

# XX ENANCIB

21 a 25 Outubro/2019 – Florianópolis

A Ciência da Informação e a era da Ciência de Dados

ISSN 2177-3688

GT-8 – Informação e Tecnologia

**PRESERVAÇÃO DE MÍDIAS SOCIAIS E ARQUIVAMENTO DA WEB: UM ESTUDO ACERCA DAS ELEIÇÕES PRESIDENCIAIS BRASILEIRAS DE 2018**

***SOCIAL MEDIA PRESERVATION AND WEB ARCHIVING: A STUDY OF THE 2018 BRAZILIAN PRESIDENTIAL ELECTIONS***

Lisiane Braga Ferreira - Universidade Federal do Rio Grande do Sul

Moisés Rockembach - Universidade Federal do Rio Grande do Sul

**Modalidade: Trabalho Completo**

**Resumo:** O arquivamento da *web* e mídias sociais constitui o tema central deste trabalho, bem como a preservação dos conteúdos audiovisuais publicados no *Facebook* pelos candidatos à Presidência durante a campanha eleitoral de 2018. Adotou a metodologia de revisão bibliográfica e análise documental, de desenvolvimento de códigos de programação e estruturação de base de dados, sob uma abordagem exploratória-descritiva de análise quanti-qualitativa. O referencial teórico versou, predominantemente, sobre o arquivamento da *web* e sua relação com as mídias sociais. Observou-se que todos os candidatos possuíam *fanpage* no *Facebook* na campanha de 2018, sendo ativa a publicação de vídeos e que, dos 2.821 vídeos coletados durante a campanha eleitoral, 32 não estavam mais disponíveis no formato original ou por restrições legais ou por indisponibilidades diversas. Por fim, apresentou o protótipo utilizado para realizar a coleta e preservação dos registros e metadados. Concluiu que existe uma carência geral de arquivamento da *web* e das mídias sociais de conteúdos brasileiros em plataformas de preservação da *web* internacionais, sendo vital o fomento, no Brasil, da preservação de conteúdos públicos em mídias sociais a fim de minimizar a efemeridade inerente aos registros.

**Palavras-chave:** Arquivamento da *web* e mídias sociais. Candidatos presidenciais de 2018. Eleições presidenciais de 2018. *Facebook*. Vídeos.

**Abstract:** *The web archiving and social media is the central theme of this work, as well as the preservation of the audiovisual content published on Facebook by presidential candidates during the 2018 electoral campaign. It adopted the methodology of bibliographic review and documentary analysis, development of codes of programming and database structuring, under an exploratory-descriptive approach of quanti-qualitative analysis. The theoretical reference was predominantly about web archiving and its relationship to social media social media. It was observed that all candidates had a Facebook fanpage in the 2018 campaign, with videos being active and that of the 2,821 videos collected during the election campaign, 32 were no longer available in the original format or because of legal restrictions or various unavailability. Finally, it presented the prototype used to perform the collection and preservation of the records and metadata. It was concluded that there is a general lack of web archiving of Brazilian social media on international web preservation platforms, and it is vital to promote the preservation of public content in social media in order to minimize the inherent ephemerality of the records.*

**Keywords:** *web archiving and social media. Presidential Candidates 2018. Presidential Elections 2018. Facebook. Videos.*

## 1 INTRODUÇÃO

O presente trabalho decorre de uma pesquisa de mestrado (FERREIRA, 2019), sendo aqui expostos resultados obtidos durante a investigação desenvolvida. A pesquisa foi motivada a partir de estudos acerca da preservação das informações digitais produzidas pelos candidatos políticos a presidência da república do Brasil em 2018 durante suas campanhas e, mais especificamente, as mídias sociais, bem como a efemeridade que as publicações estão sujeitas nesses meios.

Embora haja muitos estudos sobre o arquivamento da web, processo que envolve a preservação de conteúdos disponibilizados online, sobre a preservação da *web* brasileira e sobre a preservação das mídias sociais, ainda há muito o que se desenvolver e discutir sobre a preservação da web brasileira. Alguns estudos brasileiros investigando as políticas e tecnologias de arquivamento da web (ROCKEMBACH, PAVÃO, 2018) ou os seus usos científicos (FERREIRA, MARTINS, ROCKEMBACH, 2018) já apontam para a relevância destas fontes de informação, carecendo a preservação dos conteúdos das mídias sociais de uma maior investigação.

A escolha pela plataforma *Facebook* se deu pelo fato de ser a mídia social com o maior número de usuários contando, atualmente com aproximadamente 79,4 milhões de usuários no Brasil (STATISTA, 2019).

Assim, nosso objeto de pesquisa envolveu 2.821 conteúdos audiovisuais publicados pelos treze candidatos à Presidência da República do Brasil em 2018, bem como do ex-

presidente Lula, em suas *fanpages* oficiais no *Facebook*, durante a campanha eleitoral do primeiro e segundo turnos compreendidos entre 16 de agosto de 2018 a 04 de outubro de 2018 e 08 de outubro a 26 de outubro de 2018.

Com o intuito de realizar a coleta e a preservação de um grupo de informações produzidas pelos candidatos durante a campanha eleitoral, optou-se em focar nos conteúdos audiovisuais por constituírem registros ricos em informações e ainda pouco explorados nas iniciativas de arquivamento da *web*, inclusive se comparado com os estudos internacionais.

O objetivo geral deste estudo visou compreender a necessidade de preservação digital dos vídeos publicados pelos candidatos à Presidência no *Facebook*, por meio do arquivamento da *web*.

A metodologia utilizada foi composta por pesquisa bibliográfica a partir do portal de periódicos CAPES sobre o arquivamento da *web* e análise documental baseada nos dados coletados. Para a coleta dos vídeos foram utilizados recursos informatizados a partir do desenvolvimento de interfaces *web* de inserção, consulta e um *bot*<sup>1</sup> que viabilizou a coleta dos metadados. Todos estes recursos, da concepção aos códigos de programação, foram desenvolvidos especialmente para esta pesquisa, servindo de contributo para futuros projetos. Também foi realizada a estruturação de três tabelas em uma base de dados para armazenar e permitir a análise das informações coletadas. As análises foram executadas por processos de extração de informações do banco de dados (*queries*), sendo que a pesquisa foi quantitativa-qualitativa, com abordagem exploratória-descritiva.

A apresentação dos dados coletados, neste trabalho, foi dividida em duas etapas: a primeira discorre sobre os principais dados coletados e a segunda descreve a análise de disponibilidade do formato de origem dos vídeos coletados. Por fim, apresentamos o protótipo de arquivamento disponibilizado para futuras pesquisas.

## 2 ARQUIVAMENTO DA *WEB* E MÍDIAS SOCIAIS DAS ELEIÇÕES

A *web* é um canal de comunicação dos mais relevantes atualmente e que desempenha um papel fundamental no fornecimento de informações aos eleitores – não apenas durante, mas também fora da temporada eleitoral, para o exercício da cidadania. Os

---

<sup>1</sup> Aplicação de *software* que desempenha uma tarefa de forma autônoma.

políticos, em geral, têm optado cada vez mais por fornecer informações apenas em formato digital, por ser mais rápido e muito mais barato. Outras razões que impulsionam essa comunicação são as possibilidades de recrutar trabalhadores de campanha e obter apoio financeiro. Além disso, já há muito tempo que a maioria dos políticos e partidos buscam através de novas formas de participação digital envolver mais as pessoas, especialmente os jovens. (FERREIRA, 2019; VOERMAN et al, 2002).

Com o aumento do uso da *web* como plataforma de comunicação eleitoral e a consequente necessidade de pesquisa nesses canais e fontes de informação, diversas iniciativas internacionais de arquivamento da *web* têm trabalhado na criação de coleções temáticas ou por evento sobre as eleições a fim de minimizar o impacto da efemeridade a que esses registros estão sujeitos.

O arquivamento da *web*, definido por Brügger (2011), é visto como um ato deliberado e intencional, que implica consciência de que se está preservando o material e porque o material da *web* está sendo preservado. O autor explica que o início da *web* desencadeou esta necessidade de preservar o que foi criado ou encontrado no ciberespaço por usuários comuns, definidos pelo autor como amadores, os quais deram origem a uma das muitas formas do microarquivamento. Esses arquivamentos possibilitaram a única evidência existente da *web* antiga. Por outro lado, com os avanços das tecnologias de arquivamento da *web* viabilizada por grandes instituições, foi possível desenvolver o macroarquivamento. (BRÜGGER, 2011; FERREIRA, 2019).

Costa, Gomes e Silva (2017) expõem que as iniciativas de arquivamento da *web* são bastante heterogêneas em termos de tamanho e escopo, sendo que a maior parte dos arquivos da *web* contêm exclusivamente conteúdo relacionado ao seu país de hospedagem, região ou instituição.

No entanto, podem existir outros casos como o exemplo da Biblioteca do Congresso dos Estados Unidos (*Library of Congress*) que, além de preservar desde 2000 diversas coleções sobre política em geral, também criou uma coleção temática intitulada *Brazilian Presidential Election 2010 Web Archive*, composta por 48 *websites* de partidos políticos nacionais e de candidatos das eleições presidenciais brasileiras de 2010. (FERREIRA, 2019; LIBRARY OF CONGRESS, 2018).

Para a criação de uma coleção temática ou por evento é necessário estabelecer as políticas de seleção que retornarão as informações desejadas da forma mais completa

possível. Assim, é realizada a identificação e captura de objetos da *web* através de um conjunto de URLs que se acredita serem relevantes para um tema ou tópico específico. Este conjunto de URLs selecionadas, que representam os *sites* ou páginas de interesse, são utilizados como “sementes” e irão rastrear, em uma periodicidade estabelecida e com regras claramente especificadas, as páginas e objetos vinculados. (FERREIRA, 2019; SCHNEIDER et al, 2003).

Contudo, é importante ressaltar que não foram encontradas coleções que preservassem conteúdos audiovisuais publicados em mídias sociais de campanhas eleitorais brasileiras. Compreendendo assim que, segundo Pop, Vasile e Masanès (2010), é difícil projetar uma solução geral para lidar com todos os sites que hospedam conteúdo de vídeo e a técnica de coleta deve ser adaptada para cada caso particular, bem como o esforço de engenharia de rastreamento necessário para adaptar as ferramentas geralmente depende da complexidade do site. E quando nos referimos a plataformas privadas, como o *Facebook*, este desafio se mostra mais complexo e exige um conjunto de habilidades, as quais apenas a engenharia de rastreamento não é capaz de sanar sozinha. (FERREIRA, 2019; POP, VASILE E MASANÈS, 2010).

### **3 METODOLOGIA**

A metodologia utilizada consistiu em uma abordagem exploratória-descritiva de um estudo quanti-qualitativo, onde a coleta de vídeos envolveu a utilização de recursos informatizados de linguagem de programação e estruturação em base de dados. Esta abordagem se justifica a partir de Gil (2002) que descreve a análise exploratória como usual no desenvolvimento de ideias e hipóteses que poderão ser testadas em estudos posteriores, possibilitando maior flexibilidade que o dos outros tipos de pesquisa. E a pesquisa descritiva como a descrição das características do fenômeno estudado, sendo que a preocupação do pesquisador é a de descrever com precisão essas características, utilizando instrumentos que conduzem a resultados de natureza quantitativa. Contudo, também foram realizadas inferências a partir das análises documentais das publicações audiovisuais realizadas pelos candidatos em suas *fanpages*. (FERREIRA, 2019; GIL, 2002).

A coleta dos dados ocorreu a partir da identificação das *fanpages* dos candidatos na plataforma *Facebook*, do desenvolvimento do código das páginas de interface *web* (inclusão e consulta) e da estruturação da base de dados MySQL<sup>2</sup>.

Para estruturar a base de dados e estabelecer os critérios de desenvolvimento do código de programação, coletamos os números de identificação (ID) respectivos de cada candidato em sua *fanpage*, usando-os para popular a primeira tabela e atuar como elo nas demais tabelas.

Foram criadas três tabelas no MySQL, a primeira, denominada “tbl\_alvo”, que listou todos os candidatos à Presidência; a segunda, denominada “tbl\_archive”, onde foram armazenados todos os registros coletados e seus respectivos dados; e a terceira tabela, denominada “tbl\_log”, armazenou os dados e metadados atualizados diariamente no 1º turno das eleições.

O processo de desenvolvimento das rotinas de *scripts*<sup>3</sup> passou por várias alterações, uma vez que a maneira ideal e correta de se buscar dados de dentro de uma página do *Facebook* seria através do uso de *Application Programming Interface*<sup>4</sup> (APIs) disponibilizadas para tal. Porém, depois do escândalo da *Cambridge Analytica*, o *Facebook* restringiu ao máximo o uso das APIs para o público “doméstico”, deixando apenas abertas possibilidades para grandes agências de notícias, ou empresas devidamente autorizadas, com ampla restrição ao tipo de dados fornecidos.

Optou-se, então, pelo uso da técnica de *scraping*, que consiste em criar *scripts* que analisam as *strings*<sup>5</sup> retornadas das páginas em análise.

O processo de *scraping* dividiu-se em cinco etapas: 1) identificação da página pública do candidato; 2) realização de um “*dump*”<sup>6</sup> dessa página, usando técnicas de pesquisa de programação que buscam por palavras-chave que fornecem pistas do que se deseja coletar. Neste caso, por exemplo, buscamos por metadados que informaram os nomes dos arquivos mp4 (vídeos), o assunto da postagem (*description*), etc; 3) criação de uma listagem em um banco de dados, que armazenou cada link a ser investigado ao longo do tempo. Esta listagem foi constituída de uma tabela que armazenou o link, o nome do candidato o ID do

---

<sup>2</sup> Versão *Software Livre*.

<sup>3</sup> Linguagem de programação.

<sup>4</sup> Conjunto de rotinas parametrizadas que possibilitam integrações entre diferentes aplicações.

<sup>5</sup> Sequência de caracteres usados para representar uma palavra, frase ou textos de um programa.

<sup>6</sup> Exportação de todos os dados.

candidato e o nome da pasta de *download* dos registros. Esta coleta foi feita diariamente de forma manual, com base em observações das postagens dos candidatos nas suas páginas, usando a interface *web*, que foi desenvolvida especificamente para isso; 4) realização, durante a coleta inicial, do armazenamento dos metadados 'data da coleta', 'tamanho' e 'duração do vídeo', identificação da forma de transmissão, ou seja, se era *live* ou vídeo comum, reações daquele momento, bem como outros metadados contidos no *iframe* do *Facebook*; por fim, 5) realização das coletas posteriores, a partir do mecanismo automático, programado para pesquisar novamente a base a cada seis horas, limitando-se a coletar novos dados sobre as reações referentes aos vídeos e *lives*. (FERREIRA, 2019).

O período de coleta foi definido a partir da data estipulada pelo TSE para a realização da propaganda eleitoral, sendo do dia 16 de agosto a 04 de outubro o 1º turno das eleições, e do dia 08 de outubro ao dia 26 de outubro para os dois candidatos do 2º turno.

Para identificar a persistência dos vídeos coletados na fonte de origem, isto é, nas *fanpages* dos candidatos, foi realizada a inclusão do campo “ativo na fonte” na página de consulta, onde a regra geral de população na base de dados foi definida como “S” (sim) e quando identificada, manualmente, a indisponibilidade do conteúdo, foi alterada para “N” (não). Esta identificação foi realizada logo após o fim das eleições, nos dias 15, 16 e 17 de novembro de 2018. (FERREIRA, 2019).

#### **4 COLETA, ANÁLISE E INTERPRETAÇÃO DOS DADOS**

A campanha eleitoral brasileira iniciou no dia 16 de agosto, mas a Justiça Eleitoral julgou até o dia 17 de setembro todos os pedidos de registro de candidatos que concorreram ao pleito de 2018. (TRIBUNAL SUPERIOR ELEITORAL, 2018). Em decorrência disso, até o dia 17 de setembro, o candidato à Presidência pelo Partido dos Trabalhadores (PT) era o ex-presidente Luiz Inácio Lula da Silva (Lula) e o candidato a vice-presidente era Fernando Haddad. Contudo, ao não ser aceita a candidatura de Lula pelo TSE, em função da Lei Complementar n.135/2010, houve a troca de candidatos do PT – passando a ser o candidato à Presidência Fernando Haddad e, a vice-Presidência, Manuela D’Ávila.

Desta forma, buscamos preservar o contexto das eleições presidenciais de 2018, mantendo a coleta das informações e dos dados publicados na *fanpage* do ex-presidente Lula durante todo o 1º turno das eleições.

**XX ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO – ENANCIB 2019**  
**21 a 25 de outubro de 2019 – Florianópolis – SC**

Assim sendo, a partir de Ferreira (2019), listamos na **Tabela 1** os nomes de todos os candidatos à Presidência, a identificação de suas páginas/*fanpages* no *Facebook*, o partido ao qual são filiados, os vice-presidentes da chapa e seus partidos de filiação.

**Tabela 1 - Lista dos candidatos à Presidência - informações gerais.**

CANDIDATOS À PRESIDÊNCIA				
CANDIDATO	ID FACEBOOK	PARTIDO	VICE	PARTIDO
Álvaro Dias	@ad.alvarodias	Podemos	Paulo Rabello de Castro	PSC
Cabo Daciolo	@depudadocabodaciolo	Patriota	Suelene Balduino Nascimento	Patriota
Ciro Gomes	@cirogomesoficial	PDT	Kátia Abreu	PDT
Fernando Haddad	@fernandohaddad	PT	Manuela D'Ávila	PCdoB
Geraldo Alckmin	@geraldoalckmin	PSDB	Ana Amélia	PP
Guilherme Boulos	@guilherme.boulos	PSOL	Sônia Guajajara	PSOL
Henrique Meirelles	@hmeirellesoficial	MDB	Germano Rigotto	MDB
Jair Bolsonaro	@jairmessias.bolsonaro	PSL	Hamilton Mourão	PRTB
João Amoêdo	@JoaoAmoedoNOVO	NOVO	Christian Lohbauer	NOVO
João Goulart Filho	@joaogoulart54	PPL	Léo Alves	PPL
José Maria Eymael	@eymaelOficial	DC	Helvio Costa	DC
Marina Silva	@marinasilva.oficial	Rede	Eduardo Jorge	PV
Vera Lúcia	@verapstu	PSTU	Hertz Dias	PSTU
Lula	@Lula	PT		

Fonte: Dados da pesquisa.

A coleta foi realizada a partir dos recursos desenvolvidos na interface *web* da **Figura 1**, onde era inserido manualmente o link do vídeo que se queria preservar e os *scripts* da página de inclusão coletavam as demais informações contidas no *iframe* do vídeo do candidato. (FERREIRA, 2019).

**Figura 1 - Interface web de inclusão dos registros audiovisuais na base de dados.**

Fonte: Dados da pesquisa.

Nossa análise se deu sobre os conteúdos audiovisuais publicados como vídeos e *lives* pelos candidatos à Presidência. No entanto, a coleta não se limitou a estes conteúdos, reunindo também informações que contextualizam a publicação, tais como a descrição que resume uma ideia central que será tratada no vídeo, o tipo de conteúdo, se consistia em um vídeo convencional ou uma transmissão ao vivo, a data da publicação, o número de curtidas,

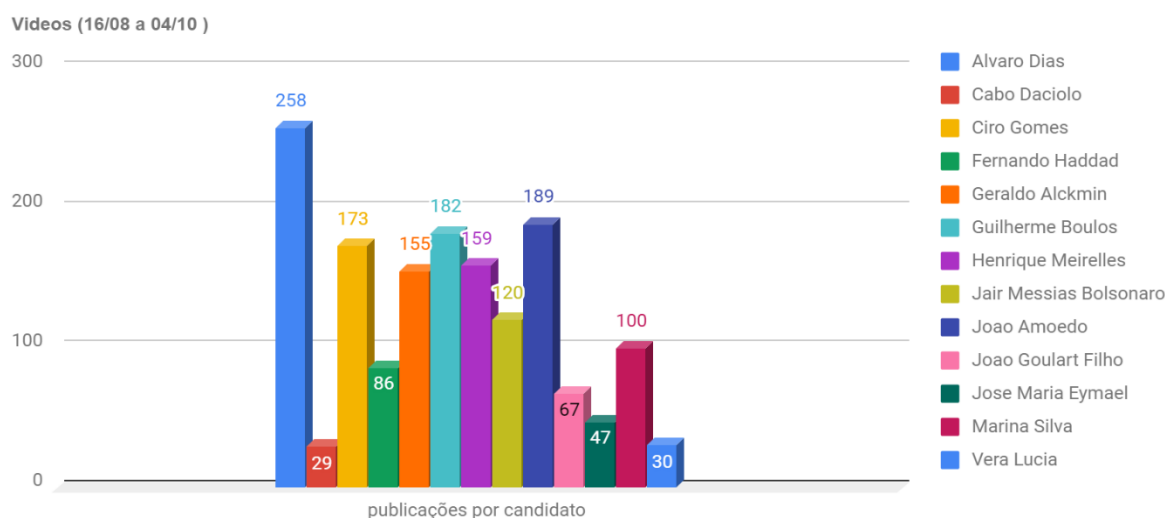


comentários, compartilhamentos e visualizações que a publicação obteve até a data da coleta. (FERREIRA, 2019).

A coleta resultou no total de 3.052 conteúdos audiovisuais divididos em: 404 registros do ex-presidente Lula e 2.142 registros dos treze candidatos oficiais durante o 1º turno das eleições e 275 registros dos dois candidatos do 2º turno. Além destes, 231 registros foram coletados fora dos períodos de campanha, pois – durante a coleta, observou-se que estes registros poderiam enriquecer o contexto por conterem informações pertinentes quanto ao apoio aos candidatos ao 2º turno, bem como suas manifestações de agradecimento aos eleitores. (FERREIRA, 2019).

Dos 1.595 vídeos convencionais publicados, observamos que sete candidatos obtiveram um número de publicações acima de 100. Como ilustra o **Gráfico 1**, o candidato Álvaro Dias atingiu o maior número de publicações – com 258 vídeos. Na sequência, apareceu João Amoêdo em 2º lugar com 189 vídeos, Guilherme Boulos em 3º lugar com 182 vídeos, Ciro Gomes em 4º lugar com 173 vídeos, Henrique Meirelles em 5º lugar com 159 vídeos, Geraldo Alckmin em 6º lugar com 155 vídeos e Jair Bolsonaro em 7º lugar com 120 vídeos. Além disso, seis candidatos aparecem com menos publicações de vídeos, Marina Silva em 8º lugar com 100 vídeos, Fernando Haddad em 9º lugar com 86 vídeos, João Goulart Filho em 10º lugar com 67 vídeos, Jose Maria Eymael em 11º lugar com 47 vídeos, Vera Lucia em 12º lugar com 30 vídeos e Cabo Daciolo em 13º lugar com 29 vídeos. (FERREIRA, 2019).

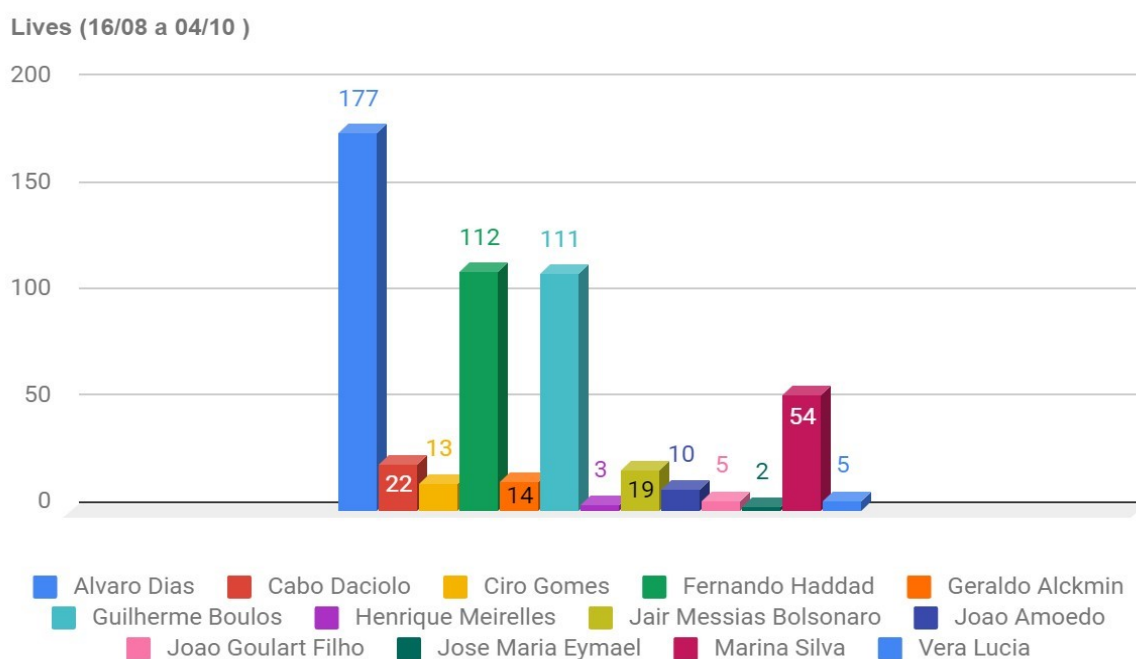
**Gráfico 1 - Quantitativo de vídeos publicados no Facebook no 1º turno por candidato.**



Fonte: Dados da pesquisa.

Embora a utilização do recurso de *live* tenha sido percebido na *fanpage* de todos os candidatos, o número de publicações nessa modalidade, no geral, foi mais baixo em todos os casos com exceção do candidato Fernando Haddad que teve mais publicações de *lives* em sua página. Como mostra o **Gráfico 2**, os candidatos que tiveram mais publicações de *lives* foram Álvaro Dias com 177 registros, Fernando Haddad com 112 registros e Guilherme Boulos com 111 registros. Marina Silva aparece com 54 registros, Cabo Daciolo com 22 registros, Jair Bolsonaro com 19 registros, Geraldo Alckmin com 14 registros, João Amoêdo com 10 registros, João Goulart Filho e Vera Lucia com 5 registros cada, e, por fim, Henrique Meirelles e Jose Maria Eymael com 3 e 2 registros respectivamente. (FERREIRA, 2019).

**Gráfico 2 - Quantitativo de *Lives* publicados no *Facebook* no 1º turno por candidato.**



Fonte: Dados da pesquisa.

Finalizamos com os dados dos 404 registros coletados da página do ex-presidente Lula, durante o primeiro turno, onde 154 eram vídeos e 250 eram *lives*, representando 38% e 62% respectivamente. Observou-se, contudo, que a maior parte dos *lives* foram produzidos por Haddad, que compartilhou os registros em ambas as páginas ou eram conteúdos antigos onde o ex-presidente aparecia gravados como *live*.

**XX ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO – ENANCIB 2019**  
**21 a 25 de outubro de 2019 – Florianópolis – SC**

O segundo turno das eleições presidenciais iniciou no dia 08 de outubro de 2018 e terminou no dia 26 de outubro de 2018. Entre os dois candidatos, Fernando Haddad se mostrou mais ativo nas publicações de conteúdos audiovisuais do que o candidato Jair Bolsonaro. Foram 101 registros disponibilizados na *fanpage* de Haddad e 58 registros disponibilizados na *fanpage* de Bolsonaro, gerando uma proporção de 64% e 36% respectivamente.

O segundo ponto importante de nossa pesquisa consistiu na identificação do formato original dos registros coletados após o término das eleições. Esta análise foi realizada considerando a quantidade total dos registros coletados durante o período de campanha eleitoral, ou seja, os 2.821 registros audiovisuais. (FERREIRA, 2019).

A análise ocorreu manualmente, nos dias 15, 16 e 17 de novembro de 2018 – aproximadamente 18 dias após o término das eleições e teve como resultado os dados da **Tabela 2**, onde temos um total de 32 vídeos e/ou *lives* indisponíveis no link original da *fanpage*. Dos 32 registros indisponíveis, 27 correspondem aos 13 candidatos e 5 são de origem da página do ex-presidente Lula. (FERREIRA, 2019).

**Tabela 2 - Persistência dos conteúdos audiovisuais nas *fanpages* dos candidatos no Facebook**  
**Novembro**

<b>Candidatos</b>	<b>Vídeos/<i>lives</i> off</b>
Álvaro Dias	0
Cabo Daciolo	2
Ciro Gomes	0
Fernando Haddad	2
Geraldo Alckmin	5
Guilherme Boulos	2
Henrique Meirelles	2
Jair Bolsonaro	3
João Amoêdo	4
João Goulart Filho	1
José Maria Eymael	5
Marina Silva	0
Vera Lucia	1
Lula	5
<b>Total</b>	<b>32</b>

**Fonte: Dados da pesquisa.**

Apenas três candidatos, em novembro, ainda mantinham todos os conteúdos audiovisuais disponíveis em suas *fanpages*: o candidato Álvaro Dias, o candidato Ciró Gomes e a candidata Marina Silva. Os demais candidatos tiveram, pelo menos, um registro

indisponível, chegando a ter até cinco registros indisponíveis de Geraldo Alckmin, José Maria Eymael e do ex-presidente Lula.

A partir desta primeira análise é possível perceber a efemeridade a que estas publicações estão sujeitas e como o fomento de pesquisas e desenvolvimentos de projetos sobre o arquivamento da *web* brasileira em redes sociais como o *Facebook*, principalmente no campo da política, é pertinente.

Considerando que, possivelmente, ainda existem estudos acerca dos impactos das comunicações políticas em mídias sociais durante as eleições presidenciais de 2018 e o próprio contexto histórico e político que nortearam estas eleições, a indisponibilidade de 32 registros, após apenas 18 dias da coleta, é reflexo da falta de compromisso dos políticos para com seus eleitores e da cultura do descarte que a *web* alimenta.

As mensagens padrão disponibilizadas no *Facebook* quando uma publicação não se encontra no seu formato original descrevem que: 1) o conteúdo não está disponível no momento ou 2) o conteúdo está indisponível devido a restrições legais. (FERREIRA, 2019).

No primeiro caso, o conteúdo pode ter sido removido da página do candidato; ter restringido o acesso a um grupo específico de usuários; ou pode ter sido removido e repostado nas mesmas condições ou com alterações, dentre outras possibilidades que esta pesquisa não se aprofundou.

No segundo caso, uma restrição legal no Brasil significa que uma pessoa ou um grupo de pessoas entrou com um processo na justiça solicitando a retirada do conteúdo da rede por entender ou provar que o registro continha informações prejudiciais.

É importante salientar que nem todos os registros indisponíveis na data de análise, necessariamente, estarão indisponíveis permanentemente uma vez que podem ter tido seu acesso apenas restringido e, assim – quando de interesse do candidato, voltarem a serem visualizados publicamente de maneira ampla ou ainda terem sido objeto de outro tipo de problema de visibilidade e acesso momentâneo. (FERREIRA, 2019).

## **5 PROTÓTIPO DE PESQUISA DOS VÍDEOS DO FACEBOOK PRESERVADOS**

O Protótipo foi um recurso necessário para realizar a coleta autônoma, estrita e semi-automatizada de registros audiovisuais disponibilizados em *fanpages* públicas do *Facebook*. E em decorrência desta necessidade, optamos por tornar livre e disponível o acesso aos

registros e dados coletados a fim de fornecer subsídios às futuras pesquisas sobre este contexto. (FERREIRA, 2019).

Como mostra a **Figura 2**, a interface *web* de pesquisa do protótipo é bem simples e possui alguns recursos flexíveis para a consulta das informações.

**Figura 2 - Interface *web* de pesquisa dos registros audiovisuais preservados.**

Protótipo - Sistema de Arquivamento da WEB

Nome Candidato  entre  e  Live?(S/N)  Ativo?(S/N)

**Fonte: Dados da pesquisa.**

A consulta pode ser realizada pelo nome do candidato ou pela opção todos, onde retornarão os registros de todos os candidatos. Ainda, é possível limitar o período de consulta, o tipo de registro que se deseja pesquisar (vídeo, live ou todos os tipos) e, ainda, a disponibilidade do formato original (sim, não ou todos). Outro recurso de consulta disponibilizado é pelo link do vídeo na origem.

Os principais recursos de visualização consistem em uma interface *web* que apresenta o(os) vídeo(s) em forma de lista a partir dos filtros selecionados, onde aparecem (**Figura 3**): nome do candidato, data da coleta, descrição do vídeo, número de curtidas, número de comentários, número de compartilhamentos, número de visualizações, link do vídeo preservado, *iframe* do vídeo no *Facebook* e se estava disponível no formato original em novembro.

**Figura 3 - Apresentação de consulta por candidato.**

Nome Candidato  entre  e  Live?(S/N)  Ativo?(S/N)

Fernando Haddad	27-10-2018 12:45	Há 43 anos, Vladimir Herzog foi assassinado pela ditadura no DOI-CODI. Essa mesma ditadura que Bolsonaro defende com convicção, sem remorso algum. Definitivamente, o Brasil não pode voltar para trás, como quer o deputado. Nós queremos construir um futuro sem violência e com esperança.	 8300	3100	10000	176893	<a href="#">2132678283716568.mp4</a>		nov/ 18: S
Fernando Haddad	27-10-2018 12:43	Coletiva de imprensa com Fernando Haddad em São Paulo #Haddad13 #ViraVoto	 16000	32000	8600	302581	<a href="#">283774408924080.mp4</a>		nov/ 18: S
Fernando Haddad	27-10-2018 12:41	Marieta Severo, Osmar Prado, Antônio Pitanga e diversos artistas se somam à nossa campanha por um Brasil sem ódio e intolerância.	 27000	6900	37000	757807	<a href="#">1135423039916129.mp4</a>		nov/ 18: S
Fernando Haddad	27-10-2018 12:40	Aprendendo uma música nova ao lado de amigos e amigas. É a onda da virada. É importante intensificar a agenda de atividades e conquistar mais votos para o Brasil sair da crise e voltar a ser feliz de novo. #Haddad13 #ViraVoto Video: Ricardo Stuckert	 26000	5900	25000	632709	<a href="#">491287991385047.mp4</a>		nov/ 18: S

**Fonte: Dados da pesquisa.**

XX ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO – ENANCIB 2019  
21 a 25 de outubro de 2019 – Florianópolis – SC

Ao selecionar/clicar no link do vídeo preservado, o usuário é direcionado para outra página que irá abrir o registro em mp4 tal qual como ele foi coletado, como mostra a **Figura 4**.

**Figura 4 - Execução do link armazenado.**





Fonte: Dados da pesquisa.

Já aonde aparece o ícone do *Facebook*, quando acionado irá abrir uma página pop-up com o *iframe* da postagem original atualizada, **Figura 5**.

**Figura 5 - Visualização do *iframe* da publicação na origem.**

Protótipo - Sistema de Arquivamento da WEB

Nome Candidato	Jair Messias Bolsonaro	entre	16/08/2018	e	27/03/2019	Live?(S/N)	Sim	Ativo?(S/N)	Todos	consultar
Jair Messias Bolsonaro	28-08-2018	11:39	Continuação Bolsonaro no Ceasa 2	12662	7134					
Jair Messias Bolsonaro	28-08-2018	11:39	Bolsonaro no CEASA 1	8701	5705					
Jair Messias Bolsonaro	22-08-2018	22:42	Desembarque no aeroporto de Presidente Prudente - SP	55256	35556					
Jair Messias Bolsonaro	17-08-2018	11:36	Formatura de Sargentos da Polícia Militar do estado de São Paulo. (Live)	34944	25038	12745	364206	<a href="https://www.facebook.com/428854424188636.mp4">428854424188636.mp4</a>		nov:18.S



Fonte: Dados da pesquisa.

Neste caso, quando o vídeo não está disponível em seu formato original aparece uma tela de fundo escuro com o aviso “vídeo indisponível”.

Todas estas informações coletadas serão migradas para a base de dados do projeto AWEB (Arquivamento das Eleições Brasileiras), em desenvolvimento pelo Núcleo de Pesquisa em Arquivamento da *web* e Preservação Digital - NUAWEB, ([www.ufrgs.br/nuaweb](http://www.ufrgs.br/nuaweb)).

No dia 03 de abril de 2019, foi realizado um registro gratuito de redirecionamento de domínio para o Protótipo no *No-IP Free Dynamic DNS*, que passa a ser nomeado como **Dumpres2018**. A sigla é uma conjunção da expressão inglesa *dump* com o prefixo da palavra presidenciais, acrescida do ano das eleições (2018). Assim, os dados poderão ser acessados, mesmo que migrando da base de dados atual pelo endereço eletrônico: **[dumpres2018.redirectme.net](http://dumpres2018.redirectme.net)**

## 6 CONSIDERAÇÕES FINAIS

A inserção dos candidatos às eleições presidenciais de 2018 de maneira ampla e ativa nas principais mídias sociais demonstra que existe uma busca pela proximidade dos candidatos com seus eleitores através da *web*. Contudo, mesmo com o uso massivo das mídias sociais pela população em geral, ainda são escassos os projetos e iniciativas voltados para a preservação dos conteúdos publicados nessas plataformas.

No entanto, a preservação de conteúdos audiovisuais na íntegra publicados em mídias sociais como o *Facebook* mostrou-se uma tarefa muito difícil e durante esta pesquisa, em diversas tentativas de automatização da coleta, foram impostas barreiras difíceis para um pesquisador transpor. Esta dualidade deixou evidente o real compromisso do *Facebook* com a parte comercial, desconsiderando os interesses social e histórico de seus usuários.

No decorrer deste trabalho surgiram diversos questionamentos que podem guiar novas discussões sobre a possibilidade de automatizar os processos de coleta em uma plataforma comercial, como o *Facebook*, que se mostra cada vez mais restritiva.

Por fim, ressaltamos que a preservação neste trabalho privilegiou os registros audiovisuais na íntegra e, como adendo, os metadados das publicações e não obrigatoriamente a estrutura da origem, como o que acontece nas iniciativas internacionais de arquivamento da *web*. Este tipo de preservação, por ser mais específico, pode fornecer resultados mais contextualizados e completos. No entanto, requer mais espaço de memória para o seu armazenamento que, embora os custos tenham diminuído no decorrer dos anos, ainda são mais proeminentes do que se fossem armazenados *snapshots*. Outro fator

importante a ser considerado é a capacidade de processamento necessária ao servidor, que – quando disponibilizado amplamente o acesso aos registros, deve proporcionar um tempo de resposta rápido para não congestionar a plataforma. Isto pode ser atenuado quando a hospedagem dos dados é realizada em servidores comerciais ou institucionais de alta performance.

## REFERÊNCIAS

BRÜGGER, Niels. *web archiving: Between past, present, and future*. **The handbook of Internet studies**, [s.l.], p. 24-42, 2011.

COSTA, Miguel; GOMES, Daniel; SILVA, Mário J. The evolution of *web archiving*. **International Journal on Digital Libraries**, [s.l.], v. 18, n. 3, p. 191-205, 2017.

FERREIRA, Lisiane Braga. **Arquivamento da *web* e mídias sociais: preservação digital de vídeos da campanha presidencial brasileira de 2018**. 2019.

FERREIRA, Lisiane Braga, MARTINS, Marina Rodrigues, ROCKEMBACH, Moisés. Usos do Arquivamento da Web na Comunicação Científica. **Prisma.com**, [s.l.], n.36, 2018.

Gil, Antônio Carlos. **Como elaborar projetos de pesquisa**. São Paulo: Atlas, 4. ed., 2002.

LIBRARY OF CONGRESS. **Brazilian Presidential Election 2010 *web* Archive**. Disponível em: <https://www.loc.gov/collections/brazilian-presidential-election-2010-web-archive/about-this-collection/>. Acesso em: jun. de 2018.

POP, Radu; VASILE, Gabriel; MASANÈS, Julien. Archiving *web* video. In: **International *web* Archiving Workshop IWAW 2010**. 2010.

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

SCHNEIDER, Steven M. et al. Building thematic *web* collections: challenges and experiences from the September 11 *web* Archive and the Election 2002 *web* Archive. **Digital Libraries**, ECDL, p. 77-94, 2003.

STATISTA. **Number of Facebook users in Brazil from 2017 to 2023 (in millions)**. Disponível em: <https://www.statista.com/statistics/244936/number-of-facebook-users-in-brazil/>. Acesso em: jul. de 2019. 2019

TRIBUNAL SUPERIOR ELEITORAL. **Confira as principais datas do calendário eleitoral das Eleições Gerais de 2018**. [Brasília: TSE], 2017. Disponível em: <http://www.tse.jus.br/imprensa/noticias-tse/2017/Dezembro/confira-as-principais-datas-do-calendario-eleitoral-das-eleicoes-gerais-de-2018>. Acesso em: abr. de 2018.

VOERMAN, Gerrit et al. Archiving the *web*: Political party *web* sites in the Netherlands. **European Political Science**, [s.l.], v. 2, n. 1, p. 68-75, 2002.