

SOCIOMATERIALIDADES DO ESQUECIMENTO E POLÍTICAS DAS PLATAFORMAS: desafios para preservar rastros digitais do 8 de janeiro¹

SOCIOMATERIALITIES OF ERASURE AND PLATFORM POLITICS: challenges to preserve the digital traces of January 8th

Marcelo Alves dos Santos Junior ²

Resumo: Em 08 de janeiro, pessoas vandalizaram os prédios dos Três Poderes em Brasília, transmitindo seus atos ao vivo por plataformas digitais. Meses depois, boa parte desses vídeos e do conteúdo que fomentou essas ações foram deletados. Este artigo aborda as políticas de esquecimento digital, discutindo as condições sociomateriais, as agências da governança de plataformas e o impacto do apagamento sistemático de vídeos. Aponta desafios associados às affordances sociotécnicas como a efemeridade e o atual declínio na transparência. Propõe apontar caminhos para a preservação e compartilhamento de fontes primárias digitais, destacando a importância de arquivos digitais acessíveis para pesquisas futuras. O estudo visa contribuir para a preservação de acervos de contextos históricos em plataformas digitais, considerando questões éticas, tecnológicas e de regulação.

Palavras-Chave: Governança de Plataformas. Humanidades digitais. Sociomaterialidades. Métodos Digitais. Esquecimento.

Abstract: On January 8, protesters vandalized the buildings of the Three Powers in Brasília, broadcasting their actions live on digital platforms. Months later, much of these videos and the content that fueled these actions was deleted and can no longer be found. This article addresses the politics of digital forgetting, discussing sociomaterial conditions, platform governance agencies and the impact of the systematic erasure of digital content. It points out challenges associated with socio-technical affordances such as ephemerality and the current decline in transparency. It proposes ways of preserving and sharing digital primary sources, highlighting the importance of accessible digital archives for future research. The study aims to contribute to the preservation of collections of historical contexts on digital platforms, considering ethical, technological and regulatory issues.

Keywords: Platform Governance. Digital humanities. Sociomaterialities. Digital methods. Forgetting.

1 Trabalho apresentado ao Grupo de Trabalho Materialidades Digitais e Práticas Comunicacionais. 34º Encontro Anual da Compós, Universidade Federal do Paraná (UFPR). Curitiba - PR. 10 a 13 de junho de 2024.

2 Professor do Programa de Pós-Graduação em Comunicação da PUC-Rio, Doutor em Comunicação pela UFF, marcelo_alves@puc-rio.br.

1. Introdução

No dia 08 de janeiro, uma multidão insurgente vandalizou os prédios dos Três Poderes em Brasília. Durante a depredação das instituições democráticas nacionais, os autointitulados patriotas gravaram cada detalhe, transmitindo vídeos e imagens em suas contas nas plataformas digitais em tom de triunfo. A insurreição golpista foi um fenômeno político-midiático, em grande parte, construído nas plataformas digitais. Mais de dois anos depois do 08 de janeiro, contudo, os rastros dessa história estão rapidamente sendo apagados e a extrema-direita empreende um esforço de revisionismo e negacionismo para normalizar os atos antidemocráticos e pautar um projeto de anistia. O problema de pesquisa desse artigo são as condições sociomateriais dos rastros digitais e as políticas do esquecimento, a fim de construir um arcabouço teórico-metodológico para fomentar projetos de preservação e compartilhamento de fontes primárias para a pesquisa acadêmica.

Levantamento realizado pela organização não-governamental SumOfUs identificou dezenas de coberturas sendo transmitidas ao vivo no Youtube direto da Esplanada durante o 08 de janeiro, algumas atingindo audiências de cerca de 1 milhão visualizações. No entanto, na semana seguinte, a maior parte dos vídeos já estava indisponível em função da remoção pelas plataformas digitais ou por decisões do TSE. Além disso, os próprios autores dos conteúdos começaram a deletar as gravações e as imagens da insurgência em resposta às notícias das prisões de milhares de pessoas. A remoção dos vídeos como estratégia de evasão de responsabilização jurídica não foi acidental ou isolada. Análise da Novelo Data publicada pelo jornal O Globo demonstrou que mais de 1700 vídeos haviam sido apagados³.

Dessa forma, fontes relevantes para o estudo desse acontecimento estão sendo rapidamente deletadas, implicando na deterioração das condições materiais de reconstruir essa história. Estudos no campo da memória tradicionalmente utilizam fontes secundárias como as reportagens publicadas pela imprensa jornalística para reconstituir um período. Todavia, no caso do 08 de janeiro, esse recorte é insuficiente, na medida em que perde de vista o processo de mobilização e construção de discursos, que teve como meio principal as plataformas digitais.

Em pesquisa anterior, identificamos que, em aferição realizada em fevereiro de 2023, 31% da produção de uma rede de canais de extrema-direita no Youtube não estavam

3 Disponível em: <https://oglobo.globo.com/blogs/sonar-a-escuta-das-redes/noticia/2024/02/14/canais-bolsonaristas-deletam-mais-de-15-mil-videos-apos-operacao-da-pf-contr-o-ex-presidente.ghtml>. Acesso em 02 de junho de 2024.

disponíveis na plataforma, totalizando 26 mil vídeos e mais de 500 milhões de visualizações perdidas (Autoria, 2023). Isso quer dizer que uma pesquisa iniciada anos depois do acontecimento chegaria a resultados completamente diferentes, encontrando muito menos desinformação e discursos antidemocráticos.

Quando uma publicação é deletada de uma plataforma digital, ela também é removida de todos os mecanismos de busca disponíveis para a pesquisa acadêmica. Milligan (2013) aponta que a internet se tornou um grande cemitério de histórias perdidas e links quebrados. Se esse processo de apagamento é feito sistematicamente e em escala massiva, há um grande risco de distorção e enviesamento de processos de reconstrução da memória sobre eventos históricos. No caso do 08 de janeiro, o revisionismo aconteceu quase imediatamente quando parlamentares e influenciadores digitais bolsonaristas construíram o discurso de que as pessoas eram “presas políticas” de um “regime ditatorial” e os atos de violência foram provocados por “infiltrados de esquerda”.

Não há um acervo de pesquisa sobre as plataformas digitais no Brasil. O objetivo deste artigo é debater as implicações das condições epistêmicas dos estudos de acontecimentos mediados pelas plataformas digitais. Particularmente, visa articular o problema a partir de um marco teórico que relaciona conceitos da sociomaterialidade das infraestruturas, estudos de plataformas, métodos digitais e humanidades digitais nas linhas de história e memória. Em diálogo com os estudos de plataformas e os métodos digitais, argumenta-se que múltiplos agenciamentos, sociais, tecnológicos e políticos, influenciam as condições concretas de pesquisa e de construção de bases de dados sobre eventos históricos e que diversas *affordances*, entre elas, a efemeridade dos rastros digitais e efeitos difusos da moderação de conteúdo constituem as condições de recuperabilidade das fontes primárias disponíveis para resgate a posteriori.

Dessa forma, a contribuição desse artigo é lançar bases iniciais para refletir sobre as possibilidades e limitações para a preservação de publicações, vídeos e imagens circulantes em plataformas digitais para consulta futura. A produção de acervos digitais e compartilhamento de dados nesses espaços enfrenta uma série de questões que envolvem dimensões de regulação, ética, tecnologia e geopolítica. Assim, pretende-se somar esforços para o resgate, preservação e compartilhamento de fontes primárias de publicações para fomentar redes de pesquisa nacionais.

2. Sociomaterialidades de plataformas digitais enquanto acervos de pesquisa histórica

Este artigo dialoga com arcabouços teórico-metodológicos da comunicação, métodos digitais e história digital. Partimos do princípio de que fenômenos e eventos altamente mediatizados em nossa sociedade contemporânea, sobretudo aqueles de comunicação política, são co-constituídos pelas estratégias de comunicação digital (ROGERS, 2015). Dessa forma, os rastros digitais compostos por dados, metadados, capturas de tela e outras formas de arquivamento são centrais para preservar a história política de acontecimentos como o de 08 de janeiro. É a partir da construção de acervos digitais de bancos de dados que se pode reconstituir a memória da contestação eleitoral e dos atos intervencionistas e produzir conhecimento acadêmico sobre a comunicação digital da extrema-direita (MAYNARD, 2014).

Nesse sentido, é importante situar as plataformas digitais e suas sociomaterialidades como interagentes importantes não somente porque detêm um acervo de publicações que representam um momento histórico de diversas culturas e civilizações no mundo inteiro, mas também porque suas *affordances* e políticas de governança influenciam na natureza desses dados. Ao estudar o processo de digitalização das imagens pessoais na web, van Dijck (2007) cunhou o conceito de memórias mediadas para analisar o processo de digitalização das práticas que ativam a memória, focando, sobretudo, na relação das inscrições realizadas nas mídias digitais com os contextos culturais em que estão inseridas.

Assim como a construção da memória pode ser pensada de forma sociotécnica, é possível argumentar no sentido das sociomaterialidades do esquecimento, na medida em que a destruição, manipulação e enviesamento de arquivos são táticas antigas de revisionismo historiográfico. Por sociomaterialidades do esquecimento, entende-se as agências tecnológicas, políticas e individuais sobre o estatuto de vivência de um determinado conteúdo online em uma conjuntura em que a “guarda” desses documentos tangencia a institucionalidade de bibliotecas e do bem comum em direção de uma detenção compartilhada entre a plataforma e as complementadoras que criaram o conteúdo. Nesse sentido, a

preservação desse material é instável e sujeita a múltiplas tentativas de apagamento, ainda que inicialmente bem-intencionadas no sentido de combate à desinformação.

As bibliografias da história digital e dos métodos digitais têm se aprofundado na concepção e nos procedimentos aplicados do fazer historiográfico a partir de fontes digitais. Milligan (2018) argumenta que o processo historiográfico é transformado pelo uso da web como fonte primária de estudos da história, particularmente, por meio das análises sistemáticas de arquivos digitais que preservam as materialidades dos rastros digitais. Brügger (2018) define a história digital como um arcabouço conceitual que encapsula diversas práticas de uso de tecnologias digitais na produção de relatos historiográficos. Parte da literatura sobre história digital se debruça sobre a digitalização das obras e fontes históricas para a web (BRASIL; NASCIMENTO, 2020). Esse artigo situa seu objeto a partir do que Brügger (2018) chama de *web history*”, ou seja, uma prática historiográfica que considera os rastros digitais como fontes primárias para reconstituir um contexto sociopolítico a partir da web arquivada.

Debatendo as quatro tradições historiográficas digitais, Rogers (2018) define as coleções baseadas em eventos como uma forma de preservar redes ou esferas da web que se formam em torno de acontecimentos históricos marcantes como o ataque terrorista de 11 de setembro de 2001 nos Estados Unidos. Nesse sentido, essa pesquisa expande essa concepção de arquivamento construído em torno de eventos para uma abordagem focada em plataformas digitais como ponto de partida, compreendendo-as como infraestruturas sociotécnicas distintas, ainda que com conexões, da web em geral.

No Brasil, começou a se consolidar, nas últimas duas décadas, uma subárea de estudos das humanidades digitais voltada para a historiografia. Lucchesi (2014) sustenta que ocorre uma certa “virada digital” nas ciências humanas e analisa a bibliografia que nasce no começo dos anos 2000 para compreender os desafios da memória digital. A abordagem adotada nesta pesquisa é configurada pelo registro baseado em eventos (ROGERS; BRÜGGER;

MILLIGAN, 2018), na medida em que busca preservar a história da cultura digital que levou ao 08 de janeiro no Brasil.

Teles (2014) sustenta que a história do tempo presente é desafiada pela rápida circulação de informações na internet e que é necessário um rigor metodológico para evitar revisionismos e incompreensões. “A Web tem uma quantidade infinita de conteúdo, mas, ao mesmo tempo, uma experiência comum é que o que estava na Web ontem ou há um ano não está mais lá. Ele pode ter sido movido, alterado ou excluído” (BRUGGER, 2018, p. 75).

Nesse sentido, os arquivos da web são fontes relevantes de pesquisa sobre os fenômenos que são, em alguma medida, constituídos digitalmente. Webster (2018) explica que a preocupação em preservar a web está diretamente conectada com a necessidade de compreender as expressões culturais de sociedades que, cada vez mais, se comunicam utilizando meios digitais. Ele explica que o Internet Archive é o primeiro e mais relevante projeto de arquivamento da web, criado em 1996 e em operação ininterrupta até hoje. Embora seja um acervo de extrema importância, foi desenhado metodologicamente para armazenar a web por meio da técnica de *crawling* dos hiperlinks e possui grandes limitações quando se trata da web plataformizada (HELMOND, 2015).

Dessa forma, é necessário compreender as sociomaterialidades específicas da web e das plataformas digitais para construir e gerenciar os acervos que analisam a história a partir da internet. Outros estudos já argumentaram em favor das mídias sociais como fontes históricas. Bruns e Weller (2016), por exemplo, defendem a preservação do Twitter como um artefato histórico que constrói um “primeiro rascunho do presente”, dada a característica da plataforma de troca de opiniões durante o desenrolar dos acontecimentos.

As particularidades das plataformas digitais enquanto infraestruturas privadas que estão apartadas da web mais geral demandam procedimentos metodológicos e instrumentos técnicos específicos para a elaboração de arquivos digitais. Embora ainda seja uma temática muito recente, algumas propostas vêm surgindo na literatura especializada para estruturar o armazenamento sistemático da web. Pehlivan; Thièvre e Durgeon (2021) explicam que um

ponto central é a seleção, tendo em vista que a decisão sobre o que arquivar é fundamental dada a vastidão das plataformas digitais. As pesquisas sugerem protocolos que levam em consideração recortes de listas de atores relevantes, *hashtags* e delimitações geográficas.

Ben-David (2020) adota uma perspectiva crítica sobre a preservação da web. Segundo ela, em contexto de fechamento dos projetos de compartilhamento de dados das plataformas digitais com os pesquisadores, a prática de arquivamento emerge como uma forma de resistência contra a apropriação de dados públicos por estas empresas. Ben-David (2020, p. 255) sugere o conceito de contra-arquivamento, derivado de estudos críticos sobre o colonialismo, como uma ação de resistência epistêmica contra as plataformas digitais. “Eu proponho a construção de arquivos do Facebook que são desenhados para contrariar os protocolos da plataforma de acesso ao conhecimento [...] que questionam o direito do público ao arquivo das mídias sociais”.

3. Problemas para estudos historiográficos das plataformas digitais

O problema de pesquisa é atravessado por uma transformação no paradigma da comunicação a partir de uma ampla infraestrutura midiática digital que opera coordenando diversas plataformas digitais (D’ANDREA, 2022; DOS SANTOS, 2021; PASQUETTO et al., 2022). Compreender detalhadamente as estruturas e processos de comunicação digital é central para elucidar como se configuraram o discurso e as mobilizações antidemocráticas. Todavia, um conjunto de limitações dificulta o desenvolvimento de estudos de fôlego sobre essas questões. A primeira seção debate a questão das políticas e sociotécnicas do acesso às bases de dados. A segunda expande sobre a governança das plataformas digitais e seus efeitos sobre os fenômenos empíricos.

3.1 Políticas dos dados: acesso e recuperabilidade

Em geral, estudos que se dedicam a compreender fenômenos e culturas digitais são episódicos e voltados para recortes temporais sincrônicos e recentes, como a última eleição ou um ciclo de manifestações. Quando se tem a finalidade de lançar olhares longitudinais ou analisar momentos relativamente distantes do ponto de partida temporal da pesquisa, todavia,

emergem uma série de questões metodológicas relevantes. Nesse sentido, dois aspectos são centrais para conduzir a pesquisa: o acesso e a recuperabilidade.

Os estudos empíricos sobre plataformas digitais, sobretudo os que se fundamentam, em alguma medida, nos métodos digitais historicamente se constituíram em torno do que Venturini e Rogers (2019) chamam de “pesquisa baseada em APIs”. Isto é, uma apropriação de materialidades tecnológicas das plataformas repropostas para fins de investigações acadêmicas, como a construção de bases de dados por meio das consultas às interfaces programáticas. Mais recentemente, esse modelo vem sofrendo diversas descontinuações e interrupções no que se convencionou chamar de APIcalypse e pesquisa pós-API (Periam, Birkbak e Freeman, 2020; Bruns, 2021).

D’Andrea (2021) debate as políticas da APIs situando um conjunto de *affordances*, governanças e mediações algorítmicas que agem no sentido de predefinir ações possíveis, delimitar o que é visível e como esses traços são reconfigurados em formatos que podem ser compartilhados com terceiros para diversos fins. Nos últimos anos, desde o escândalo da Cambridge Analytica em 2016 e, mais especificamente, com a compra do Twitter por Elon Musk em outubro 2022, vem sendo instaurado um paradigma contrário à transparência e à observabilidade programática nas plataformas digitais.

Assim, o primeiro desafio é o atual cenário de “apagão de dados” das plataformas digitais. Nos últimos anos, Meta e X (Twitter) implementaram políticas de opacidade que impedem ou dificultam a transparência sobre os fenômenos que ocorrem em seus sites. Chamado na literatura acadêmica de pesquisa pós-API, somente o Youtube e o Telegram oferecem acesso aberto a fontes de dados para pesquisadores (BRUNS, 2019; FREELON, 2018; TROMBLE, 2021). Facebook e Instagram encerraram o programa de compartilhamento de dados, instaurando um procedimento de certificação de pesquisa centralizado na Universidade de Michigan, o X de Elon Musk descontinuou os projetos de colaboração e impôs um custo de 5 mil dólares mensais; e o TikTok só autoriza requerimentos de pesquisadores vinculados a centros dos Estados Unidos e Europa.

Assim, há um retrocesso nas condições concretas de estudo que impossibilita a elaboração de bases de dados nas plataformas digitais por meios oficiais. Esse retrocesso estimulou um exercício metodológico na proposição de alternativas, como a doação de dados e um resgate de técnicas de interface gráfica como a raspagem, que implica na quebra do termo de serviço das empresas e potenciais implicações legais e éticas (Freelon, 2018).

Levando em conta esses desafios, Mancosu e Vegetti (2020, p. 8) propõem uma rotina de raspagem que visa proteger a privacidade dos usuários ao inserir etapas de pseudo-anonimização das identidades durante a extração: “não acreditamos que os pesquisadores que coletam dados públicos do Facebook para fins científicos corram o risco de sofrer repercussões muito negativas. No entanto, é importante estar ciente das possíveis consequências que podem ocorrer”.

3.2 Efemeridade da web: múltiplas agências e declínio do conteúdo

O segundo entrave é que, mesmo que o problema do acesso seja endereçado de algum modo, boa parte das publicações sobre o período histórico em questão já não existe. Estudos sobre história digital apontam que quanto mais se distancia do período de interesse, menor a qualidade da amostra de dados ainda disponível online (BRÜGGER, 2018). Quando um conteúdo é removido, não é possível resgatar a forma original; nenhum programa de transparência acadêmica preserva postagens que foram deletadas (BEN-DAVID, 2020). Por vezes, nem mesmo o sistema de justiça consegue resgatar as provas de um processo. A Meta respondeu ao Ministério Público Federal (MPF) não ter uma cópia do vídeo compartilhado, e posteriormente apagado, por Bolsonaro no Facebook contestando o resultado das eleições. O material foi preservado e cedido ao MPF pelo site de arquivamento digital MetaMemo.

Assim como as plataformas digitais agem sobre as condições de acesso, formatação, seleção e consumo de dados por meio de suas interfaces programáticas, também incidem de diversas maneiras sobre a própria disponibilidade dessas publicações online. Novamente, os métodos digitais contribuem com um arcabouço robusto para situar essas agências tecnológicas com o objetivo de se apropriar de forma ativa desses efeitos para consolidar desenhos de pesquisa. De Keulenaar e Rogers (no prelo) propõem uma abordagem de métodos digitais para resgatar os estudos dos traços digitais por meio dos efeitos das plataformas na moderação de conteúdo e as reações das pessoas contra a aplicação de sanções e, no limite, o banimento de suas contas.

Nesse sentido, os estudos sobre as políticas e práticas de moderação de conteúdo destacam a não-organicidade dos dados digitais a partir de um olhar refinado para as diferentes formas pelas quais as sociomaterialidades influenciam a própria ontologia desses

rastros. Para o contexto desta pesquisa, essa perspectiva é fundamental na medida em que a disponibilidade de dados digitais nas plataformas é instável e absolutamente contingente de múltiplas agências (De Keulenaar, 2021; Autor, 2024). Sejam elas tecnológicas, as próprias empresas podem decidir por remover o conteúdo; institucionais, como ordens judiciais para deletar ou suspender o acesso em determinada localidade, ou mesmo individuais, na medida em que as próprias audiências elaboram estratégias baseadas em imaginários algorítmicos para contornar a governança de plataformas e eventuais responsabilizações judiciais.

Além disso, a própria gramatização das ações nas plataformas pré-estrutura as dinâmicas de sociabilidades em determinadas práticas e modelos de interface (Gerlitz e Rieder 2018). Nesse sentido, uma das principais *affordances* indicada por Boyd (2010), a noção da persistência dos registros das sociabilidades digitais precisa ser revista em função da ideia de efemeridade e da contingência. Com a popularização do Snapchat, as *affordances* da temporalidade inconstante e do apagamento programado das publicações incentivaram novas culturas de uso (Bayer, 2016). Mais recentemente, a adoção de *affordances* temporárias em plataformas como o Instagram nos *stories* e na opção de não arquivar transmissões em tempo real introduz mais desafios para a preservabilidade das interações mediadas.

Esses desafios não configuram necessariamente uma impossibilidade de estudar rastros digitais, mas demandam maior protagonismo, crítica tecnológica e agência da pesquisa para a construção de bancos de dados. Por vezes, essas dinâmicas podem relevar outras formas de reconstituir peças do quebra-cabeça de eventos em tempo real a partir das culturas de dispositivos, como a identificação de lives apagadas em seu canal original sendo retransmitidas e remixadas ao vivo por youtubers que fazem *react*.

A efemeridade da web, e das plataformas digitais em particular, é bem conhecida pelos estudos de arquivos digitais. Bruns e Weller (2016) citam a iniciativa da Biblioteca do Congresso dos Estados Unidos, instituição pública que armazena o maior arquivo do Twitter, todavia, sem disponibilizar para pesquisadores. Considerando o dinamismo das atualizações,

os Bruns e Weller (2016, p. 4) explicam que existem várias dificuldades para arquivar o site, sobretudo quando uma publicação é apagada e seu conteúdo é perdido. “Isso cria um grande problema [...] *tweets* apagados podem ser os mais interessantes para historiadores do futuro, como em casos de declarações controversas de políticos importantes”.

4. Caminhos possíveis: Acervo Digital do 08 de Janeiro

Dada a relevância e a efemeridade dos registros visuais sobre o 08 de janeiro, algumas iniciativas se preocuparam em arquivar diferentes tipos de dados. Beiguelman et al (2024) desenvolveram um arquivo digital⁴ das imagens das câmeras de vigilância captadas no Palácio do Planalto no dia da depredação ao prédio da presidência. O projeto Lupa nos Golpistas, da Agência Lupa, realizou reportagens jornalísticas e com base em um banco de dados de 3,6 mil publicações enviadas pelo público por meio de um formulário⁵. No entanto, ainda não há um repositório com fins de pesquisa acadêmica que preserve, organize, catalogue e ofereça acesso para publicações de plataformas digitais sobre esse evento.

Essencialmente, uma das principais formas de preservar as informações digitais é guardar cópias de dados ou de metadados, registradas antes da publicação original ser removida (BEN-DAVID, 2020; BRÜGGER, 2018; DE KEULENAAR; BURTON; KISJES, 2021). Desde setembro de 2021, mantenho um sistema de coleta, processamento e armazenamento de publicações que arquivou diariamente um vasto acervo de informações. Nesse sentido, trata-se de um conjunto de informações extremamente valiosas porque não podem ser recuperadas a posteriori e já foram inteiramente armazenadas fisicamente. Essa técnica preserva o acervo independente da existência digital do conteúdo original na plataforma em que foi postado inicialmente.

A ideia é construir um Acervo Digital do 08 de Janeiro como arquivo contendo metadados, textos, imagens e vídeos das publicações que foram compartilhadas nas plataformas digitais entre outubro de 2022 e fevereiro de 2023. O arquivo será organizado em torno de três eixos. O primeiro é composto pelas principais plataformas digitais do Brasil - Facebook, Instagram, X (Twitter), Youtube, TikTok e Kwai; o segundo explora as chamadas

4 Disponível em: <https://www.acervosdigitais.fau.usp.br/8-1-narrativas-dissidentes-da-arte-da-arquitetura-e-da-cultura-digital/>

5 Disponível em: <https://lupa.uol.com.br/institucional/2023/01/13/lupa-nos-golpistas-recebe-mais-de-3-6-mil-denuncias>

alt-tech que abrigam discursos extremistas e conspiratórios, como Gettr, BitChute e Rumble; e o terceiro apresenta os websites do ecossistema midiático de extrema-direita.

O acervo possui três objetivos específicos:

a) Preservar: armazenar um grande volume de dados e metadados contendo textos, imagens, vídeos e estatísticas das publicações que foram compartilhadas durante o período de análise. O arquivo deve estar em formato digitalizado em cópia física, de forma a ser preservado em caso de apagamento do conteúdo original nas plataformas digitais;

b) Catalogar: considerando o vasto conjunto de informações extraídas, é necessário adotar técnicas de gestão para limpar, organizar, tratar, de-duplicar e indexar o arquivo, a fim de permitir buscas e filtragens por contas, palavras-chave, plataformas, recorte geográfico e intervalo temporal. Além disso, a maior parte do acervo é composta por imagens e vídeos que receberão tratamento utilizando técnicas de inteligência artificial para realizar o Reconhecimento Ótico de Caracteres (OCR), extração de entidades, transcrição dos áudios e catalogação de categorias (PEARCE et al., 2020);

c) Compartilhar: criar ferramentas de bancos de dados e mecanismos buscáveis para facilitar a difusão de conhecimento para pesquisadores de forma rápida e gratuita.

5. Desafios para um projeto de preservar rastros do 08 de Janeiro

Uma revisão da bibliografia acadêmica sobre arquivos de plataformas digitais e compartilhamento de bases de dados para replicabilidade e produção de novos estudos aponta para um conjunto de entraves, em menor ou maior medida, regulatórios e de governança que dificultam a sua execução. As seções abaixo resgatam esses debates e propõem possibilidades levando em consideração dois dilemas principais relacionados às formas de obtenção desses rastros e das dificuldades de compartilhamento com a comunidade acadêmica. Para além de questões técnicas, o foco principal será a insegurança relativa à governança

5.1 Dilemas da obtenção dos dados

Com a restrição das permissões para acessar as interfaces programáticas das plataformas digitais para consultas com fins de pesquisa, a principal, senão a única, alternativa para investigar os rastros utilizando métodos digitais é a raspagem da interface gráfica. Trezza (2023) identificou uma migração da comunidade acadêmica italiana para *scraping*, debatendo a técnica como um “mal necessário” e que deve ser conduzida sob

estritos termos de proteção de dados privados. Marres e Weltevrede (2013) definiram a raspagem como uma prática analítica para coletar dados fundamentada em uma infraestrutura sociotécnica prévia que é reproposta para as investigações, como as pesquisas sobre os algoritmos de ordenamento do Google (Rogers, 2009). Técnica frequentemente utilizada nas indústrias digitais para finalidades comerciais, ela ainda possui fundamentação jurídica incerta e é desautorizada pelos termos de serviço das plataformas digitais, o que aumenta as incertezas da prática acadêmica.

Nesse sentido, Leerssen, Heldt e Kettermann (2023) classificam a situação atual como uma precaridade legal e técnica do acesso aos dados de plataformas digitais com fins de interesse público. A obtenção de dados por meios não autorizados pelas gigantes tecnológicas gera uma insegurança jurídica de acordo com potenciais violações dos termos de serviço que podem ser interpretadas legalmente, como as legislações que se concentram em *hacking*, propriedade intelectual, direitos autorais e de privacidade. Ainda que essas normas não enderecem questões especificamente de pesquisa ou de interesse público, elas podem ser instrumentalizadas para perseguição de cientistas com fins de reduzir a transparência e a responsabilização das plataformas digitais.

Um dos primeiros casos de repercussão legal da raspagem de dados foi o processo movido pela American Civil Liberties Union (ACLU) para contestar a constitucionalidade do uso da lei de Computer Fraud and Abuse Act (CFAA) nos Estados Unidos para criminalizar a pesquisa acadêmica. O caso movido pelo pesquisador Christian Sandvig em 2016 retrata um estudo de auditoria de sistemas digitais para identificar vieses de discriminação racial e de gênero na contratação de pessoas. Sandvig (2017) argumenta que a pesquisa não deve ser criminalizada, seja objetivamente em processos legais, ou como um efeito inibidor que desestimula os estudos digitais e o jornalismo investigativo. Em 2020, a Corte Federal decidiu que a pesquisa em função do interesse público não pode ser criminalizada por violar os termos de serviço de empresas digitais⁶.

Ainda assim, as gigantes tecnológicas implementam uma série de medidas extraoficiais para dificultar análises independentes. Uma equipe da New York University (NYU) teve suas contas bloqueadas no Facebook em função do projeto Ad Observatory, que extraía informações sobre anúncios políticos exibidos para as pessoas durante a navegação na

6 Disponível em: <https://www.aclu.org/press-releases/federal-court-rules-big-data-discrimination-studies-do-not-violate-federal-anti>. Acesso em 15 de janeiro de 2025.

plataforma. Mesmo com o consentimento da raspagem, a empresa instrumentalizou preocupações com a privacidade para interromper projetos exercendo um poder de veto unilateral⁷.

Esse modelo de autogovernança reproduz uma assimetria de poder que contribui para a perda da transparência sobre esses sistemas sociotécnicos e dificulta análises independentes. Assim, as demandas pela soberania digital e as disputas geopolíticas num contexto de imperialismo de plataformas se traduzem, também, na restrição da transparência, lobby contra as medidas regulatórias e tentativa de criminalizar a pesquisa acadêmica de interesse público e independente das grandes corporações tecnológicas.

Compartilhamos do argumento de Ben-David (2020, p. 256) sobre o contra-arquivamento como uma prática de resistência contra a dominação dessas empresas: “explicitamente propõe coletar, reorganizar e republicar dados do Facebook que explicitamente vão contra a ordem de arquivamento do Facebook”. Contudo, é necessário cautela e nuance ao ler o texto da autora⁸, que colapsa a legalidade com os termos de serviço. No limite, é necessário resgatar o debate crítico sobre a imposição dos interesses das plataformas digitais contra o interesse público a fim de salvaguardar práticas de pesquisa que obtêm bases de dados sem autorização das empresas, desde que respeitando cuidadosamente o direito à privacidade das pessoas.

5.2 Dilemas do compartilhamento de dados

Uma segunda linha de desafios a serem endereçados são pertinentes à própria lógica de compartilhamento de dados. Do ponto de vista científico, a publicação da base empírica é considerada boa prática para garantir a replicabilidade dos resultados, validação externa e, até mesmo, estimular novas pesquisas. No entanto, as precariedades jurídicas dos rastros de plataformas digitais suscitam dificuldades que vêm coibindo o fomento de bases de dados comuns e a articulação de redes de pesquisa em torno dessas fontes primárias.

⁷ Disponível em: <https://www.theverge.com/2021/8/4/22609020/facebook-bans-academic-researchers-ad-transparency-misinformation-nyu-ad-observatory-plug-in>. Acesso em 15 de janeiro de 2025.

⁸ O trecho do texto “Como coleções voltadas para o público, os contra-arquivos responsáveis do Facebook devem ser legais e não devem violar os termos de uso da plataforma” (Ben-David, 2022, p. 256) enfraquece o próprio argumento da autora, na medida em que busca exercer uma espécie de “resistência responsável” sem debater profundamente os termos de serviço enquanto contrato comercial que visa proteger os interesses das empresas e não o bem comum.

Zimmer (2015) analisa a coleção de dados cedida pelo Twitter à Biblioteca do Congresso dos Estados Unidos, que guardou todos os tweets públicos na íntegra de 2006 até o final de 2017⁹. O projeto tinha o objetivo de preservar um acervo valioso da memória cultural do período digital e compartilhar com pesquisadores certificados para fins de pesquisa, proibindo a reprodução do conteúdo ou propósitos comerciais. A iniciativa, no entanto, foi limitada por questões técnicas e de ordem política. O autor debate como restrições do acordo entre a Biblioteca e o Twitter impossibilitaram o acesso aberto ao acervo e desafios acerca da expectativa de privacidade e direito ao esquecimento das pessoas que postaram na plataforma durante esse período. Zimmer (2010) argumenta que a justificativa de que os dados “já são públicos” é insuficiente para desabonar as questões éticas que surgem dessas coleções ao ignorar as normas contextuais de privacidade.

Já do ponto de vista de dados compartilhados para fins de pesquisa, os termos de serviço do Twitter autorizavam apenas a publicação dos identificadores únicos dos tweets, desconsiderando quaisquer outros metadados. “Embora isso pode ter sido estabelecido como uma forma de proteger a privacidade dos usuários, limita significativamente a preservação de dados do Twitter e trabalhar com coleções arquivadas” (Bruns e Weller, 2016, p. 3). Isso gerou uma prática chamada de reidratação dos tweets, em que pesquisadores publicavam uma lista de IDs para terem seus metadados extraídos novamente. A interrupção da API pública do Twitter em 2023, todavia, impediu essa possibilidade, que também nunca foi muito eficiente (Chan, Schatto-Eckrodt, Gruber, 2024).

Outro ponto relevante é a requisição de deletar metadados de *tweets* que foram removidos da plataformas por quem as publicou. Divulgada como uma iniciativa que reforça a autonomia de escolha das pessoas, Bruns e Weller (2016) destacaram como essa prática arrisca inviabilizar arquivos sobre plataformas digitais. Considerando um projeto de arquivamento do 8 de janeiro, essa remoção destruiria grande parte do conteúdo que foi apagado para evitar responsabilização legal, contribuindo para apagar os rastros desses acontecimentos, efetivamente como uma antítese do projeto. Nesse sentido, eles propõem uma série de medidas para evitar a perda de publicações das plataformas digitais, como coletar informações disponíveis, identificar atores e instituições públicos, monitorar eventos

⁹ Em dezembro de 2017, a instituição informou que a partir de 2018 somente armazenaria publicações com base em um recorte amostral. Disponível em: <https://blogs.loc.gov/loc/2017/12/update-on-the-twitter-archive-at-the-library-of-congress-2>

públicos e construir um conjunto de regramentos robusto com base em conceitos de humanidades digitais.

6. Considerações finais

É fundamental para a análise do 8 de janeiro ter acesso às fontes primárias que documentaram a ação de atores e discursos enunciados nas plataformas digitais por atores de extrema-direita. A ideia é fomentar uma rede de pesquisadores para preservar a história dos atos antidemocráticos e fomentar estudos de interesse público. No entanto, o apagamento sistemático das falas, discursos, imagens e atos são parte do revisionismo histórico que nega o papel de atores para incitar a depredação de símbolos democráticos.

Esse artigo resgatou o debate teórico sobre arquivamento e compartilhamento de dados de plataformas digitais para levantar alguns desafios que vem sendo enfrentados para o estudo de eventos históricos em sites como o Instagram e o Youtube. Em nossa perspectiva, é fundamental considerar aspectos éticos e legais dessas pesquisas aplicadas, sem perder de vista a instrumentalização dessas normativas e dos contratos com as grandes empresas de tecnologia para, efetivamente, inviabilizar a prática arquivista e, no limite, tentar criminalizar a pesquisa acadêmica.

A compreensão desses sistemas sociotécnicos como infraestruturas dinâmicas e em evolução é central para promover atravessamentos entre o debate sobre direito ao esquecimento com perspectivas críticas dos estudos de informação e humanidades digitais. Nesse sentido, o estatuto de removido ou apagado pelos próprios criadores de conteúdo deve ser interpretado de forma crítica como uma medida de mitigação das responsabilidades pelo potencial estímulo da violência política no 08 de janeiro. É dever fundamental da pesquisa acadêmica de interesse público e independente dessas corporações debater as melhores práticas de preservação de coleções digitais que salvaguardam questões éticas, ao mesmo tempo em que resistem ao abuso e à instrumentalização para evitar a sistematização do esquecimento enquanto projeto político.

Referências

- BAYER, Joseph B. et al. Sharing the small moments: ephemeral social interaction on Snapchat. **Information, communication & society**, v. 19, n. 7, p. 956-977, 2016.
- BEN-DAVID, A. Counter-archiving Facebook. **European Journal of Communication**, v. 35, n. 3, p. 249-264, jun. 2020.
- BOYD, Danah. Social network sites as networked publics: Affordances, dynamics, and implications. In: **A networked self**. Routledge, 2010. p. 47-66.
- BRASIL, E.; NASCIMENTO, L. F. HISTÓRIA DIGITAL: REFLEXÕES A PARTIR DA HEMEROTECA DIGITAL BRASILEIRA E DO USO DE CAQDAS NA REELABORAÇÃO DA PESQUISA HISTÓRICA. **Estudos Históricos (Rio de Janeiro)**, v. 33, n. 69, p. 196-219, abr. 2020.
- BRÜGGER, N. **The archived web: doing history in the digital age**. Cambridge, Massachusetts: The MIT Press, 2018.
- BRUNS, A. After the 'APIcalypse': social media platforms and their fight against critical scholarly research. **Information, Communication & Society**, v. 22, n. 11, p. 1544-1566, 19 set. 2019.
- BRUNS, A.; WELLER, K. **Twitter as a first draft of the present: and the challenges of preserving it for the future**. Proceedings of the 8th ACM Conference on Web Science. **Anais... Em: WEBSCI '16: ACM WEB SCIENCE CONFERENCE**. Hannover Germany: ACM, 22 maio 2016. Disponível em: <<https://dl.acm.org/doi/10.1145/2908131.2908174>>. Acesso em: 1 jun. 2024
- CHAN, Chung-hong; SCHATTO-ECKRODT, Tim; GRUBER, Johannes. What makes computational communication science (ir) reproducible?. **Computational Communication Research**, v. 6, n. 1, p. 1, 2024.
- D'ANDRÉA, Carlos. Para além dos dados coletados: Políticas das APIs nas plataformas de mídias digitais. **Matrizes**, v. 15, n. 1, p. 103-122, 2021.
- DE KEULENAAR, E.; BURTON, A. G.; KISJES, I. Deplatforming, demotion and folk theories of Big Tech persecution. **Fronteiras - estudos midiáticos**, v. 23, n. 2, p. 118-139, 14 set. 2021.
- FREELON, D. Computational Research in the Post-API Age. **Political Communication**, v. 35, n. 4, p. 665-668, 2 out. 2018.
- GERLITZ, Carolin; RIEDER, Bernhard. Tweets Are Not Created Equal: Investigating Twitter's Client Ecosystem. **International Journal of Communication (19328036)**, v. 12, 2018.
- GILLESPIE, T. **Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions That Shape Social Media**. [s.l.] Yale University Press, 2019.
- GOMES, D. et al. (EDS.). **The Past Web: Exploring Web Archives**. Cham: Springer International Publishing, 2021.
- HELMOND, A. The Platformization of the Web: Making Web Data Platform Ready. **Social Media + Society**, v. 1, n. 2, p. 205630511560308, 1 jul. 2015.
- LEERSEN, Paddy; HELDT, Amélie P.; KETTEMANN, Matthias C. Scraping By? Europe's law and policy on social media research access. **86272**, v. 12, p. 405-425, 2023.
- LUCCHESI, A. Por um debate sobre História e Historiografia Digital. **Boletim Historiar**, n. 2, 2014.

MANCOSU, Moreno; VEGETTI, Federico. What you can scrape and what is right to scrape: A proposal for a tool to collect public Facebook data. **Social Media+ Society**, v. 6, n. 3, p. 2056305120940703, 2020.

MARRES, N.; WELTEVREDE, E. SCRAPING THE SOCIAL?: Issues in live social research. **Journal of Cultural Economy**, v. 6, n. 3, p. 313–335, ago. 2013.

MAYNARD, D. C. S. Intolerância ao Sul da América: estudo comparado de grupos fascistas do Brasil e da Argentina na Internet (1996-2007). **Revista Tempo e Argumento**, v. 6, n. 12, p. 54–84, 29 set. 2014.

MILLIGAN, I. Mining the ‘Internet Graveyard’: Rethinking the Historians’ Toolkit. **Journal of the Canadian Historical Association**, v. 23, n. 2, p. 21–64, 23 maio 2013.

MILLIGAN, I. Historiography and the Web. **The SAGE handbook of web history**. SAGE Publications Ltd, London, 2018.

PASQUETTO, I. V. et al. Disinformation as Infrastructure: Making and maintaining the QAnon conspiracy on Italian digital media. **Proceedings of the ACM on Human-Computer Interaction**, v. 6, n. CSCW1, p. 1–31, 2022.

PEARCE, W. et al. Visual cross-platform analysis: digital methods to research social media images. **Information, Communication & Society**, v. 23, n. 2, p. 161–180, 28 jan. 2020.

PEHLIVAN, Z.; THIÈVRE, J.; DRUGEON, T. Archiving Social Media: The Case of Twitter. Em: GOMES, D. et al. (Eds.). **The Past Web**. Cham: Springer International Publishing, 2021. p. 43–56.

PERRIAM, Jessamy; BIRKBAK, Andreas; FREEMAN, Andy. Digital methods in a post-API environment. **International Journal of Social Research Methodology**, v. 23, n. 3, p. 277–290, 2020.

ROGERS, R. Digital Methods for Web Research. Em: SCOTT, R. A.; KOSSLYN, S. M. (Eds.). **Emerging Trends in the Social and Behavioral Sciences**. 1. ed. [s.l.] Wiley, 2015. p. 1–22.

ROGERS, R. Deplatforming: Following extreme Internet celebrities to Telegram and alternative social media. **European Journal of Communication**, v. 35, n. 3, p. 213–229, 2020.

ROGERS, R.; BRÜGGER, N.; MILLIGAN, I. Periodizing web archiving: Biographical, event-based, national and autobiographical traditions. **The SAGE handbook of web history**, p. 42–56, 2018.

SANDVIG, C. Heading to the courthouse for Sandvig v. Sessions. 2017. Disponível em: <https://socialmediacollective.org/2017/10/19/heading-to-the-courthouse-for-sandvig-v-sessions/>

TELES, P. O discurso histórico e as suas multiplicidades: uma abordagem sobre as novas demandas da historiografia contemporânea a partir da história do tempo presente. **Boletim Historiar**, n. 3, 2014.

TREZZA, Domenico. To scrape or not to scrape, this is dilemma. The post-API scenario and implications on digital research. **Frontiers in Sociology**, v. 8, p. 1145038, 2023.

TROMBLE, R. Where Have All the Data Gone? A Critical Reflection on Academic Digital Research in the Post-API Age. **Social Media + Society**, v. 7, n. 1, p. 205630512198892, jan. 2021.

VAN DIJCK, José. **Mediated memories in the digital age**. Stanford University Press, 2007.

VAN DIJCK, J.; DE WINKEL, T.; SCHÄFER, M. T. Deplatformization and the governance of the platform ecosystem. **New Media & Society**, p. 14614448211045662, 2021.

WEBSTER, P. Existing web archives. **The SAGE handbook of web history**, p. 30–41, 2018.

ZIMMER, Michael. The Twitter Archive at the Library of Congress: Challenges for information practice and information policy. **First Monday**, 2015.