web/20130609131808/https://www.estadao.com.br/</a>	Aparecem 05 links possíveis.
<b>10 de junho de 2013</b>	07	<a href="https://web.archive.org/web/20130610140926/https://www.estadao.com.br/">https://web.archive.org/web/20130610140926/https://www.estadao.com.br/</a>	Aparecem 06 links possíveis.
<b>11 de junho de 2013</b>	08	<a href="https://web.archive.org/web/20130611215834/https://www.estadao.com.br/">https://web.archive.org/web/20130611215834/https://www.estadao.com.br/</a>	Aparecem 07 links possíveis. Era o terceiro dia de manifestações em SP.
<b>12 de junho de 2013</b>	11	<a href="https://web.archive.org/web/20130612232719/https://www.estadao.com.br/">https://web.archive.org/web/20130612232719/https://www.estadao.com.br/</a>	Aparecem 09 links possíveis.
<b>13 de junho de 2013</b>	08	<a href="https://web.archive.org/web/20130613205103/https://www.estadao.com.br/">https://web.archive.org/web/20130613205103/https://www.estadao.com.br/</a>	Aparecem 06 links possíveis.
<b>14 de junho de 2013</b>	10	<a href="https://web.archive.org/web/20130614211211/https://www.estadao.com.br/">https://web.archive.org/web/20130614211211/https://www.estadao.com.br/</a>	Aparecem 08 links possíveis.



15 de junho de 2013	13	<a href="https://web.archive.org/web/20130615221033/https://www.estadao.com.br/">https://web.archive.org/web/20130615221033/https://www.estadao.com.br/</a>	Aparecem 07 links possíveis. Abertura da Copa das Confederações.
16 de junho de 2013	14	<a href="https://web.archive.org/web/20130616212709/https://www.estadao.com.br/">https://web.archive.org/web/20130616212709/https://www.estadao.com.br/</a>	Aparecem 10 links possíveis.
17 de junho de 2013	13	<a href="https://web.archive.org/web/20130617224917/https://www.estadao.com.br/">https://web.archive.org/web/20130617224917/https://www.estadao.com.br/</a>	Aparecem 11 links possíveis. <sup>27</sup>
18 de junho de 2013	09	<a href="https://web.archive.org/web/20130618222237/http://www.estadao.com.br/">https://web.archive.org/web/20130618222237/http://www.estadao.com.br/</a>	Aparecem 06 links possíveis. Aqui aparece uma chamada de capa tão importante quanto a do dia 17 de junho, mas não verificamos ainda <sup>28</sup> .
19 de junho	13	<a href="https://web.archive.org/web/20130619233614/https://www.estadao.com.br/">https://web.archive.org/web/20130619233614/https://www.estadao.com.br/</a>	Aparecem 08 links possíveis. Chamada de capa sobre

<sup>27</sup> Aqui repete-se o que foi verificado no dia 07 de junho, porém sem o aviso de erro no link. É sobre a chamada da página principal “Ao vivo: manifestantes chegam na Paulista”. O link para essa matéria, que veicula como o anterior numa página do “blog do Estadão”, conta com 43 capturas. Verificar o link: <https://web.archive.org/web/20130617214101/http://blogs.estadao.com.br/estadao-urgente/manifestantes-fazem-quinto-ato-contra-o-aumento-da-tarifa-de-onibus-em-sao-paulo/> **Informação importante! (visto dia 17/08/2018)**. A barra de informação do wayback sinaliza que a última captura da reportagem foi feita dia 03 de março de 2017. Coletamos duas imagens sobre esse novo link. Nas coletas para esse link vários círculos verdes aparecem, o que representam um erro na captura.

<sup>28</sup> Acesse link para a chamada de capa <https://web.archive.org/web/20130615191821/http://blogs.estadao.com.br/estadao-urgente/> **Informação importante! (visto dia 17/08/2018)**. Para esse link aparecem 53 capturas, a maior percebida até agora para esses dias de junho de 2013. A última captura data de 20 de março de 2017.

de 2013			manifestações aparecem com as palavras “o país nas ruas” <sup>29</sup> .
20 de junho de 2013	18	<a href="https://web.archive.org/web/20130620230032/https://www.estadao.com.br/">https://web.archive.org/web/20130620230032/https://www.estadao.com.br/</a>	Aparecem 11 links possíveis. Manifestações em mais uma capa <sup>30</sup> .
21 de junho de 2013	11	<a href="https://web.archive.org/web/20130621223000/https://www.estadao.com.br/">https://web.archive.org/web/20130621223000/https://www.estadao.com.br/</a>	Aparecem 08 links possíveis. Manifestações em reportagem de capa <sup>31</sup> .
22 de junho de 2013	11	<a href="https://web.archive.org/web/20130622234449/https://www.estadao.com.br/">https://web.archive.org/web/20130622234449/https://www.estadao.com.br/</a>	Aparecem 09 links possíveis <sup>32</sup> .

<sup>29</sup> Chamada de capa sobre manifestação aparece com o erro do wayback. A reportagem de capa agora redireciona para esse endereço: [https://web.archive.org/web/20130624084813/http://www.estadao.com.br:80/noticias/nacional/haddad-e-alkmin-anunciam-reducao-de-tarifas-do-transporte-publico-em-sp\\_1044416.0.htm](https://web.archive.org/web/20130624084813/http://www.estadao.com.br:80/noticias/nacional/haddad-e-alkmin-anunciam-reducao-de-tarifas-do-transporte-publico-em-sp_1044416.0.htm), e conta com 38 capturas no total, sendo a última captura realizada dia 28 de junho de 2017. A reportagem conta com a data de 19 de junho de 2013, mas a captura ‘principal’ é do dia 24 de junho de 2013. Capturamos uma imagem que mostra a discrepância nas datas. **Informação importante! (visto 17/08/2018).**

<sup>30</sup> Sobre a reportagem de capa das manifestações: <https://web.archive.org/web/20130620231637/http://blogs.estadao.com.br/estadao-urgente/protostos-ocorrem-em-75-cidades-e-terao-seguranca-reforcada/>, conta com 29 capturas, a última realizada em 20 de abril de 2017. **Informação importante! (visto 17/08/2018).**

<sup>31</sup> Sobre a reportagem de capa das manifestações: <https://web.archive.org/web/20130615191821/http://blogs.estadao.com.br/estadao-urgente/>, é o mesmo link para o dia 15 de junho de 2013, e a primeira captura data de 26 de janeiro de 2013. É o link com as informações mais desconstruídas até agora, em termos de datas. Conta com 53 capturas, a última no dia 20 de março de 2017. **Informação importante! (visto 17/08/2018).**

<sup>32</sup> Sobre a reportagem de capa das manifestações: <https://web.archive.org/web/20130626153600/http://blogs.estadao.com.br:80/estadao-urgente/ao-vivo-manifestantes-voltam-as-ruas-pelo-pais-neste-sabado>, conta com 22 capturas no total, a primeira realizada dia 26 de junho de 2013. A última captura feita dia 18 de agosto de 2016. **Informação importante! (visto 17/08/2018).**

<b>23 de junho de 2013</b>	12	<a href="https://web.archive.org/web/20130623221906/https://www.estadao.com.br/">https://web.archive.org/web/20130623221906/https://www.estadao.com.br/</a>	Aparecem 09 links possíveis. Primeira chamada de capa sobre a PEC 37 <sup>33</sup> .
<b>24 de junho de 2013</b>	10	<a href="https://web.archive.org/web/20130624230400/https://www.estadao.com.br/">https://web.archive.org/web/20130624230400/https://www.estadao.com.br/</a>	Aparecem 06 links possíveis. Capa é sobre reunião que Dilma convocou sobre o Plebiscito da Reforma Política.
<b>25 de junho de 2013</b>	12	<a href="https://web.archive.org/web/20130625182118/http://www.estadao.com.br/">https://web.archive.org/web/20130625182118/http://www.estadao.com.br/</a>	Único com link marcado como erro no wayback, mas é somente o primeiro link disponível. Aparecem 09 links possíveis. A chamada de capa é “Dilma revê proposta de convocar constituinte”.
<b>26 de junho de 2013</b>	10	<a href="https://web.archive.org/web/20130626234955/https://www.estadao.com.br/">https://web.archive.org/web/20130626234955/https://www.estadao.com.br/</a>	Aparecem 06 links possíveis. Chamada de capa é “Senado aprova projeto que define corrupção como crime hediondo”.
<b>27 de junho</b>	04	<a href="https://web.archive.org/web/20130627194511/https://www.estadao.com.br/">https://web.archive.org/web/20130627194511/https://www.estadao.com.br/</a>	Aparecem 04 links possíveis. Capa é sobre o plebiscito.

<sup>33</sup> Sobre a reportagem de capa das manifestações: <https://web.archive.org/web/20130626163742/http://blogs.estadao.com.br:80/estadao-urgente/ao-vivo-manifestantes-voltam-as-ruas-pelo-pais-neste-domingo/>, captura semelhante a nota anterior, mas com total de capturas reduzido: 17 capturas no total, sendo que a primeira data de 26 de junho de 2013, e a última data de 05 de dezembro de 2013.

<b>de 2013</b>			
<b>28 de junho de 2013</b>	02	<a href="https://web.archive.org/web/20130628230214/https://www.estadao.com.br/">https://web.archive.org/web/20130628230214/https://www.estadao.com.br/</a>	Aparecem 02 links possíveis. Capa sobre Copa das Confederações.
<b>29 de junho de 2013</b>	02	<a href="https://web.archive.org/web/20130629133143/https://www.estadao.com.br/">https://web.archive.org/web/20130629133143/https://www.estadao.com.br/</a>	Aparecem 02 links possíveis. Capa sobre a Copa das Confederações.
<b>30 de junho de 2013</b>	01	<a href="https://web.archive.org/web/20130630041834/https://www.estadao.com.br/">https://web.archive.org/web/20130630041834/https://www.estadao.com.br/</a>	Apenas um link disponível. Capa sobre a Copa das Confederações.

## 5.2. O arquivamento da *web* e a construção da narrativa do jornal O Estado de São Paulo.

Analisando pontualmente cada data, no dia 06 de junho de 2013 apareceram as primeiras menções importantes sobre as manifestações na capa do jornal, abrindo o período de cobertura especial que o jornal fez do momento. Na disposição do jornal, é a partir dessa data que as manifestações passam a ser acompanhadas pelo setor “Última Hora”, e existem dissonâncias consideráveis nesse primeiro momento. O dia inteiro conta com 09 capturas e 07 *links* disponíveis para análise, o primeiro deles com horário de 01:28:43 (AM)<sup>34</sup>. Pelo horário, podemos aferir que as notícias desse primeiro *link* são a última capa criada no dia 05 de junho, e as reportagens referem-se a indicação de Luís Roberto Barroso como ministro do STF<sup>35</sup>, sobre a regulamentação da Lei dos Portos, sobre vazamento de óleo na bacia amazônica, e sobre a renda do Campeonato Brasileiro de Futebol<sup>36</sup>, e apenas a primeira e última reportagens mencionadas foram arquivadas pelo IA, o que demonstra que a qualidade do arquivamento para o dia 06 nesse primeiro link é baixa.

A última capa do dia 06 de junho conta com o horário de 23:16:31, e nele apresentam a reportagem da manifestação na capa<sup>37</sup>, no setor “Última Hora”<sup>38</sup>. Acessando a reportagem, o *Wayback Machine* descreve o conteúdo como parte de uma coleção do IA chamada “Estadão”, e a página foi capturada 14 vezes entre os dias 07 de junho de 2013 e 16 de novembro de 2013. A partir desse dado, a pesquisa encontra novo desdobramento, o da captura maciça de alguns dos *links* disponíveis na hierarquia do site do jornal. Logo, passamos a verificar os *links* das reportagens relacionadas especificamente às manifestações e verificamos que a maioria destas tem mais capturas do que o resto do jornal, inclusive as capas.

<sup>34</sup> O *link* para essa primeira capa é <https://web.archive.org/web/20130606012843/http://www.estadao.com.br/>. Acessado em 12 de novembro de 2018.

<sup>35</sup> <https://web.archive.org/web/20130606003226/http://www.estadao.com.br/noticias/nacional,senado-aprova-barroso-para-vaga-no-stf,1039206,0.htm>. Acessado em 12 de novembro de 2018.

<sup>36</sup> <https://web.archive.org/web/20130609140120/http://blogs.estadao.com.br:80/jamil-chade/2013/06/05/brasileirao-ja-e-6o-campeonato-mais-rico-do-mundo>. Acessado em 12 de novembro de 2018.

<sup>37</sup> Para acessar a capa: <https://web.archive.org/web/20130606231631/http://www.estadao.com.br/>. Acessado em 12 de novembro de 2018.

<sup>38</sup> Para acessar a reportagem: <https://web.archive.org/web/20130607140721/http://www.estadao.com.br/noticias/cidades,manifestacao-contra-aumento-da-tarifa-de-onibus-fecha-vias-em-sao-paulo,1039657,0.htm>. Acessado em 12 de novembro de 2018.

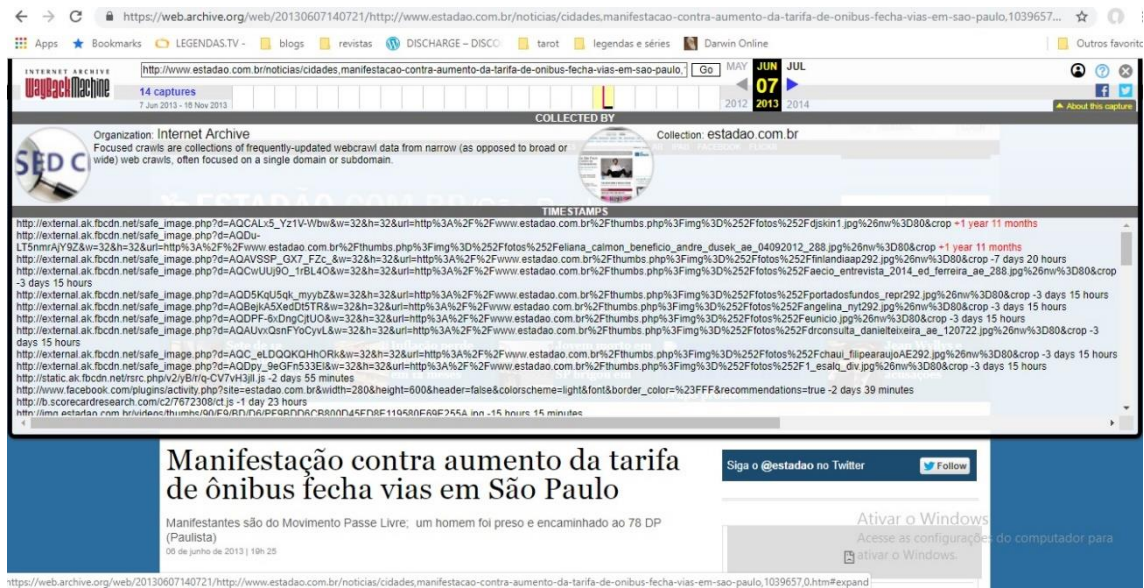


Figura 3: Elaboração própria. Detalhamento do arquivamento. Reportagem sobre as manifestações em São Paulo publicada no dia 06 de junho de 2013.

Na figura acima podemos verificar que a reportagem “Manifestação contra aumento da tarifa de ônibus fecha vias em São Paulo” foi publicada no dia 06 e arquivada pela primeira vez no IA no dia 07, o que indica que o arquivamento foi feito com uma das versões finais do documento. A reportagem acima é localizada como um site, parte do setor de ‘notícias’ e ‘cidades’ do jornal. Entretanto, se vamos buscar a reportagem referente às manifestações a partir do último *link* disponível para o dia 07 de junho<sup>39</sup>, o segundo dia da cobertura do jornal no setor “Última Hora”, observamos que a cobertura é feita por um *blog* do jornal, com texto escrito minuto a minuto.

É significativa a mudança de narrativa, considerando que o *blog* detém uma forma específica de narrar, mais livre e rápida, com a possibilidade de modificar rapidamente o conteúdo uma vez que a construção da linguagem *html* é mais simplificada. Weber afirma que *blogs* permitiram que jornalistas publicassem conteúdo mais bruto *online* fora da hierarquia de um jornal tradicional, se destacando como uma das formas de produção de conteúdo que reinventou a forma como as notícias diárias são produzidas (WEBER, 2017 In BRÜGGER & SCHROEDER, 2017, p. 88). Existe uma teoria jornalística que desde o final dos anos 1990 tem

<sup>39</sup> Para acessar a última capa do dia 07 de junho de 2013: <https://web.archive.org/web/20130607233554/http://estadao.com.br/>. Para acessar a reportagem referente às manifestações do último link para o dia 07 de junho: [https://web.archive.org/web/20130611002856/http://blogs.estadao.com.br:80/estadao-urgente/acompanhe-a-manifestacao-contr-aumento-da-passagem-de-onibus-em-sp?doing\\_wp\\_cron=1370910518.9690771102905273437500](https://web.archive.org/web/20130611002856/http://blogs.estadao.com.br:80/estadao-urgente/acompanhe-a-manifestacao-contr-aumento-da-passagem-de-onibus-em-sp?doing_wp_cron=1370910518.9690771102905273437500). Ambas acessadas em 12 de novembro de 2018.

discutido a existência do “*citizen journalism*”, ou ‘jornalismo cidadão’<sup>40</sup>. Segundo a perspectiva, a tecnologia como novo *médium* e a proliferação de seu uso abriu novos canais de mídia que, combinado com a precarização do trabalho do jornalista, iniciou uma demanda corporativa por um perfil específico do jornalista, em que o profissional precisa atuar de maneira independente e ‘livre’. Numa corporação, isso significa que o jornalista não tem tempo nem investimento para realizar uma reportagem investigativa de longo prazo. Esse é um dos motivos pelos quais as mídias corporativas optam por notícias ligadas ao comércio, entretenimento, e notícias governamentais. Atualmente, um grande jornal não se mantém sozinho, mesmo lucrando com publicidade, se mantendo apenas estando atrelado a uma corporação ou empresa maior, com outros objetivos que vão além da manutenção de um jornal. A partir disso, os autores Bruns, Highfield e Lind (BRUNS, HIGHFIELD & LIND, 2012 In LIND, 2012) afirmam que isso significa que textos editoriais (em alguns casos) e conteúdo jornalístico (na maioria dos casos) são direcionados para causas selecionadas, relacionados com os interesses políticos e econômicos da corporação que mantém o jornal, o que garante em certa medida a livre existência do jornal, mas não da produção jornalística individual (Op. cit.).

Segundo o histórico disponível no site O Estado de São Paulo<sup>41</sup>, o jornal foi fundado por um grupo de republicanos no ano de 1875, com o nome “A Província de São Paulo”, que foi alterado em 1890, quando São Paulo teve seu *status* modificado de província para estado. A fundação teve um claro objetivo político, colocado pelo site: um grupo de republicanos liderados por Manoel Ferraz de Campos Salles e Américo Brasiliense “decidiram criar um diário de notícias para combater a monarquia e a escravidão” (ESTADÃO, 2018, documento eletrônico). Em sua dissertação de mestrado, Moreira resgata as ponderações de Ribeiro que afirma que o jornal apresenta ambiguidade em suas falas, em que a direção superior nega-se a uma adaptação à opinião do leitor, e a linha editorial é de corte político representando os interesses da família Mesquita, dona do jornal (RIBEIRO, 1994, p. 91; apud MOREIRA, 2006, p. 83). Logo, para o país, O Estado de São Paulo é uma das principais e mais tradicionais corporações jornalísticas do país, que defende a liberdade de imprensa, porém escreve a partir de claros interesses sociais, políticos e econômicos.

É fundamental ter presente que toda corporação jornalística escreve a partir de alguma subjetividade, embora a imprensa se coloque como imparcial. Logo, cabe questionar o uso da

---

<sup>40</sup> Para nós, um equivalente ao ‘jornalismo cidadão’ seria a mídia alternativa.

<sup>41</sup> Ver: [https://acervo.estadao.com.br/historia-do-grupo/decada\\_1870.shtm](https://acervo.estadao.com.br/historia-do-grupo/decada_1870.shtm). Acessado em 12 de novembro de 2018.

ferramenta *blog* para noticiar um movimento político importante. Nesta data, as manifestações ainda não haviam tomado a proporção que conhecemos hoje, então, é possível aferir que o jornal optou pelo uso dessa ferramenta em função da praticidade, em detrimento da produção de conteúdo mais trabalhado. Entretanto, o uso de *blog* para veicular informações importantes acerca da política naquele momento se tornou comum no jornal. Na primeira capa do jornal do dia 11 de junho de 2013<sup>42</sup>, consta uma reportagem com o título “Em maio, número de votos a favor de Dilma na Câmara despenca”<sup>43</sup>, que redireciona para um outro *blog*, conhecido como “estadão dados”, uma proposta interativa de multimídia para demonstrar os assentos na Câmara Federal, quantos eram a favor do governo e quanto se colocavam contra. Porém, o site se encontra vazio, como podemos aferir na figura 4.



Figura 4: Blog Estadão Dados do dia 11 de junho de 2013.

Essa página conta com 7 capturas no total, a primeira da data 27 de agosto de 2013 e a última do dia 17 de agosto de 2017. As datas que constam na figura acima são de 26 de junho de 2013 (“atualizado em”) e 07 de junho de 2013 (“atualizado em”). Os dados seriam gráficos sobre as votações nominais na Câmara de Deputados Federais e no Senado ocorridas no mês de maio de 2013. O “pulso do governo Dilma” é medido pelo número de deputados federais votaram a favor dos projetos propostos pelo governo, e aqueles que votaram contra.

<sup>42</sup> A primeira capa do dia 11 de junho de 2013 consta o horário de 02:35:58 (AM) e pode ser acessada por esse link: <https://web.archive.org/web/20130611023559/http://www.estadao.com.br/>. Acessado em 12 de novembro de 2018.

<sup>43</sup> Ver através do link: <https://web.archive.org/web/20130827031011/http://blog.estadaodados.com/o-pulso-do-governo-dilma-na-camara-e-no-senado/>. Acessado em 12 de novembro de 2018.



Verificamos as 07 capturas realizadas desse link: apenas uma feita no ano de 2013, em 27 de agosto; uma realizada no ano de 2014, no dia 10 de julho; nenhuma realizada no ano de 2015; 03 realizadas no ano de 2016, nos dias 07 de abril, 12 de maio e 03 de agosto; e duas realizadas no ano de 2017, nas datas de 28 de julho e 17 de agosto.

Outro ponto importante é o de uma empresa bem-sucedida e histórica de jornalismo no país optar pelo uso dessas ferramentas simplificadas para veicular suas reportagens. Afora o problema da perda de informação através do arquivamento, em função do uso de ferramentas multimídia que são difíceis para o *crawler* arquivar (gerando quebra de *link*), existe a opção do jornal em utilizar ferramentas que são rotineiramente usadas por empresas com poucos recursos ou até mesmo por jornalistas independentes. Encontramos um outro exemplo dessa situação, em que o jornal usa uma outra ferramenta em formato *blog* para veicular fotos das manifestações. No mesmo dia 11 de junho, no centro da capa do jornal existe uma sessão de fotos das manifestações, com foco na cidade do Rio de Janeiro. Acessando a página das fotos através da capa, ela redireciona para uma ferramenta *blog* largamente utilizada nos EUA, conhecida por ser um espaço que majoritariamente compartilha imagens e vídeos, o *Tumblr*<sup>44</sup>. Capturada pelo menos 15 vezes entre os dias de 10 de agosto de 2012 e 30 de janeiro de 2017<sup>45</sup>, o *link* redirecionado não tem qualquer relação com as manifestações. Existem apenas 03 datas de publicação na página, dos dias 23 a 25 de junho de 2013 em sequência, e os temas das publicações são: número de mortos no Complexo da Maré, Dilma Rouseff e ministros se encontrando com representantes do Movimento Passe Livre de São Paulo, e a observação da Lua Cheia. Todas as publicações contariam com imagens, mas nenhuma foi arquivada pelo IA. Como o *Tumblr* é compreendido como rede social e o *Internet Archive* tem problemas com o arquivamento das redes, pode ser que em função do uso da ferramenta grande parte do conteúdo tenha se perdido.

---

<sup>44</sup> *Tumblr* é uma ferramenta de *microblog* que aparece na esteira de outras plataformas de *blogs* como o Blogspot. O diferencial dessa ferramenta é a proposta “micro” para a geração do seu conteúdo. Muito mais rápido do que ferramentas de Blog tradicional, o *Tumblr* propõe uma interface simplificada, porém moderna, em que seus usuários podem publicar fotos, imagens, vídeos, áudios, citações, textos e interagir com outros *posts* a partir de respostas direcionadas que podem ser lidas por toda a comunidade de usuários. O Blogspot ficou conhecido como um espaço de produção de conteúdo denso, com a capacidade de criação textual mais ampla e menos interação direta. O *Tumblr* despontou a partir de 2007 como a ferramenta *blog* mais usada entre jovens e mantém até hoje grande parte de seus usuários abaixo dos 25 anos de idade. A plataforma começou a alcançar seu apogeu no ano de 2010 e em 2013 foi comprada pela *Yahoo!*. Para maiores informações ver o texto de CHANG *et al.* (2014).

<sup>45</sup> Ver link: <https://web.archive.org/web/20130625201903/http://estadaofotos.tumblr.com/>, Acessado em 12 de novembro de 2018.

Retomando a colocação resgatada pelos autores Capurro e Hjørland (2007) citada no capítulo 4, que afirma “não há informação falsa, mas há significado sem verdade” (p. 170), fica a reflexão sobre a qualidade das informações arquivadas. Nesse pequeno exemplo exposto, analisando até aqui apenas 03 dias de publicações virtuais do jornal, além da significativa perda de informação, há também informações que não estão completamente relacionadas ao tema analisado. As escolhas de ferramentas de publicação não comprometem apenas o arquivamento, dão a tônica da construção narrativa sobre o episódio das manifestações. Com textos em formatos que não contextualizam o acontecimento, as manifestações de junho foram encaradas pela mídia como uma grande incógnita. Mesmo com reivindicações municipais a resposta para a demanda parece vir apenas da esfera federal, uma vez que o nome da então presidente Dilma Rousseff é um dos mais citados nas capas e nas reportagens. Com uma narrativa tão simplificada, qualquer perda de informação parece ser grande. Com relação aos dois últimos objetos analisados, não é possível aferir do se tratavam os dados expostos nas páginas, é possível apenas deduzir algumas questões. Pensando na Arquivística, nos conceitos de organização de documentos, é quase como se num arquivo de jornais estivessem arquivados exemplares com títulos, todavia sem seus respectivos textos.

Com relação ao posicionamento do jornal frente à preservação da sua produção jornalística, O Estado de São Paulo disponibiliza todo seu conteúdo impresso digitalizado via *web*. No entanto, essa informação é acessível mediante uma assinatura junto ao jornal. A presente pesquisa não teve acesso a base de dados por não ter efetuado a assinatura, entretanto é sabido que o Acervo Estadão<sup>46</sup> disponibiliza todos os exemplares impressos, desde 1875 até os dias atuais. Fica incerto se o acervo trata a produção *online* da mesma forma.

### **5.3. Abundância do silêncio.**

Ao tecer a ética da memória em Walter Benjamin, Mattos (2016) recorre ao pesquisador Ricardo Timm de Souza para chamar a atenção de um hábito do pensamento representacional, que “não consegue lidar senão com presenças”, em que a sociedade estaria vivendo em delírio prisioneiro dessa presença (MATTOS, 2016, p. 92-93). Relacionando Walter Benjamin com Jacques Derrida, Mattos afirma que a proposta do seu trabalho, da ética da memória, escaparia

---

<sup>46</sup> Ver <https://acervo.estadao.com.br/>. Acessado em 12 de novembro de 2018.

das ideias representadas no presente (MATTOS, 2016, p. 93), se colocando nos estilhaços que estão para além do presente. A autora Pires (2014) coloca:

Ao redefinir o conceito de verdade e recuperar a linguagem como campo para a resignificação do sujeito e da história, a obra de Benjamin apresenta-nos caminhos que levam a um diálogo entre o conhecimento e a verdade; a sensibilidade e o entendimento: ‘Benjamin reivindica para as ciências humanas outra forma de expor a verdade, forma que se distingue profundamente do que chamamos conhecimento empírico do real e, portanto, questiona os limites rígidos da racionalidade técnica, preconizando um tipo de conhecimento que inclui as paixões e as utopias indispensáveis à vida, sem as quais não há humanidade possível’. (SOUZA, 2009, p. 187, Apud PIRES, 2014, p. 816)

Pires (2014) analisa a questão da linguagem em Walter Benjamin ressaltando que a palavra é um meio para o autor, uma forma de comunicar conteúdos ou transmitir informações, em que a atividade intelectual capaz de gerar ideias e conceitos se comunica *na* linguagem, ou seja, toda a atividade intelectual só se realiza *na* linguagem (PIRES, 2014, p. 820-821, grifos da autora). Vimos até aqui que as fontes analisadas recordam de momentos do presente, com fotos, análises políticas com informações diagramadas, narrações minuto a minuto sobre acontecimentos de um dia, num esforço para reter partes daquele presente para o futuro. Contudo, as ausências dos dados podem ser vistas como um escape da construção desse presente, como os estilhaços colocados por Mattos (2016).

Nas fontes analisadas, o dia 17 de junho de 2013 conta com 13 capturas no total, com 11 *links* disponíveis para acesso, e dispõe capturas maciças na hierarquia da página, muito maior do que realizado sobre a capa. No último *link* disponível para o dia<sup>47</sup>, a cobertura da sessão “Última Hora” acompanhava as manifestações daquele dia. A reportagem central, “Ao vivo: manifestantes chegam à Paulista”<sup>48</sup>, foi capturada 43 vezes pelo IA, do dia 17 de junho de 2013 até o dia 03 de março de 2017. A página mantém as características analisadas anteriormente: publicada através de um *blog* com texto escrito minuto a minuto. Era a quinta manifestação para a redução da tarifa do transporte público e seus efeitos já eram amplamente sentidos pela cidade: mais de 30 mil manifestantes nas ruas, congestionamento histórico na cidade de São

<sup>47</sup> Ver <https://web.archive.org/web/20130617224917/http://www.estadao.com.br/>. Acessado em 12 de novembro de 2018.

<sup>48</sup> Ver <https://web.archive.org/web/20130617214101/http://blogs.estadao.com.br/estadao-urgente/manifestantes-fazem-quinto-ato-contra-o-aumento-da-tarifa-de-onibus-em-sao-paulo/>, Acessado em 12 de novembro de 2018.

Paulo, o segundo dia de manifestantes presos por portarem vinagre, violência policial como resposta às ruas tomadas.

Investigando como o arquivamento foi feito, averiguamos que essa reportagem não faz parte das coleções anteriormente mencionadas. Parte das capturas foram arquivadas numa coleção intitulada “*Spring 2013 Protests in Turkey*” (Primavera 2013 Protestos na Turquia), que tem o objetivo de coletar “novos artigos” que realizaram a cobertura dos protestos na Turquia. É importante lembrar que o arquivamento é realizado através de robôs que respondem especificamente a algumas demandas criadas em algoritmos. De maio a agosto de 2013, a Turquia passou por uma convulsão social desencadeada por protestos pacíficos de cunho ambiental que rapidamente se transformaram em contestação nacional ao poder federal turco. A Primavera Árabe iniciou um novo processo de derrubada de poderes estabelecidos a partir do ano de 2010, modificando o cenário político de países do Oriente Médio e norte da África, em que as redes sociais e a *internet* tiveram um papel fundamental para organização do movimento e disseminação de informação.

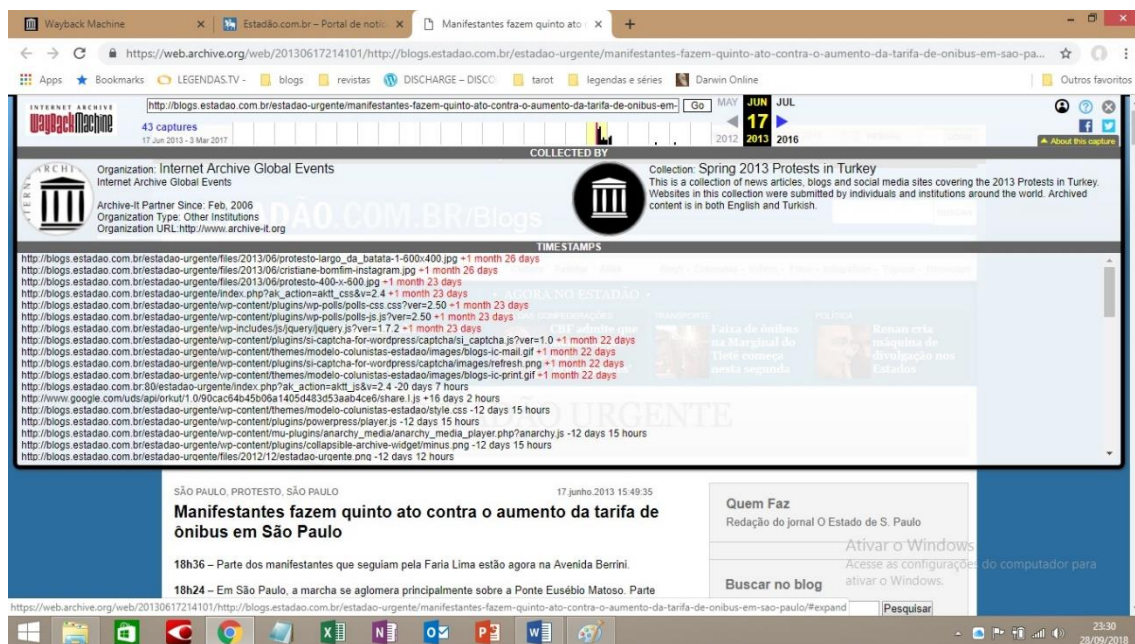


Figura 5: Detalhamento do arquivamento da matéria "Ao vivo: manifestantes chegam à Paulista".

Como sabemos, o processo político nacional brasileiro tomou rumos diferentes do que àqueles experimentados pela Turquia ou pelos países árabes, mesmo que por algum momento as Jornadas de Junho tenham sido lidas como “primavera”. Contudo, através desse exemplo podemos aferir que mesmo um arquivamento robotizado sem seleção humana prévia contém uma narratividade própria. A relação das manifestações brasileiras com os protestos na Turquia

parece estranha observada hoje. Entretanto, existe uma narração exterior criada sobre essas manifestações populares em países emergentes, que qualifica arranjos políticos de formas que não imaginamos. Parte dessa narrativa é criada justamente porque o *Internet Archive* é estadunidense, com uma agenda própria de interesses ao arquivar informações de todo o mundo.

Apesar da perda de informação em alguns campos do arquivamento da *web*, estamos lidando não somente com o excesso de informação, como também o excesso de narrativas possíveis sobre essa informação. Uma notícia sobre manifestações brasileiras arquivada numa coleção de notícias sobre a Turquia em 2013 é uma forma de qualificar a informação, que induz a uma determinada narrativa. Entre a abundância da informação, para narrar as possibilidades de produção desse arquivamento, teremos que atentar para os silêncios nas fontes, olhar criticamente para o que está presente no arquivo e questionar o que não está.

Benjamin (2012) afirma que a memória é uma faculdade épica por excelência e, não obstante, é o ato de rememorar a musa da poesia épica para os gregos. Segundo o autor, “a *rememoração* funda a cadeia da tradição, que transmite os acontecimentos de geração em geração. Ela corresponde à musa mais épica no sentido mais amplo. Entre elas, encontra-se em primeiro lugar a encarnada pelo narrador” (BENJAMIN, 2012, p. 228). Narração enquanto ato é fundamental para o rememorar, revisitando uma história ela passa a ter sentido a cada nova geração.

Um Arquivo dispõe de fontes. Pode ser entendida como memória bruta, que ganha sentido através de uma narração criada a partir delas. Uma análise amostral da plataforma *Internet Archive* demonstra o potencial que um arquivo tem. Todavia, os Arquivistas não são agentes passivos do processo. Nós mesmos temos que narrar nossos processos organizativos, que dão sentido a guarda, seleção, avaliação e organização dos documentos. Porque fazemos o que fazemos, e como o fazemos. Rememorar o passado com o auxílio dos arquivos também atravessa uma narrativa do nosso fazer profissional. Se a linguagem ressignifica o sujeito, a narração analítica sobre como um arquivo é organizado também é capaz de ressignificar nossa profissão.

## 6 Conclusão

Essa monografia de conclusão de curso teve como principal objetivo analisar a construção da memória das Jornadas de Junho de 2013 nos arquivos da *Internet Archive*, uma plataforma estadunidense de arquivamento da *web*, a partir do seguinte questionamento: “Que memória das Jornadas de Junho estamos resguardando?”. Organizamos nosso texto para melhor responder aos três objetivos propostos na metodologia, tendo a certeza de que o trabalho não está terminado. Este texto é, antes de tudo, um esforço narrativo sobre um arquivo: como ele é organizado, suas potencialidades, seus problemas, defeitos, falhas e como ele auxilia na construção da memória.

No terceiro capítulo visitamos brevemente a história da *web* mundial para compreender o funcionamento e premissas da plataforma *Internet Archive*. Com o propósito inabalado de se tornar a nova Biblioteca de Alexandria, vimos como o IA tem preservado desde os primórdios da história da *web*. Para manter seu objetivo principal, a instituição optou por arquivamentos robotizados em massa através dos *crawlers*, numa tentativa de captar o máximo de informação produzida na Rede. Atentamos para as características dos *crawlers*, suas formas de arquivamento, a repetição de informação arquivada, a perda de informação multimídia, a qualidade de arquivamento estabelecida pela captura da hierarquia de um site e não apenas de sua página inicial, vimos formas em que *sites* alteraram seu conteúdo por motivações políticas bem como parte do acervo pode ser usado em caráter probatório. Oferecemos uma breve revisão de literatura sobre o tema, cientes de que o assunto é recente e pouco trabalhado em textos de língua portuguesa, num esforço para oferecer uma pequena base bibliográfica para futuras pesquisas.

Para construir o debate acerca do conceito de memória no capítulo 04, visitamos autores da Arquivística e Ciência da Informação para compreender como o assunto é trabalhado na área. A partir do texto de Oliveira e Rodrigues (2011) entendemos que o conceito de memória não é muito trabalhado pela Arquivística, apesar de amplamente citado. Memória é relacionada qualitativamente a outros conceitos, não sendo trabalhado em profundidade. Para contornar o problema sem exceder o escopo do trabalho, julgamos interessante visitar outros conceitos importantes para as áreas, em especial o conceito de informação. Revisando o conceito de informação a partir de Capurro e Hjørland (2007), partimos para um diálogo com Dodebei (2006 e 2008) que relaciona os conceitos de patrimônio, memória e digital. Sobre o conceito de

memória, arriscamos trabalhar com um autor não utilizado na Arquivística para tratar do tema, ainda que seja um expoente nome na área das Ciências Humanas e Sociais. Trabalhamos com Walter Benjamin (2012) a partir de seu texto “O Narrador”, escrito no ano de 1936, e outros autores que dialogaram de forma potente com os textos do autor.

No capítulo 05 trouxemos nossa análise, construída novamente a partir da introdução ao objeto de pesquisa, dessa vez com novos detalhes das fontes primárias. Fizemos o uso de figuras para ilustrar o arquivo *online*, bem como apresentamos dados relacionados as capturas dos *links* em uma tabela, com intenção pedagógica de exemplificar determinados pontos da nossa análise. A partir da observação de duas datas específicas, demonstramos como o jornal construiu sua narrativa para o evento no período delimitado. Verificamos como o uso de determinadas ferramentas para publicação de conteúdo *online* interferiu na qualidade do arquivamento no IA. Ao final, procuramos expor como a rememoração pode atuar em dois extremos: na abundância e na ausência de informação.

Este texto foi construído a partir da crença na potencialidade da Arquivística enquanto Ciência. A falta de estudos sobre o tema e a falta de iniciativa nacional para tratar do arquivamento da *web* não nos levou a narrativa da crise na Arquivística nacional, tão comum nos trabalhos acadêmicos recentes. Entretanto, não negamos nem silenciemos frente à crise. Entendemos que somos forjados profissionalmente por ela, e devemos trabalhar sempre a partir dela.

Nosso caminhar profissional está intimamente ligado às escolhas intelectuais feitas nesse trabalho. Do entendimento da Arquivística como parte integrante da Ciência da Informação à escolha de Walter Benjamin como referencial teórico sobre memória, tudo serviu não apenas para analisar um objeto de pesquisa, mas serve também para nos colocarmos como pesquisadores profissionais dentro da área. É possível perceber através da análise que o excesso de informação pode levar a certa confusão. Porém é fundamental entender as diferenças entre informação e narrativa, e informação e memória. Essas escolhas também serviram para iniciar um esforço pedagógico de marcar a importância do profissional arquivista se narrar dentro de suas atividades. A narrativa de um objeto histórico social não é completa sem a narrativa de si, do pesquisador profissional por trás do texto.

O materialismo histórico tem presente o posicionamento do pesquisador. A pesquisa científica não começa pelo objeto, ela começa sempre pelo próprio pesquisador. A

potencialidade de questionar um objeto será um reflexo da potencialidade de questionar a si mesmo. É por isso que Benjamin (2012) afirma a necessidade de “escovar a história a contrapelo”. Ao caminhar na contramão da produção científica majoritária, está aberta a possibilidade de construir uma narrativa histórica autenticamente política.



## 7 Referências Bibliográficas.

- BENJAMIN, Walter. **Magia e técnica, arte e política: ensaios sobre literatura e história da cultura**. Tradução Sérgio Paulo Rouanet; prefácio de Jeanne Marie Gagnebin – 8ª edição. São Paulo: Editora Brasiliense, 2012.
- BRASIL. ARQUIVO NACIONAL. **Programa permanente de preservação digital**. 2016.
- BRÜGGER, Niels. Web archiving – Between Past, Present and Future. IN: CONSALVO, Mia; ESS, Charles (Ed.). **The handbook of internet studies**. John Wiley & Sons, 2011.
- BRÜGGER, Niels; SCHROEDER, Ralph (Ed.). **Web as History: Using Web Archives to Understand the Past and the Present**. UCL Press, 2017.
- CAPURRO, Rafael & HJORLAND, Birger. O conceito de informação. The concept of information as we use in everyday. IN: **Perspectivas em Ciência da Informação**. Vol. 12, n. 01, p. 148-207. Jan/abril de 2007.
- CHANG, Yi *et al.* What is tumblr: A statistical overview and comparison. IN: **ACM SIGKDD explorations newsletter**, v. 16, n. 1, p. 21-29, 2014.
- DODEBEI, Vera. Patrimônio e Memória Digital. **Revista Eletrônica em Ciências Humanas**, Rio de Janeiro, ano 04, n. 08, 2006.
- DODEBEI, Vera. Digital virtual: o patrimônio no século XXI. IN: **E o patrimônio?** Vera Dodebei & Regina Abreu (orgs.). Rio de Janeiro: Contra Capa & Programa de Pós Graduação em Memória Social da Universidade Federal do Estado do Rio de Janeiro, 2008.
- DURANTI, Luciana. Registros documentais contemporâneos como provas de ação. **Revista Estudos Históricos**, v. 7, n. 13, p. 49-64, 1994.
- FERREIRA, Lisiane Braga; MARTINS, Marina Rodrigues; ROCKEMBACH, Moisés. Usos do Arquivamento da Web na Comunicação Científica. **PRISMA. COM**, n. 36, p. 78-98, 2018.
- GONÇALVES, José Reginaldo Santos. O patrimônio como categoria de pensamento. IN: **Memória e patrimônio: ensaios contemporâneos**. Rio de Janeiro: DP&A, p. 21-29, 2003.
- BRUNS, Axel; HIGHFIELD, Tim; LIND, Rebecca Ann. Blogs, Twitter, and breaking news: The produsage of citizen journalism. IN: LIND, Rebecca Ann. **Producing theory in a digital world: The intersection of audiences and production in contemporary theory**, v. 80, n. 2012, p. 15-32, 2012.
- MATTOS, Manuella Sampaio de. **Ética da memória em Walter Benjamin. Um ensaio**. Porto Alegre, Editora Bestiário, 2016.
- MONTEIRO, Fernanda. Reflexões epistemológicas dos arquivos e do fazer arquivístico enquanto instrumentos de poder. IN: **Acervo**, v. 27, n. 1 jan-Jun, p. 313-322, abr. 2014.
- MOREIRA, Fabiane Barbosa. **Os valores-notícia no jornalismo impresso: análise das ‘características substantivas’ das notícias dos jornais Folha de São Paulo, O Estado de São Paulo e o Globo**. 157f. Dissertação (Mestrado em Comunicação e Informação). Faculdade de

Biblioteconomia e Comunicação. Porto Alegre: Universidade Federal do Rio Grande do Sul, 2006.

NIU, Jinfang. An overview of web archiving. IN: **D-Lib magazine**, v. 18, n. 3/4, 2012.

O'REILLY, Tim. What is web 2.0. IN: **Online Communication and Collaboration: A Reader**. Editado por Helen Margaret Donelan, Karen Kear, Magnus Ramage. New York, 2005.

DE OLIVEIRA, Eliane Braga; RODRIGUES, Georgete Medleg. O conceito de memória na Ciência da Informação: análise das teses e dissertações dos programas de pós-graduação no Brasil. IN: **Liinc em Revista**. Rio de Janeiro, vol. 07, n. 01, março de 2011, pp. 311-328.

PINSKY, Carla Bassanezi (org.). **Fontes históricas**. São Paulo: Editora Contexto, 2006.

PIRES, Eloiza Gurgel. Experiência e linguagem em Walter Benjamin. IN: **Educação e Pesquisa**. São Paulo, v. 40, n. 3, p. 813-828, set. 2014.

PRICE, Ludovica. **Internet Archiving-The Wayback Machine**. 2011.

PRIMO, Alex. O aspecto relacional das interações na Web 2.0. In: **E-Compós**. 2007.

REIGERT, Natali Moniqui. **Arquivamento e preservação da web**: procedimentos de coleta e armazenamento de sites institucionais dos arquivos públicos estaduais brasileiros. 2017. 98f. Trabalho de conclusão (Graduação em Arquivologia). Faculdade de Biblioteconomia e Comunicação, Universidade Federal do Rio Grande do Sul, 2017.

ROCKEMBACH, Moisés. Inequalities in digital memory: ethical and geographical aspects of web archiving. IN: **Information Ethics from a Marxian Perspective**, v. 26, p. 138, 2017.

ROCKEMBACH, Moisés. Arquivamento da Web: estudos de caso internacionais e o caso brasileiro. IN: **RDBCI: Revista Digital de Biblioteconomia e Ciência da Informação**. Campinas, SP. Campinas, SP. Vol. 16, n. 1 (jan./abr. 2018),[18] p., 2018.

ROCKEMBACH, Moises; PAVÃO, Caterina Marta Groposo. Políticas e tecnologias de preservação digital no arquivamento da web. IN: **Revista Ibero-Americana de Ciência da Informação**, v. 11, n. 1, p. 168-182, fev. 2018. ISSN 1983-5213. Disponível em: <http://periodicos.unb.br/index.php/RICI/article/view/27950> .

RONDINELLI, Rosely Curi. **Gerenciamento arquivístico de documentos eletrônicos: uma abordagem teórica da diplomática arquivística contemporânea**. Rio de Janeiro: FGV Editora, 2005.

RONDINELLI, Rosely Curi. **O documento arquivístico ante a realidade digital: uma revisão conceitual necessária**. Rio de Janeiro: Editora FGV, 2013.

SILVA, Armando Malheiro da; RIBEIRO, Fernanda; RAMOS, Júlio; REAL, Manuel Luís. **Arquivística – Teoria e Prática de uma Ciência da Informação**. Volume 1. Porto, Portugal: Edições Afrontamento, 1999.

WEBER, Matthew S. The tumultuous history of news on the web. IN: BRÜGGER, Niels; SCHROEDER, Ralph (Ed.). **Web as History: Using Web Archives to Understand the Past and the Present**. UCL Press, 2017.

WINTERS, Jane. Web archives for humanities research – some reflections. IN: BRÜGGER, Niels; SCHROEDER, Ralph (Ed.). **Web as History: Using Web Archives to Understand the Past and the Present**. UCL Press, 2017.