

Counter-archiving Facebook

European Journal of Communication
2020, Vol. 35(3) 249–264

© The Author(s) 2020



Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/0267323120922069

journals.sagepub.com/home/ejc



Anat Ben-David

The Open University of Israel, Israel

Abstract

The article proposes archival thinking as an analytical framework for studying Facebook. Following recent debates on data colonialism, it argues that Facebook dialectically assumes a role of a new archon of public records, while being unarchivable by design. It then puts forward counter-archiving – a practice developed to resist the epistemic hegemony of colonial archives – as a method that allows the critical study of the social media platform, after it had shut down researcher's access to public data through its application programming interface. After defining and justifying counter-archiving as a method for studying datafied platforms, two counter-archives are presented as proof of concept. The article concludes by discussing the shifting boundaries between the archivist, the activist and the scholar, as the imperative of research methods after datafication.

Keywords

Application programming interface, archive, datafication, Facebook, methods

Introduction

In recent years, data extraction was a popular *modus operandi* in social media research. Platforms' application programming interfaces (APIs) allowed the extraction of large volumes of data, which were then analysed using various computational, quantitative and qualitative methods to answer a broad set of research

Corresponding author:

Anat Ben-David, The Open University of Israel, The Dorothy De Rothschild Campus, One University Road, P.O. Box 808, Ra'anana 4353701, Israel.

Email: anatbd@openu.ac.il

questions on the political, societal, economic and technological aspects of datafication (Rieder, 2013). In 2018, this *modus operandi* had to change, after Facebook – in response to misuse of users' data as part of the Cambridge Analytica scandal – shut down hundreds of thousands of applications that used its API to extract both public and personal data, including the majority of the tools used by the research community. The revoked access had immediate consequences for media and communication scholars, who could no longer conduct independent, ethical and public interest research into the societal effects of social media, such as the formation of online communities, information disorders, political mobilization or discriminatory practices. Without access to data, researchers also cannot conduct independent auditing or monitoring of the role platforms themselves play in shaping social and political processes, such as the use of political advertising around election campaigns. The revoked access to data extraction sparked methodological debates, as researchers were seeking alternatives for studying massively datafied, algorithmic and computational platforms without the convenience and totality afforded by APIs. Among the methods that have been proposed to address what is now called 'post-API' research are collaboration with external companies (Puschmann, 2019), web scraping (albeit violating platforms' terms of use; Bruns, 2019; Freelon, 2018) or returning to the digital fieldwork to think of new methods (Venturini and Rogers, 2019). This article contributes to this debate by suggesting archival theory as a framework for understanding Facebook as contemporary archons, and by introducing counter-archiving as a method of dissent to platforms' appropriation of public data after datafication.

On its face, archival thinking is unfitting for studying algorithmic media. Moreover, declarations on the end of theory which have characterized uncritical views of big data also entailed a death sentence to the archive (Bowker, 2014). In the late 20th century, the humanities and social sciences were inflicted by 'archive fever' (Derrida, 1996), expressed as a renewed interest in archives, archiving and materialities as foundations of post-modern critique. Twenty years later, this 'archive fever' was replaced by an uncritical 'data fever' (Agostinho, 2016). The datafication of everything turns everything into an archive. Why bother with appraisal, description and ordering when the data can answer any question?

The justification for applying archival thinking as a framework for studying Facebook (and other data-driven companies) is that these companies meticulously collect user data at unprecedented scale, thereby forming new forms of commercial archives documenting every aspect of human life (Gehl, 2011). Couldry and Mejias (2019) propose the notion of 'data colonialism' to describe the monopolization of data collection as a new form of capitalism. They draw parallels between Western colonial powers' appropriation of natural resources in past centuries, and the contemporary datafication and commodification of everyday life by digital platforms. Following Couldry and Mejias, this article argues that data colonialism is not only manifested in the datafication of personal and social behaviour, but also in the monopolization of the public record. Although the argument may apply to other social media platforms, the article focuses on Facebook, and draws further

parallels between colonial archives and the social media platform to argue that by negotiating varying levels of access to their data, and by monopolizing their power to discern between private and public records, Facebook dialectically functions as a new 'archon', all the while being unarchivable by design. I subsequently propose counter-archiving as a 'post-API' method for studying Facebook. Counter-archiving has previously been conceived as a form of epistemic resistance that questions colonial archives' hegemonic order, and that calls to understand them as sites of knowledge production, rather than knowledge retrieval (Stoler, 2002). Therefore, in the context of data colonialism, it is proposed to counter-archive Facebook to provide alternatives to the platform's appropriation of public records, and to critique the epistemic affordances of the data it makes available as public.

To justify counter-archiving as a method, I begin by outlining the potential contribution of archival thinking to the study of social media and contextualize the argument on Facebook's appropriation of public records in wider discussions on web archiving. Subsequently, I introduce counter-archiving as a method of dissent that borrows from epistemic responses to colonial archives and define how it may be applied as a post-API method for studying Facebook. After providing examples to counter-archives of Facebook as proof of concept, I conclude by discussing the limits of counter-archives as methods that are agonistic by design.

Facebook as archon

Derrida (1996) traces the origins of the archive in the *Arkeion* of Greek antiquity and refers to it primarily as a space of privilege commanded by *archons* – superior magistrates acting as guardians of documents. These archives gained their ability to represent the law by being situated at the intersection between private and public spaces: the documents were signed and stored in the private households of the archons, on account of the public recognition of their authority. In providing a post-colonial critique of the concept of the archive, Ariela Azoulay (2011) argues that Derrida's emphasis on archive-as-place downplays the archon's power. For Azoulay (2011), the archon's role is not only realized as the guardian of documents in his domicile, but also as the one in charge of 'distancing those wishing to enter the archive too early, before the materials stored within would become history, dead matter, the past' (n.p.).

Both Derrida's etymology of the archive as a public/private space, and Azoulay's understanding of the archon's role in distancing citizens from information that may be of political relevance in real time, are useful frameworks for understanding Facebook as self-appointed archon in the context of data colonialism and post-API research. In this section, I attempt to situate the company's appropriation of the notion of the public archive in the context of the history of web archiving, and in the company's recent attempts to brand temporary access to data sets of political advertisements as archives and libraries of transparency.

To web archivists and internet historians, post-API debates are not new. After nearly two decades of archiving large proportions of the open web, these practitioners and scholars were early to notice that social media platforms, and Facebook in particular, are unarchivable. Web archiving differs from data extraction in the sense that it is less concerned with how the data will be used now, and more with how to preserve data for access and use in the future (Brügger, 2012). Digital preservation experts argue that the digital cultural heritage of our times is at risk since digital media are prone to decay and deletion (UNESCO, 2003). Web archiving methods were developed to fight web decay, operating under the premise that the web constitutes an important part of humanity's public record, and that there is eminent need in its preservation for posterity (Brügger and Milligan, 2018). From as early as 1996, the internet archive and national libraries around the world have been preserving petabytes of archived websites, and continue to do so on a daily basis (Costa et al., 2017). A growing community of researchers depends on web archives as scarce and reliable born-digital primary sources that support historical Internet research, for without web archives, it would be nearly impossible to find online evidence of the web's past (Ben-David, 2016).

However, both the logic of web archiving and the ability to apply historical thinking in web research collapsed when the majority of the web's content migrated to commercial social media platforms. Due to the conditions specified in platforms' terms of use, social media data are no longer in the public domain, and while API access allows data extraction to some extent, archiving Facebook is legally impossible. To address the unarchivability of social media, web archivists began seeking 'post-API' workarounds years before Facebook's API lockout. The solutions that have been proposed resemble the methodological solutions to post-API research described above and include attempts to reach collaborative agreement between the platform and the cultural heritage institution, the use of third-party services, and crowd-sourcing (Hockx-Yu, 2014). Most of these solutions have registered partial success. Most notorious of them is the Twitter archive at the Library of Congress. In 2010, the library had reached an agreement with the social media company to archive every public tweet posted since 2006. The collected tweets were meant to be made available for viewing after a 2-year embargo. After several years of data collection, the initiative did not bear fruit, primarily since the library was unable to find solutions to the copyright and privacy challenges involved in republishing the data (Zimmer, 2015). Other creative examples include the national library of New Zealand's initiative to create a crowdsourced 'time capsule' of Facebook by asking citizens to donate their data (Deguara, 2019), and the Internet Archive's use of the fictive Facebook account 'Charlie Archivist' to archive logged-in Facebook pages of public figures. This account has zero friends, thereby ensuring that users' private data will not be compromised during the capture. Nevertheless, since current web archiving crawlers cannot fully capture the dynamic content of social media, the eventual capture of the logged-in pages is rather incomplete (see Figure 1).

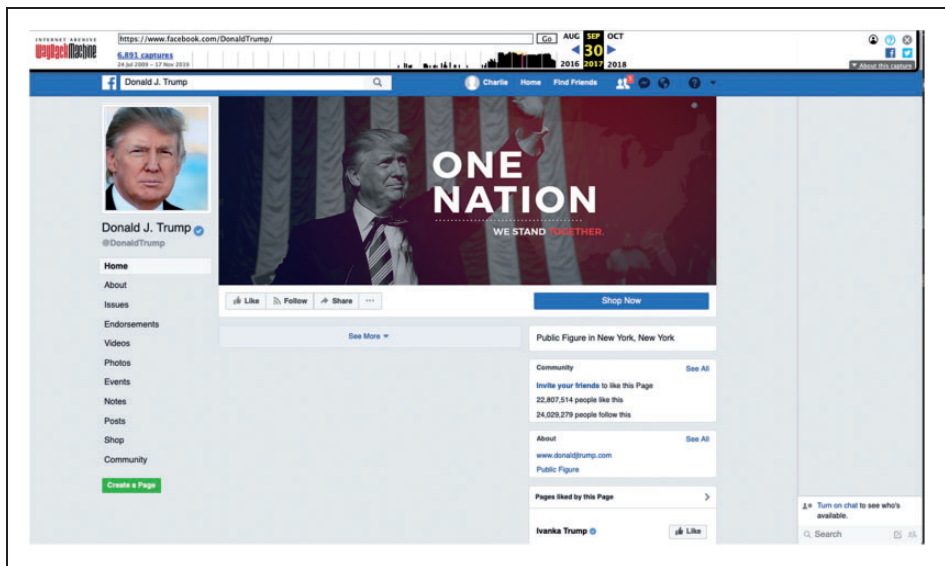


Figure 1. An archived snapshot taken from Donald Trump's Facebook page, dated 30 September 2017, captured by the Internet Archive as a logged-in page, using the fictive user account 'Charlie Archivist'.

Sources: <https://web.archive.org/web/20170930232640> and <https://www.facebook.com/DonaldTrump/>

Such creative 'workarounds' reflect institutional archives' attempts to reclaim their role as archons, in light of growing commercialization of data that hitherto was considered public. But as cultural heritage institutions were losing grip on access to public data, Facebook started lending access to new types of archives. Parallel to the API lockout in 2018, Facebook launched what it termed the 'Ad Archive API', a search interface and API to access a collection of political advertisements in the United States, and promoted it as a transparency tool that would help research manipulation (Facebook Newsroom, 2018). Researchers were quick to note, however, that these collections are 'heavily edited and appraised for reputation management' (Acker and Donovan, 2019, p. 1597), and they exclude information about the targeting categories used to reach individual Facebook users. A few months later, the Ad Archive was rebranded as the 'Ad Library' (Facebook.com, n.d.). More countries were added to the collection, and access to ads was limited to 7 years. Parallel to the launch of the Ad Archive/Library, Facebook quietly blocked browser add-ons developed by civic initiatives such as the American news organization Propublica and the British NGO 'Who Targets Me' (Merrill and Tobin, 2019). Facebook users who installed the add-ons gave permission to these initiatives to automatically and anonymously collect the political ads they were being served on the platform, along with the targeting

information attached to each ad's 'why am I seeing this ad' feature. The collected data were then made available to the public through a search interface (Waterson, 2019). Thus, while Facebook brands itself as a benevolent archon that lends immediate access to contemporary data in the name of transparency, it also keeps away citizens from accessing information that may cause political scandal (i.e. excluding targeting information, shutting down researchers' access to its API), sets a time limit to the availability of records and ensures its monopoly on record-keeping.

If we are to accept that Facebook functions as one of the archons of data colonialism, then post-API research methods may also borrow from methods of resistance to colonial powers, such as 'counter-mapping' and 'counter-archiving'. The epistemic premise behind these methods affirms the power of colonial instruments, such as maps, archives and museums in shaping knowledge, subjects, nations, and geopolitical and racial boundaries according to colonial interests (Anderson, 1983), yet uses the same techniques to uncover injustice, reclaim rights or to propose epistemic alternatives to such hegemonic structures (Peluso, 1995). Ann Stoler's (2018) work on archiving as dissensus is a case in point. In imagining what a Palestinian archive would/ought to be, Stoler (2018) argues that

at issue is an archival assembly that is not constrained by the command – in form and content – that is dictated by colonial state priorities or even by Palestinian authorities. It is rather one that is authored and authorized by a constituent as yet unspecified Palestinian public. (p. 43)

Stoler (2018) further conceptualizes the counter-archive as anticipating possible uses and possible connections, and as an invitation to make it possible to 'actualize connectivities that are dimly visible or on the horizon' (p. 46). Also writing in the context of Palestinian archiving, Ariela Azoulay argues that 'archive fever' – Derrida's notion of the obsession with archives and archiving described above – is itself a method of dissent. For her, 'archive fever' is

partaking in the practice of the archive through founding archives of new sorts, such that do not enable the dominant type of archive, founded by the State, to go on determining what the archive is. Archive fever challenges traditional protocol by which official archives have functioned and continue to do so. It proposes new models of sharing the documents stored therein in ways that requires one to think the public's right to the archive not as external to the archive but rather as an essential part of it, of its character, of its *raison d'être*. (Azoulay, 2011, n.p.)

Following Stoler and Azoulay, I propose building archives of Facebook that are designed to counter the platform's protocols of access to knowledge, that allow anticipating possible invisible connections, and that question the public's right to the social media archive.

Counter-archiving Facebook is a call for action in as much as it is a methodological solution to platform lockout. It blurs boundaries between archiving as

action and a scholarly method; between the archive as (re)source and object of study; and between the researcher as archivist, scholar and activist. Although these blurred boundaries are intentional, they require justification. In the following section of this article, I attempt to make the case for counter-archiving Facebook as a post-API method.

Counter-archiving as a method

How does one distinguish between archiving as a profession, an activist counter-practice and a scholarly method? According to the Society of American Archivists (n.d.), an archivist is an individual responsible for ‘appraising, acquiring, arranging, describing, preserving, and providing access to records of enduring value, according to the principles of provenance, original order, and collective control to protect the materials’ authenticity and context’ [...] – and for ‘management and oversight of an archival repository or of records of enduring value’. Researchers are certainly not archivists. Their scholarly work may involve consulting archives or engaging in source critique, but they are not expected to acquire, appraise, order, describe and lend access to data. Why, then, consider counter-archiving Facebook as a scholarly method?

Arguably, API access to social media data has conflated the notion of data extraction and collection-making, with archiving and preservation. API-based research offers methodological comfort in providing researchers with structured data, a clear demarcation of the types of data that can be extracted, their volume and the restrictions on their use, which are specified in the platform’s terms of use (Venturini and Rogers, 2019). Although previous research proposed using APIs for archiving social media (Acker and Kreisberg, 2019; Littman et al., 2018; Lomborg, 2012), the majority of social media researchers have used API data for immediate analysis, rather than for long-term preservation, appraisal of sources or lending access to others.

Critics of API-based research have pointed at the risks of making truth claims using social media data without taking into account that instead of mediating social or political phenomena, these data are originally created to meet specific corporate goals and ideologies (John and Nissenbaum, 2019; Marres and Gerlitz, 2016). The further lockout of researchers’ access to API data therefore necessitates thinking about new ways for data critique.

One such way is expressed in Crawford’s (2016) use of Chantal Mouffe’s notion of ‘agonistic pluralism’ as both a design ideal and a provocation for studying algorithms in broad social contexts. The political theory of agonism acknowledges the importance of conflict to politics. Unlike antagonism, in which one side aims to win over the other, agonism respects the adversary’s right to existence, yet the sides remain in perpetual disagreement. In Crawford’s work, agonism is brought to study the contestations that shape public discourse and the logic of calculated publics in various algorithmic platforms. Similarly, the call to counter-archive Facebook is agonistic by design, in the sense that it affirms the perpetual

contestation of Facebook-as-archive by its epistemic alternatives. Therefore, the method of counter-archiving Facebook explicitly proposes to collect, reorganize and republish Facebook data to go against Facebook's archival order. It differs from other methods of data mining and public data sharing or dumping, in the sense that it does not simply make data sets available (Weller and Kinder-Kurlanda, 2016) but rather consciously incorporates appraisal, facilitates use and is transparent about its definition of 'publicness' and provenance. I also distinguish it from other critical social media research methods, such as scraping, monitoring, repurposing and tinkering (Marres and Weltevrede, 2013; Venturini et al., 2018). Counter-archiving may borrow from all of the above, yet instead of immediate use, its analytical value is in reclaiming the parts of social data that may be regarded as public, which Facebook-as-archon decided to keep away from public scrutiny. For this reason, contrary to digital methods that unobtrusively follow the medium (Rogers, 2013), counter-archiving is obtrusive as it remediates and republishes public Facebook data in ways that extend their epistemic capacities and reveal more than the platform had intended.

Counter-archiving Facebook is not strictly a post-API method, as theoretically (and as shown below), it is possible to build counter-archives using API data. However, the implication of collection-making post-API access is that prospective counter-archives will neither be structured nor exhaustive. Data collection might become cumbersome and manual. There will be errors. These counter-archives will inevitably be incomplete; they will lack researchers' ability to ground truth claims on the scale, or representativeness of what they had collected. Counter-archives of Facebook are also demanding as they are not designed to meet a pre-defined analytical use. As Stoler (2018) noted,

It matters less what we do than how we do it. For in the end we task ourselves to thicken the present with such alternatives. It is those who contribute to this archive in the making who have it in their collective hands to forge an archive not of the past but of the vibrant present studded with possibilities for the future. (p. 55)

Hence the analytical value of counter-archives of Facebook could be found less in their content and more in the epistemic alternatives that they propose: the possibility to ask questions other than what the platform had intended, the ability to imagine alternative histories and alternative analytical possibilities that could perhaps do better service to the mediation of public debates and public facts. The counter-archives of Facebook should be built ethically and responsibly. The republishing of Facebook data should only focus on materials that are of public interest (e.g. publicly funded Facebook pages of politicians, sponsored ads of medical institutions or the official page of the Facebook app) and take strict measures to exclude information that can be regarded as private. Instead, the boundaries and mission statements of each collection should be demarcated, transparent and justified. As public-facing collections, responsible counter-archives of Facebook should be lawful, and should not violate the platform's terms of use.

Proof of concept to counter-archives

To illustrate the manifold shapes and forms that counter-archives of Facebook could wear, this section outlines two examples of public archives of Facebook data that I created before and after the platform’s API lockout. The first example is *Polibook* (Polibook.online, n.d.), a public archive of the Israeli parliament on Facebook in the years 2015 to 2019, demarcated as a longitudinal issue space. The second is *Meturgatim* (Hebrew for ‘Targeted’), a crowd-sourced archive of screenshots of political ads, collected during the two rounds of the Israeli general election in 2019.

Polibook is an archive containing every Facebook post of every Israeli parliament member during the 20th Knesset (15 March 2015–15 March 2019). During these years, data were extracted daily using Facebook’s Graph API ($N = 71,250$). New posts were added to an accumulative index. The archive’s public search interface uses a term frequency–inverse document frequency (TF–IDF) ranking algorithm (Ramos, 2003) to assign a relative and comparative ‘relevance’ score to each politician. That is, instead of calculating frequencies, the algorithm weighs the keywords’ relevance in relation both to the politician’s past posts and comparing them with all other politicians’ posts. After the user selects the time period of interest and the search query, the results are visualized as a star-hub network, displaying politicians with higher relevance scores closer to the search term than others (see Figure 2). Coalition and opposition members are marked in blue and green, respectively, and upon clicking on their profile images, the posts’ texts are displayed, along with their relevance scores.

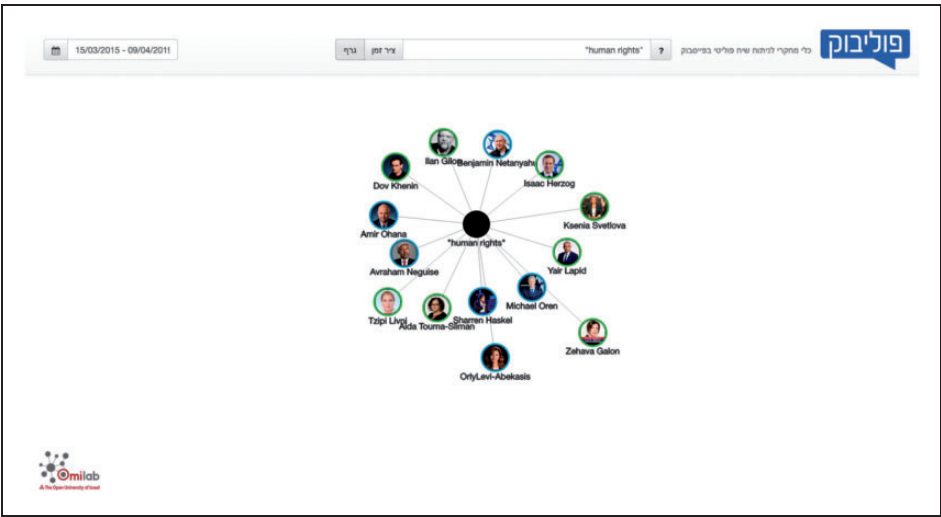


Figure 2. A screenshot from Polibook.online depicting its relational ranking of search terms.

In putting the Facebook pages of parliament members in a public issue space populated by actors from all political camps, Polibook breaks away from Facebook’s constraints on personalized newsfeed algorithms, personalized search and from the API’s affordances that limit data extraction to individual pages. De-personalized search allows imagining new connections that Facebook’s affordances do not make visible: it affords comparative and longitudinal analyses of the framing of political debates and the formation of coalitions of actors around specific issues. For example, Polibook’s data show that with the exception of three men, only women discuss gender-related issues, such as sexual harassment or parenting leave. Politicians from both left and right refrain from using the term ‘Occupation’ in the context of Palestinians, and right-wing politicians consistently refer to asylum seekers as ‘infiltrators’ (see Figure 3). As an archive, it is perhaps the most complete public evidence of the role Facebook had played in mediating Israeli politics in the studied years. Yet Polibook’s completeness was derived from access to Facebook’s Graph API, which was discontinued in June 2019.

The second example is *Meturgatim* (meaning ‘targeted’ in Hebrew), a screenshots archive of political ads during the two 2019 election rounds in Israel. This archive is a response to the platform’s framing of transparency

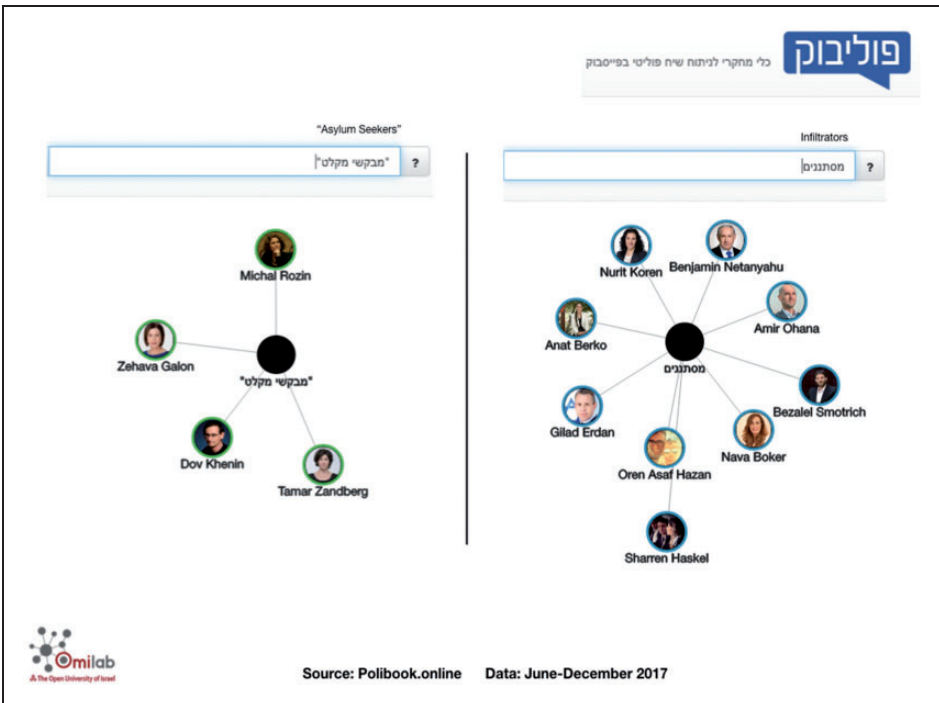


Figure 3. Screenshots from Polibook.online displaying query results for ‘infiltrators’ (right) and ‘asylum seekers’ (left).

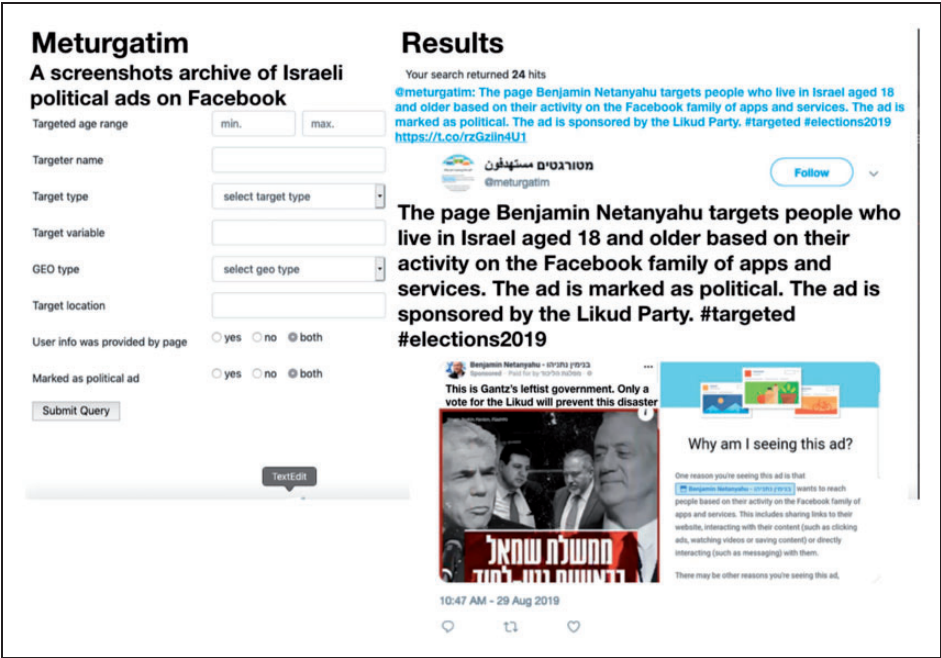
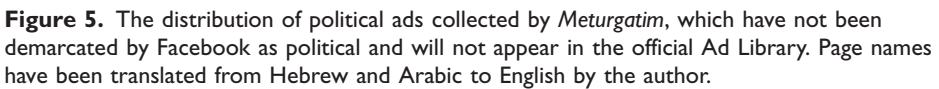


Figure 4. The interface of Meturgatim’s screenshot archive, displaying the first result of the query ‘Ads by Benjamin Netanyahu that target users based on their activity on the Facebook family of apps and services’. Displayed text has been translated from Hebrew to English by the author.

with regards to political advertising during election campaigns. Ahead of the Israeli election in 2019, Facebook announced the launch of the local version of its ad library in the country, and framed it as a transparency tool (Surkes, 2019). However, the library only displays aggregated demographics and estimate costs of the ads’ reach and does not include any information about the categories used for targeting audiences or users. Since access to automated data extraction is prohibited, we decided to build a crowdsourced archive of screenshots of political ads sent by anonymous users. Compared with other methods of data extraction, the rather ‘primitive’ screenshot offers limited analytical possibilities and is difficult to collect. However, the archive is centred around the screenshot due to its agonistic qualities as an object of ‘photographic witnessing’, and as a digital gesture of cultural resistance (Frosh, 2018). The screenshot represents users’ agency to publicly (and legally) share data that are only made visible to them individually, in ways that Facebook cannot yet control. In this case, data are manually extracted from Facebook as images, but the information the images bear is then re-transformed into searchable, textual data.



In February 2019, 2 months before the first round of the Israeli elections, we issued public calls through various news outlets and asked Facebook users to send us screenshots of political ads they were served, along with a second screenshot of the ‘why am I seeing this ad?’ feature that each user is served individually. The screenshots were then transformed into a searchable archive using a two-step process: (1) De-platformization: images were manually cropped to remove any user data. Each pair of screenshots was tweeted in real time by the project’s account (@meturgatim), along with a verbal transcription of the content of the ‘why am I seeing this ad’ screenshot. The republishing of the screenshots taken from

Facebook on another social media platform was done to make them immediately available for public scrutiny, as well as to preserve them outside of Facebook. (2) Re-datafication: each tweet was subsequently automatically indexed into a search engine, which allows retrieval of the ads based on the transcribed targeting information (see Figure 4).

Overall, we collected over 3000 political ads contributed by over 100 users. The data are by all means partial, non-representative and hence of limited quantitative analytical value. However, in providing additional evidence that cannot be inferred from the official Ad Library, this archive may serve various analytical purposes, such as studying the imagined audiences and campaign strategies of political advertisers, the types of advertising categories that Facebook makes available for political use, or inferring why certain issue ads are marked as political, while others are not. For example, initial analysis of ads served ahead of the September election shows that over 35% of the collected ads were not marked by the platform as political. These unmarked ads were targeted primarily by politicians, anonymous pages and NGOs. Ads ran by pages whose administrators are anonymous are of special importance for studying manipulation or disinformation within campaigns. However, since they are not marked as political, they will not appear in Facebook's official Ad Library, therefore, doubting the very utility of making truth claims on the completeness of Facebook's official collection (see Figure 5). From a historical perspective, this counter-archive aims to outlast the 7 years allotted by Facebook, while providing evidence of the evolution of Facebook's transparency features (which have already changed immediately after the election). The screenshots archive provides documentation of what had previously been afforded, in ways which would not be otherwise accessible for comprehensive research.

Conclusion

In the previous sections, I contextualized Facebook's unarchivability on one hand, and its tight control on the types of data it makes public on the other, in wider discussions on the critical importance of web archiving for historical research. After datafication, access to information that until recently was in the public domain and worthy of preservation by memory institutions, now greatly depends on platforms' benevolence. Following Couldry and Mejias' (2019) notion of data colonialism, I argued that platform's control over access to public data resembles the power of colonial archives' ability to keep citizens away from information until its contents become the realm of the depoliticized past. Since memory institutions are unable to reclaim their role as archons of public digital media after datafication, it is the role of social media researchers, who are already committed to research ethics and to respecting users' privacy, to cross disciplinary boundaries and engage in archival work through building counter-archives and making them available for public scrutiny. For without such scholarly intervention, the implication of platform lockout might be the de-historization (and subsequently, de-

politization) of social media research, and a growing dependence on Facebook's self-assumed role as the contemporary archon of public data.

Critics of the approach may argue that instead of a method, what is being proposed is a form of civic engagement and political action (Milan and van der Velden, 2016). Indeed, the question of legitimacy is raised when establishing collections that are not only intended for individual use. Compared with API-based research, the need to build collections that are public-facing puts additional burden on researchers' shoulders. Since institutional archives cannot build agonistic archives, individual researchers are now burdened with both the archivist's and the activist's tasks. It is the role of scholars to not only take responsibility for the ethics of data extraction and sharing, but also for the accessibility, transparency and sustainability of their collections. Researchers might be deterred from the legal responsibilities involving the republishing of Facebook data, and from assuming the archivists' duties to describe, order and lend long-term access to their archives. Future research infrastructure for maintaining and sharing scholarly archives of social media may take this burden off individual researchers' shoulders, for example, by developing new standards for indexing, describing and licensing access to counter-archiving projects.

Funding

The author(s) received no financial support for the research, authorship and/or publication of this article.

References

- Acker A and Donovan J (2019) Data craft: A theory/methods package for critical internet studies. *Information, Communication & Society* 22(11): 1590–1609.
- Acker A and Kreisberg A (2019) Social media data archives in an API-driven world. *Archival Science*. Epub ahead of print 24 September. Available at: <https://doi.org/10.1007/s10502-019-09325-9> (accessed 20 February 2020).
- Agostinho D (2016) Big data, time and the archive. *Symploke* 24(1–2): 435–445.
- Anderson B (1983) *Imagined Communities: Reflections on the Origin and Spread of Nationalism*. London: Verso.
- Azoulay A (2011) *Archive. Political Concepts: A Critical Lexicon* 1(5). Available at: www.politicalconcepts.org/issue1/archive/ (accessed 20 February 2020).
- Ben-David A (2016) What does the web remember of its deleted past? An archival reconstruction of the top-level Yugoslav web domain. *New Media & Society* 18(7): 1103–1119.
- Bowker GC (2014) Big data, big questions: The theory/data thing. *International Journal of Communication* 8(5). Available at: <https://ijoc.org/index.php/ijoc/article/view/2190/1156> (accessed 20 February 2020).
- Brügger N (2012) When the present web is later the past: Web historiography, digital history and Internet studies. *Historical Social Research* 37(4): 102–117.
- Brügger N and Milligan I (2018) *The SAGE Handbook of Web History*. London: SAGE.
- Bruns A (2019) After the 'APIcalypse': Social media platforms and their fight against critical scholarly research. *Information, Communication & Society* 22(11): 1544–1566.

- Costa M, Gomes D and Silva MJ (2017) The evolution of web archiving. *International Journal on Digital Libraries* 18(3): 191–205.
- Couldry N and Mejias UA (2019) Data colonialism: Rethinking big data's relation to the contemporary subject. *Television & New Media* 20(4): 336–349.
- Crawford K (2016) Can an algorithm be agonistic? Ten scenes from life in calculated publics. *Science, Technology, & Human Values* 41(1): 77–92.
- Deguara B (2019) National Library creates Facebook time capsule to document New Zealand's history. *Stuff*, 5 September. Available at: <https://www.stuff.co.nz/national/115494638/national-library-creates-facebook-time-capsule-to-document-new-zealands-history> (accessed 20 February 2020).
- Derrida J (1996) *Archive Fever: A Freudian Impression*. Chicago, IL: University of Chicago Press.
- Facebook Newsroom (2018) Introducing the ad archive API, August. Available at: <https://newsroom.fb.com/news/2018/08/introducing-the-ad-archive-api/> (accessed 20 February 2020).
- Facebook.com (n.d.) Ad Library. <https://www.facebook.com/ads/library> (accessed 20 February 2020).
- Freelon D (2018) Computational research in the Post-API Age. *Political Communication* 35(4): 665–668.
- Frosh P (2018) *The Poetics of Digital Media*. Cambridge: Polity Press.
- Gehl RW (2011) The archive and the processor: The internal logic of Web 2.0. *New Media & Society* 13(8): 1228–1244.
- Hockx-Yu H (2014) Archiving social media in the context of non-print Legal Deposit. In: *The International Federation of Library Associations and Institutions World Library and Information Congress*, Lyon, 16–22 August. Available at: <http://library.ifla.org/id/eprint/999> (accessed 20 February 2020).
- John NA and Nissenbaum A (2019) An agnotological analysis of APIs: Or, disconnectivity and the ideological limits of our knowledge of social media. *The Information Society* 35(1): 1–12.
- Littman J, Chudnov D, Kerchner D, et al. (2018) API-based social media collecting as a form of web archiving. *International Journal on Digital Libraries* 19(1): 21–38.
- Lomborg S (2012) Researching communicative practice: Web archiving in qualitative social media research. *Journal of Technology in Human Services* 30(3–4): 219–231.
- Marres N and Gerlitz C (2016) Interface methods: Renegotiating relations between digital social research, STS and sociology. *The Sociological Review* 64(1): 21–46.
- Marres N and Weltevrede E (2013) Scraping the social? Issues in live social research. *Journal of Cultural Economy* 6(3): 313–335.
- Merrill JB and Tobin A (2019) Facebook quietly blocked tools that let people see how its ads are targeted. *Quartz*, 30 January. Available at: <https://qz.com/1537686/facebook-blocks-propublica-and-mozillas-ad-transparency-tools/> (accessed 20 February 2020).
- Milan S and van der Velden L (2016) The alternative epistemologies of data activism. *Digital Culture & Society* 2(2): 57–74.
- Peluso NL (1995) Whose woods are these? Counter-mapping forest territories in Kalimantan, Indonesia. *Antipode* 27(4): 383–406.
- Polibook.online (n.d.) A research tool for studying political discourse on Facebook. Available at: <http://www.polibook.online/> (accessed 20 February 2020).
- Puschmann C (2019) An end to the wild west of social media research: A response to Axel Bruns. *Information, Communication & Society* 22(11): 1582–1589.

- Ramos J (2003) Using TF-IDF to determine word relevance in document queries. In: *Proceedings of the first instructional conference on machine learning*, vol. 242, Piscataway, NJ, 3–8 December, pp. 133–142. Available at: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.121.1424&rep=rep1&type=pdf>.
- Rieder B (2013) Studying Facebook via data extraction: The Netvizz application. In: Davis H (ed.) *Proceedings of the 5th Annual ACM Web Science Conference*. New York: ACM, pp. 346–355.
- Rogers R (2013) *Digital Methods*. Cambridge, MA: The MIT Press.
- Society of American Archivists (n.d.) A glossary of archival and records terminology: Archivist. Available at: <https://www2.archivists.org/glossary/terms/a/archivist> (accessed 20 February 2020).
- Stoler AL (2002) Colonial archives and the arts of governance. *Archival Science* 2(1–2): 87–109.
- Stoler AL (2018) On archiving as dissensus. *Comparative Studies of South Asia, Africa and the Middle East* 38(1): 43–56.
- Surkes S (2019) Facebook rolls out political ad transparency less than a month before polls. *The Times of Israel* 14 March. Available at: <https://www.timesofisrael.com/facebook-rolls-out-political-ad-transparency-less-than-a-month-before-polls/> (accessed 20 February 2020).
- UNESCO (2003) *Charter on the Preservation of Digital Heritage. Adopted by General Conference on 15 October 2003*. Paris: UNESCO.
- Venturini T and Rogers R (2019) ‘API-Based Research’ or how can digital sociology and journalism studies learn from the Facebook and Cambridge Analytica data breach. *Digital Journalism* 7(4): 532–540.
- Venturini T, Bounegru L, Gray J, et al. (2018) A reality check (list) for digital methods. *New Media & Society* 20(11): 4195–4217.
- Waterson J (2019) Facebook restricts campaigners’ ability to check ads for political transparency. *The Guardian*, 27 January. Available at: <https://www.theguardian.com/technology/2019/jan/27/facebook-restricts-campaigners-ability-to-check-ads-for-political-transparency> (accessed 20 February 2020).
- Weller K and Kinder-Kurlanda KE (2016) A manifesto for data sharing in social media research. In: Nejdl W (ed.) *Proceedings of the 8th ACM Conference on Web Science*. New York: ACM, pp. 166–172.
- Zimmer M (2015) The Twitter archive at the Library of Congress: Challenges for information practice and information policy. *First Monday* 20(7). Available at: <https://firstmonday.org/article/view/5619/4653> (accessed 20 February 2020).