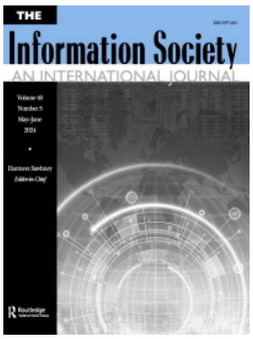

Routledge
Taylor & Francis Group

The Information Society
An International Journal

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/utis20

# Platforms as templates: Emerging datafication dynamics in digital news outlets' datawalls

Rohan Grover & Jeeyun (Sophia) Baik

**To cite this article:** Rohan Grover & Jeeyun (Sophia) Baik (31 May 2024): Platforms as templates: Emerging datafication dynamics in digital news outlets' datawalls, The Information Society, DOI: 10.1080/01972243.2024.2350015

**To link to this article:** https://doi.org/10.1080/01972243.2024.2350015

THE INFORMATION SOCIETY
https://doi.org/10.1080/01972243.2024.2350015

Routledge
Taylor & Francis Group



# Platforms as templates: Emerging datafication dynamics in digital news outlets' datawalls

Rohan Grover[a] and Jeeyun (Sophia) Baik[b]

[a]Annenberg School for Communication and Journalism, University of Southern California, Los Angeles, California, USA; [b]Department of Communication, University of San Diego, San Diego, California, USA

**ABSTRACT**

News outlets have responded to platforms disrupting audience engagement and revenue models by pursuing more stable revenue through paywall subscriptions and membership programs. However, many outlets continue to struggle, and scholars have suggested that outlets experiment with data monetization strategies by asking readers to "pay" with their data – in other words, by implementing a "datawall." We conducted an exploratory landscape analysis of digital news websites and interviewed digital news professionals to evaluate the extent to which datawalls actually exist in the field and to investigate the roles of platformization and datafication. We found that datawalls exist on a spectrum with a complicated relationship to paywalls and that platforms play an important role in the emergence of datawalls that calls for extending the existing conceptual model of platformization. We argue that datawalls represent attempts to *adopt platform logics* by accumulating audience data; thus, platforms serve as *templates* for digital news outlets. We discuss the implications of these findings for conceptualizing platformization and datafication.

**ARTICLE HISTORY**
Received 19 January 2023
Accepted 27 April 2024

**KEYWORDS**
Datafication; datawall; digital news; paywall; platformization

## Introduction

> "Third-party cookies – we knew that that was going away. So having first-party data became more attractive than the paid revenue model … We became less focused on revenue and became a little bit more focused on getting data from people. Because once we have that first-party data – you've created a login, we have your identity – we can track you and follow you and target you and all that stuff. That also became really appealing."

The above quote may appear like a logical response from an employee at one of the major platform companies, which are known for harvesting and monetizing personal data. However, it is in fact an excerpt from our interview with a former *HuffPost* employee. It expresses a common anxiety about the implications of the impending loss of third-party cookies for digital news organizations and, in particular, for their dependence on Big Tech platforms. It alludes to the perpetual uncertainty caused by long-standing processes of datafication and platformization in the digital news business.

Media scholars have conceptualized that one way news organizations could respond to these trends is by imposing a *datawall*: a mechanism to limit reader access in exchange for collecting users' data. A datawall is analogous to a paywall but collects data instead of financial payment. While datawalls have been considered and debated conceptually, they have not been examined empirically.

Therefore, we ask: to what extent do datawalls actually exist in the field as a novel strategic implementation distinct from paywalls? We address this question by conducting a landscape analysis of major digital news outlets popular among US audiences and developing a typology that captures the diversity of access mechanisms employed in the field. We argue that datawalls exist on a spectrum, from payment collection to data collection, with ambiguous hybrid models in between, and provide evidence of evolving data monetization and platform pressures.

We also evaluate the platformization framework's explanatory power to understand how and why digital news outlets have turned to datawalls. We ask: what, if any, is the role of platformization and datafication

**CONTACT** Rohan Grover rohan.grover@usc.edu Rohan Annenberg School for Communication and Journalism, University of Southern California, 3502 Watt Way, Los Angeles, CA 90089, USA.

in the emergence of datawalls? We address this question through interviews at digital news outlets and a case study of *HuffPost*. We argue that platforms play an important role in the emergence of datawalls in a way that exceeds *platformization*, defined as "the penetration of the infrastructures, economic processes, and governmental frameworks of platforms in different economic sectors and spheres of life" (Poell, Nieborg, and Van Dijck 2019, 2; see also Nieborg and Poell 2018). We suggest that datawalls may represent digital news organizations' attempts to *adopt platform logics* themselves rather than the outcome of platforms penetration in digital news. In other words, datawalls could be a form of platformization in which platforms serve not as direct infrastructure or architecture, but rather as *templates* for other sectors.

Ultimately, our article contributes to scholarship on digital journalism, platforms, and datafication by examining their intersection *via* an empirical study of datawalls. Our findings contribute analytic clarity about emerging platformization and datafication processes in digital news, and thus open up new questions about similar dynamics of *platforms as templates* in other sectors.

## Literature review

### The rise of paywalls

Previous studies have documented how platform companies such as Google and Facebook have disrupted the digital news industry primarily by reconfiguring audience distribution models. As platforms grew increasingly popular and as their platform infrastructure centralized the web (Helmond 2015), publishers became reliant upon their algorithms and digital advertising marketplaces to attain access to readers. While advertisers and economic logics have always figured in the evolution of news organizations (Napoli 2011), the relationship between platforms and publishers is unique, thereby receiving much scrutiny. Researchers have found it to be a relationship of "control" (Simon 2022; Walters 2021), "reengineering" (Bell and Owen 2017), and "dependency" (Nielsen and Ganter 2018). The stakes of this asymmetric relationship between platforms and publishers are high for the democratic role of digital news organizations, affecting their editorial identity, access to data, and revenue (Nielsen and Ganter 2018).

One increasingly common strategy that news organizations have pursued to reclaim independence and sustainability is to solicit direct reader revenue by imposing *paywalls* on content (Olsen, Kalsnes, and Barland 2021). Under this model, readers must pay in order to access digital news content. Paywalls can be distinguished between *hard* paywalls, which require a subscription to access all content, and *soft* paywalls, which permit access to some content for free and other content only with a subscription (Myllylahti 2014). Soft paywalls can be further distinguished between *metered* paywalls, which permit access to a certain number of articles for free before paying, and a *freemium* model, which distinguishes between free content and premium content that requires a subscription (Myllylahti 2014). Notably, paywalls often employ a hybrid strategy by pursuing revenue both from subscriptions paid by loyal readers and also from digital advertising simultaneously (Meese and Hurcombe 2021).[1] However, paywalls have not been universally successful. In the US, national newspapers such as the *New York Times* and the *Wall Street Journal* have enjoyed significant growth in reader revenue whereas local news outlets' audiences are often unlikely to pay (Chyi and Ng 2020).

### From paywalls to datawalls?

Recently, scholars have hypothesized about the potential rise of another strategy: the *datawall*. Evens and Van Damme (2016) examined newspaper users' willingness to share personal data with news organizations to access content and/or other experiences through a case study of Media ID – a federated identity management system supported by all the main newspaper outlets in Flanders, Belgium – where users can "register, free of charge" on any of the newspaper websites through a "single sign-in" identification (27). Based on their findings, they defined the datawall as a business model where news organizations require "readers to share personal data" to secure "access to value-added, personalized news services" (27). They described how datawalls are an attempt to reclaim independence and control over audiences, enhance targeted advertising, and improve the user experience. In another example, Bechmann, Bilgrav-Nielsen, and Korsgaard Jensen (2016) identified precursors of the datawall, including interfaces such as a "sharewall" or "permission wall" where users gain access to news in exchange for sharing news content with their social networks.

The datawall concept is compelling because of recent anxieties about data collection limitations with the phaseout of third-party cookies – which some have called the "Cookiepocalypse" (El Hana, Mercanti-Guérin, and Sabri-Zaaraoui 2022). Third-party cookies are created by advertisers and other companies to track users across websites, and

they are often used to deliver targeted or personalized advertisements. In contrast, first-party cookies are created by – and only accessible to – publishers only on their own host domains. Since third-party cookies collect data for companies whose websites have never been visited by the user, they are seen by some as violations of user privacy. In 2019, when Google Chrome – the world's most popular browser – was considering blocking third-party cookies, publishers and advertisers alike began to fear that they would lose revenue (Shields 2019). In this context, datawalls, in theory, could help websites uphold their data collection practices without third-party cookies because publishers could collect first-party data if users create accounts directly with the publisher. Specifically, datawalls would allow publishers to collect user-inputted data through registration flows and follow-up surveys, collect additional session and engagement data with first-party cookies, and defend certain data collection practices as "strictly necessary" if they are essential to operating the datawall. Indeed, a trade association for publishers recently released a playbook for publishers and instructed them to compensate for the loss of third-party cookies by driving newsletter signups, assigning unique identifiers for users, and collecting first-party data with a special focus on loyal users (Williams 2022).

Prior studies have left two key questions unanswered about datawalls: the extent to which they actually exist in the field and whether they represent a new, distinct strategy for digital news organizations. Currently, there is no empirical study evaluating the use of datawalls in the digital news industry; instead, they are conceptualized as hypothetical models and evaluated primarily through surveys of consumer attitudes (Evens and Van Damme 2016; Portilla 2018). One reason may be that the distinction between paywalls and datawalls is not always clear. Prior literature does not define paywalls narrowly as ones that require *financial* payment, so perhaps datawalls are simply "a new sort of paywall" (Evens and Van Damme 2016). Since digital news outlets already collect significant data from users (Adams 2020; Kammer 2021; Willig 2021), including implicitly through paywalls (Olsen and Solvoll 2018), it is unclear whether datawalls represent a new business strategy for digital news organizations.

This calls for analytical clarity through empirical examination about what exactly constitutes a paywall and a datawall. Since the literature suggests that datawalls represent a form of paywall, we introduce the term *financial paywall* to refer specifically to a paywall that limits access based on financial payment (see Table 1). Notably, all financial paywalls collect user data in the course of conducting financial transactions and sustaining user sessions. Then, in order to provide a mutually exclusive term, we define a *datawall* as a paywall that limits access to registered users who "pay" solely by consenting to additional data collection, but not through financial payment. Thus, our first research question is: to what extent do datawalls actually exist in the field as a novel strategic implementation distinct from financial paywalls?

## Platformization and datafication

The emergence of datawalls may represent a shift in how platformization manifests in digital media. Platformization is defined as the "penetration of the infrastructures, economic processes, and governmental frameworks of platforms in different economic sectors and spheres of life" (Poell, Nieborg, and Van Dijck 2019, 2; see also Nieborg and Poell 2018). This definition characterizes platforms as digital infrastructures that organize personalized interactions among two-sided users (e.g., individual users and advertisers in the case of Facebook) "through the systematic collection, algorithmic processing, moneti[z]ation, and circulation of data" (Poell, Nieborg, and Van Dijck 2019, 2). This model of platformization highlights how platforms induce processes that reshape cultural practices and commodities. For example, Helmond (2015) showed how social media companies reshape economic logics by extending the reach of their data architecture. This perspective sees platforms as intermediaries whose influence arises from their power to mediate between infrastructures and applications.

This conceptual model of platformization prompts another question about datawalls. On one hand, financial paywalls represent news organizations' responses to platformization by displacing platforms to reinstate a more direct relationship between user attention and revenue. On the other hand, datawalls can exemplify a response both to the mixed success of financial paywalls and to the loss of third-party cookie tracking – which is not necessarily an outcome of platformization.

This framing calls for deeper scrutiny of the underlying mechanisms of platformization. Two key mechanisms identified by van Dijck, Poell, and de

**Table 1.** Distinguishing datawalls from financial paywalls.

| | Datawall | Financial paywall |
|---|---|---|
| Collects user data? | Yes | Yes |
| Collects financial payment? | No | Yes |

Waal (2018) are *datafication* and *commodification*. Datafication refers to the collection, quantification, and processing of data on social activity (Mayer-Schönberger and Cukier 2013). Flensburg and Lomborg (2023) have called for more empirical analysis and for examining datafication as a sociotechnical phenomenon, beyond general social understandings of datafication and its technical processes. Meanwhile, commodification refers to making objects, activities, emotions, and ideas valued extrinsically for their exchange value under capitalism. These processes are intertwined because the platformization era is characterized by a relentless logic of capital accumulation in which data can function both as a commodity—valued according to financial exchange – and as capital – with value increasing through circulation (Sadowski 2019; Zuboff 2018).

A significant body of journalism scholarship has explored audience datafication in terms of journalistic practices (Bodó 2019; Carpes da Silva and Sanseverino 2020; Christin 2020; Petre 2021) and tensions between journalistic values and platform-centered marketing imperatives (Neilson and Gibson 2022). We build on this scholarship by focusing on datawalls that seem to represent a different form of audience datafication – one that is peripheral to the core activities of journalism – such as news production and audience engagement – and instead interprets data as capital whose valuation is realized through circulation. Thus, our second research question: what, if any, is the role of platformization and datafication in the emergence of datawalls?

## Methodology

To address our research questions, we collected and analyzed two types of data: digital news websites and interviews with professionals in the digital news industry. Both data types were mutually formative throughout our iterative methods, in which we addressed our two research questions through a landscape analysis and a case study on *HuffPost*.

### *Web features and landscape analysis*

We addressed the first research question – to what extent do datawalls actually exist in the field as a novel strategic implementation distinct from financial paywalls? – mainly through a landscape analysis of key digital news outlets' websites, while also being informed by the interviews and case study detailed below. Landscape analysis is commonly understood as a method to locate "the key players in a field, sector or geography and [classify] them by relevant characteristics" in many business and nonprofit settings (The Bridgespan Group 2016). In engineering and information science studies, an exploratory landscape analysis is often used to identify and understand properties, called "landscape features," of an "optimization problem" by analyzing "samples" (Mersmann et al. 2011; Škvorc, Eftimov, and Korošec 2021).

We employed this exploratory landscape analysis approach to identify key features of datawalls and evaluate their frequency in the digital news industry. We included both legacy and digital-native media frequented by US audiences. Initially, we included the top 25 outlets accessed by US audiences in May 2022 as per *SimilarWeb*, which provides rankings of the top news websites[2] by traffic and audience (Majid 2022). We excluded sites that do not publish original general news reporting. This added 18 outlets to the sample. We eventually added *Reuters* and *Slate* based on interviewees' comments. In total, 20 outlets were included in the landscape analysis (see Table 2). The exploratory landscape analysis was conducted inductively in iterative phases between June and September 2022. First, we explored a set of news publishers to classify each site based on several emergent factors, such as whether a financial paywall or datawall was present and which benefits were locked behind each program. We iteratively refined this classification system by developing more precise dimensions as we gained insights from conducting interviews and as we analyzed additional news publishers, including specific categories of benefits (such as posting comments, access to newsletters, ad-free browsing, and bookmarking articles), single sign-on options, and corporate ownership structure. We discussed the analysis and findings until reaching agreement.

We examined each website using the definitions of a financial paywall and a datawall discussed above. We found that different paywalls limited access to different types of content, including general news, a subset of news, or other content, as well as on-site product features such as commenting platforms, personalized news recommendations, and the ability to save or bookmark articles. If a website limited access to at least one of these benefits to registered (logged-in) users without requiring financial payment, we classified it as having a datawall. To simplify the analysis, we excluded off-site features, such as features specific to mobile apps, print subscriptions, community events, merchandise, and ad-free experiences.

**Table 2.** Digital news outlets by access to news content and product features.

| | | Access to product features | | |
|---|---|---|---|---|
| | | Free | Datawall | Financial paywall |
| Access to news content | Free | (A) CNN* Breitbart | (B) Fox News Daily Mail BBC NBC News HuffPost | |
| | Soft freemium financial paywall | (C) CNBC Politico* USA Today | (D) NY Post US News | |
| | Soft metered datawall | (E) Guardian | (F) Reuters | |
| | Soft metered datawall with freemium financial paywall | | (G) Insider | |
| | Soft metered financial paywall | | (H) Forbes Slate WaPo | |
| | Hard datawall with metered financial paywall | | (I) NYT | |
| | Hard financial paywall | | | (J) WSJ |

*Notes.* The horizontal dimension represents different models of regulating access to *product features*, with the most open models on the left ("Free") and the most limited on the right ("Financial paywall"). The vertical dimension represents different models of regulating access to *news content*, with the most open models on the top (A, B) and the most limited on the bottom (J). The shading reflects our argument that access controls exist on a spectrum. The outlets with the most open access to both news content and product features have no shading (A, C), whereas the outlet with the most restricted access on both dimensions in our study (J) has the heaviest shading. * These are *empty datawalls* – users are invited to create accounts and log in, but no specific benefits are offered in return.

## *Interviews and case study*

We addressed the second research question – what, if any, is the role of platformization and datafication in the emergence of datawalls? – through an exploratory case study of *HuffPost*[3] based mostly on interviews conducted between April and August 2022. This was also informed by the landscape analysis described above, which revealed several cases of digital news outlets that featured datawalls, including *CNN*, *NBC News*, and others. We used digital mastheads, LinkedIn searches, and our personal networks to identify potential participants who previously worked or are currently working in product, audience, editorial, and business strategy roles and invite them to provide an interview. In total, we conducted interviews with 10 individuals at four digital news outlets from the landscape analysis (see Appendix A). Zoom interviews lasted for one hour on average and participants were offered US$25 for their time. The recruitment and interview protocols were approved by the authors' institutional review boards. Five interviewees were current employees and five were former employees (within two years) of the outlets of interest. In addition, interviewees drew from their experiences at previous or current employers, covering a total of nine news outlets. Most interviewees requested to keep their identities anonymous in order to feel comfortable reflecting honestly on their experiences. Alongside the interview data, we also examined website pages related to *HuffPost's* membership program, including historical pages accessed through the Wayback Machine.

During the iterative processes of the landscape analysis and the interviews between April and September 2022, *HuffPost* quickly emerged as the most feasible and compelling case. First and foremost, we were able to secure the most interviews with former and current employees at *HuffPost*, which allowed us to conduct a case study on it with the richest context. This was facilitated by the first author's prior work experience at *HuffPost*. The first author's insights further helped unpack and contextualize the *HuffPost* case. Also, *HuffPost* stood out because its datawall – the free tier of its membership program, HuffPost Plus – was relatively mature, having been launched in 2019. In addition, platformization pressures at *HuffPost* are uniquely transparent since its membership program was launched under the stewardship of Verizon Media executives who also oversaw Yahoo's growth strategy. *HuffPost* further survived an ownership transition when it was sold to *BuzzFeed* in 2021. As such, we decided to more closely examine and introduce *HuffPost* as an exploratory case study, following the "purposive sampling" logic established in qualitative studies to acknowledge opportunities for intensive study and choose the case from which we can learn the most (Stake 2005).

When we analyzed and presented the findings of the *HuffPost* case study, we drew on all the interviews,

including with non-*HuffPost* employees, as well as the landscape analysis to contextualize the findings. Specifically, we traced changes to its membership program over time by analyzing the *HuffPost* website on the Wayback Machine. This allowed us to follow the evolution of specific membership program features mentioned by interviewees.

## The spectrum between financial paywalls and datawalls

In the exploratory landscape analysis, we identified various access models for news content and product features using combinations of datawalls and financial paywalls. We situated the landscape analysis on two dimensions: (i) access strategies and (ii) benefits (i.e., news content versus product features).

For the first dimension, we began with three access strategies: (1) open access (free of financial payment and registration), (2) datawall (registration only), and (3) financial paywall (financial payment and registration). We also distinguished between *hard* datawalls and financial paywalls, which limit all access, and *soft* datawalls and financial paywalls, which offer limited access, most often with a metered wall (e.g., any three news articles per month).

For the second dimension, we distinguished between news content and product features. Previous research has focused on news content, but we found many examples of news websites that attract users to register or pay for benefits that do not relate directly to editorial content. These included web-based product features (access to view and post comments, bookmarking or saving articles, personalized news recommendations, following specific topics or reporters), mobile product features (ad-free browsing, exclusive app access, reading history), off-platform benefits (discounts at third-party retailers, free merchandise), and other benefits (events with journalists, access to documents such as policy reports or research). To simplify the analysis, we focused on web-based product features because we exclusively analyzed news websites, which generally attract larger audiences than mobile apps.

Two notable categories of product features were commenting platforms and newsletter subscriptions. The ability to post comments was often the major– and sometimes the only –product feature behind datawalls. This aligned with expectations from previous research that news organizations consider user comments valuable for engaging loyal users and supporting distribution (Braun and Gillespie 2011). In contrast, access to newsletter subscriptions was often available with different levels of access for non-registered users (who could subscribe) and logged-in or paid users (who could access a newsletter management dashboard and subscribe to exclusive newsletters). Since newsletter access did not neatly fall into a single access category, it was excluded from our exploratory landscape analysis.

### *(Mostly) open content and open features (Groups A and C)*

The first category includes publishers that offer free or mostly free news content and do not limit access to product features. This includes publishers that offer niche paid products, such as Politico Pro for policy researchers or CNBC Pro for investors. The key characteristic of this group is that access to both free core news content and product features is not restricted by a financial paywall or datawall.

*Politico* and *CNN* display login buttons and do seem to maintain a registration system, but they do not restrict access – at least not yet. We refer to these as *empty datawalls* because they invite users to create accounts and log in but offer no specific, immediate benefits in return. Instead, it appears that their primary function is to collect users' email addresses and subscribe them to mailing lists for marketing emails. Furthermore, these registration flows may allude to a former or forthcoming datawall. For example, *CNN* may have launched its empty datawall in anticipation of launching its subscription service, *CNN+*, in March 2022. However, *CNN+* was short-lived; it was shut down the following month.

### *(Mostly) open content with datawalled features (Groups B and D)*

The next group of news publishers provides (mostly) free access to news content while restricting access to certain product features with a datawall. The publishers in Group D offer nearly all their news content for free, but they differ from those in Group B by also selling paid content catering to niche markets. For example, *US News* offers full access to college rankings through College Compass, and the *New York Post* offers Post Sports + for enhanced sports content.

All publishers in these two categories use a datawall to restrict feature access while the majority of their core news content is available for free. In particular, every outlet in this category offers the ability to comment only to registered users, while other product features vary. Several outlets (*BBC, HuffPost, NBC News, US News*) offer utility features such as the

ability to bookmark or save articles and access to personalized content recommendations, while others (*Fox News, Daily Mail, New York Post*) do not offer any additional features.

One interviewee from *Reuters* described the value of the feature datawall as threefold. First, it served as a "stepping stone" to launching a financial paywall one day in the future. Second, the registration form collected user data, such as occupation, country, or household income, which could be used to eventually enhance digital advertising. Third, the registration flow included a marketing opt-in so that the editorial and events teams could send newsletters and marketing promotions.

### Soft metered content datawall (Groups E, F, and G)

The next category of publishers includes the *Guardian, Reuters,* and *Insider,* which have implemented soft metered datawalls on their news content and, in some cases, on product features as well. The *Guardian* offers open access to all news content for the first few articles, but once readers exceed a limit, they must register or log in to continue reading. *Insider* has also imposed a metered datawall on its content, but it also publishes some articles that are exclusively available to paid subscribers. *Reuters's* datawall was launched as a first step to a paywall to be launched in June 2021, but the timeline has been delayed because of a dispute with financial data provider Refinitiv (see Li 2021).

The latter two outlets also encourage readers to register by highlighting specific product features locked behind datawalls. *Reuters* has a menu called My View in its header which links to three features that allow users to customize their content discovery experience, follow specific topics or reporters, and bookmark articles. These features are also described on the registration wall prompt (see Figure 1). Similarly, readers on *Insider* are occasionally invited to use a utility button to save articles, which leads them to the registration flow.

The *Guardian* is distinct because it features a porous soft metered news content datawall while providing open access to all product features. We define a *porous* datawall as one in which readers can bypass the registration wall indefinitely, in this case by opting to "do it later" (see Figure 2). The site is also unique because it offers readers the ability to either make a "contribution" akin, but not equivalent, to a charitable donation or to purchase a subscription to "support" the *Guardian* while also unlocking additional benefits, primarily on mobile. Therefore, the *Guardian* is distinct from *Reuters* and *Insider* because its product features are available for free.

### Soft metered content financial paywall with datawalled features (Group H)

The *Washington Post* (WaPo), *Forbes,* and *Slate* have all implemented metered financial paywalls on news content while locking certain features behind a datawall. The difference among the three outlets is that both the *Washington Post* and *Slate* limit access to their commenting platforms to readers who are registered, whereas *Forbes* allows anonymous readers to comment while limiting access to other features, such as the ability to follow topics and reporters.

### Hard datawall with metered financial paywall (Group I)

The next group includes the *New York Times* (NYT) as a single case. While the *New York Times* is well known for its successful financial paywall, it is unique in that it requires users to create an account and log in to access most news content. Since the datawall prompt appears immediately upon loading the article page, readers cannot use product features such as comments without logging in, either. Thus, we call this a *hard datawall*. Once a registered reader reaches the article limit for datawalled news, they must make a financial payment for a subscription to access unlimited news articles. This is the second-most restrictive model after the hard financial paywall.

### Hard financial paywall (Group J)

The final group is news publishers that have implemented a full financial paywall, restricting access to both news content and product features. The *Wall Street Journal* (WSJ) is the sole outlet in this group among the news outlets in our study. It requires readers to make a financial payment upfront to access both news content and product features such as comments and audio recordings of articles. The financial paywall requires readers to register and log in, but it does not qualify as a datawall because readers must also make a financial payment to access any benefits.

In summary, the landscape analysis shows that datawalls do exist, although they are more complex than conceptualized in the literature. Primarily, we have found that datawalls exist on a spectrum with financial paywalls – one that is difficult to describe on a linear scale. Instead, we have characterized

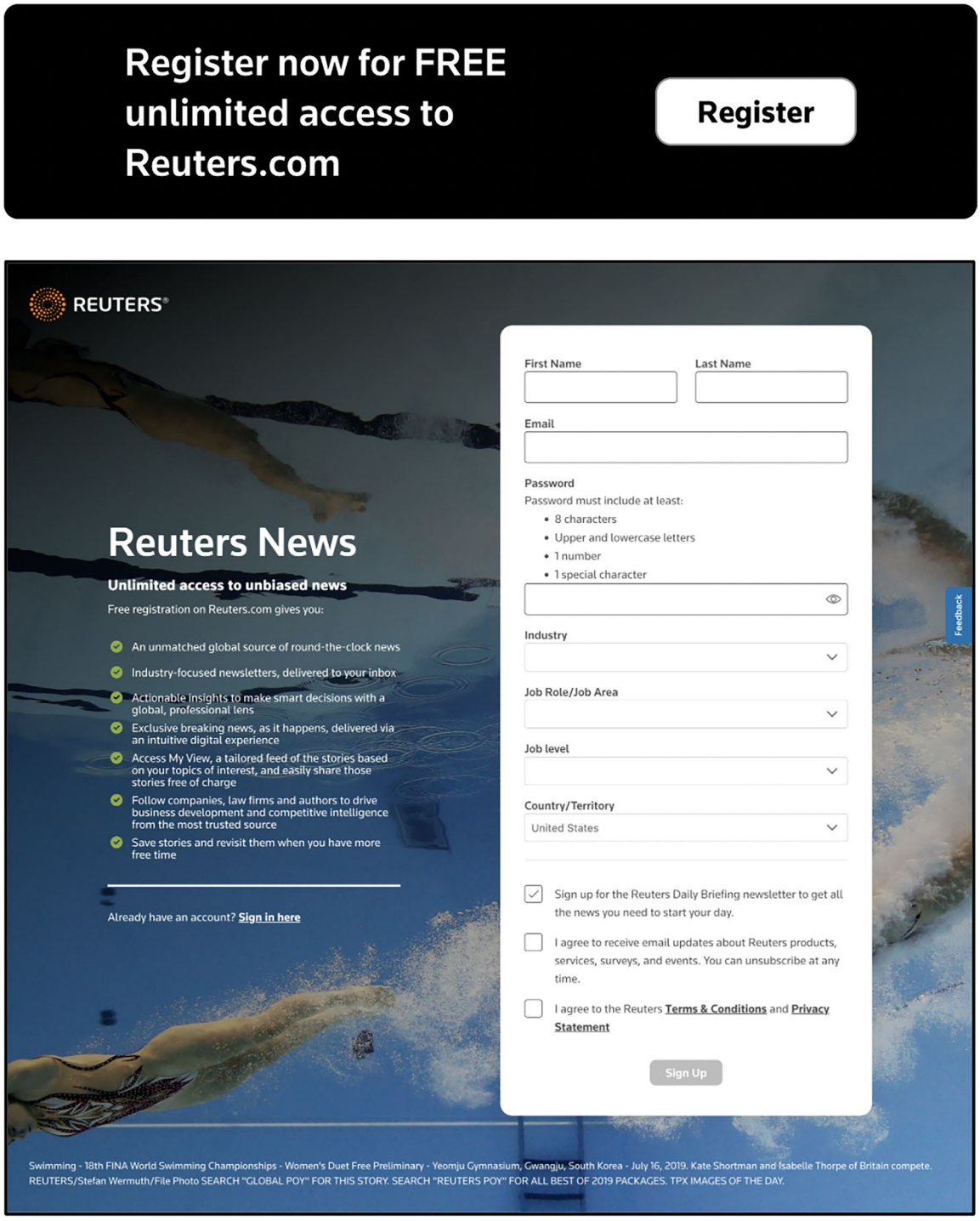


**Figure 1.** Prompt to register audience to *Reuters'* datawall.

various walls across two dimensions – news content and product features – with distinct extremes and several additional categories in between. On one end, access to content and features can be free and unlimited, as in the case of outlets such as *CNN*. On the other end, both content and features can be restricted only to users who make financial payments, which has long been the case for the *Wall Street Journal*. However, most outlets fall somewhere in between. Next, we turn to the case of *HuffPost* to examine how platformization and datafication are implicated in the emergence of datawalls.

## Case study: *HuffPost's* datawall

In the case study of *HuffPost*, we explored the roles of platformization and datafication play in the development of its membership program, HuffPost Plus.

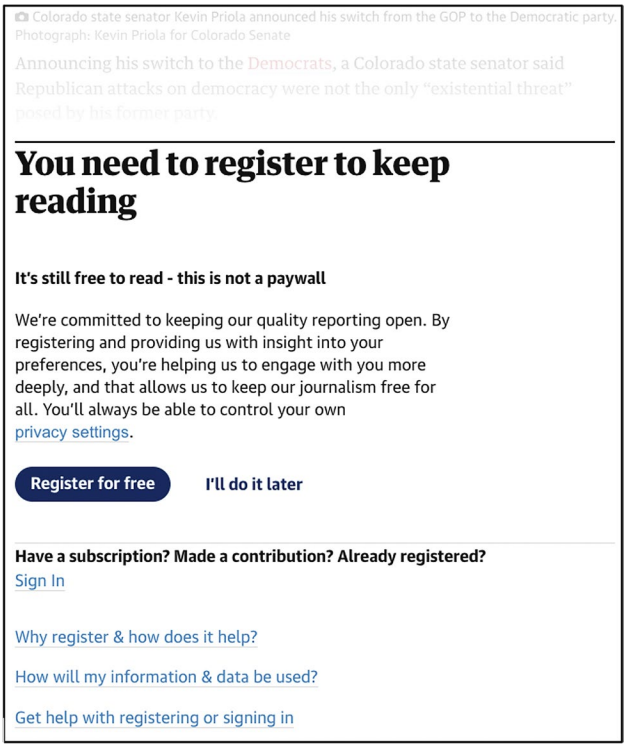


**Figure 2.** The *Guardian's* porous datawall.

*HuffPost*, founded as *The Huffington Post* in 2005, became the first digital-native media outlet to win a Pulitzer Prize in 2012. Over the past decade, it has undergone three major ownership changes. In 2011, it was acquired by AOL; then, in 2015, AOL was acquired by Verizon Communications, which then created a new subsidiary, Verizon Media, once it also acquired Yahoo! Inc in 2017. In the same year, *Huffington Post* was rebranded as *HuffPost*—with the tagline "It's Personal." During this time, *HuffPost* employees interacted with Yahoo teams and reported up to Verizon Media executives who also oversaw the full suite of Yahoo brands. In 2021, *HuffPost* eventually left Verizon Media when it was sold to *BuzzFeed*.

## Membership launch and challenges

*HuffPost* launched its membership program, HuffPost Plus, in 2019. According to a *HuffPost* senior editor, it was launched for two reasons. First, it was an opportunity to test whether readers would be willing and interested to make a financial payment for a subscription program in the future –something executives had been pushing for. Second, the leadership team did not think news content behind a financial paywall would be consistent with *HuffPost's* brand as a people-first news outlet. The membership program was launched with two tiers – a paid tier and a free tier – to reconcile these competing priorities for widespread, democratic access to news on one hand and revenue diversification on the other. Both tiers kept access to news free while offering different product and off-platform features as exclusive benefits. The free tier – which only required creating a membership account on the website or mobile app – was largely seen as a stepping stone to the paid tier as well as an opportunity to increase both advertising and direct revenue by growing the "top of the funnel,"[4] reducing the company's dependence on social media platforms for bringing in audiences. The senior editor described it as such:

> We basically tried to figure out how to connect a little bit of our audience strategy from not just only the top of the funnel but also like bringing people down the funnel and making them more engaged, more loyal. We realized membership can be one way of doing this. We wanted to kind of create a fan club – people who came to us a lot.

The free tier included product features that would excite the most loyal, engaged users: bookmarking articles, synchronizing the cross-platform experience across web and mobile apps, and managing newsletters. Meanwhile, the paid tiers – either $5.99 per month or $99.99 per year – added ad-free browsing on the mobile apps (where loyal audiences were most likely to be found), a discount on a newly launched HuffPost Stuff store, and, for members who paid annually, a t-shirt. Eventually, the editorial team also began organizing virtual events, such as Q&A sessions with the head of the politics team and with reporters after publishing noteworthy stories.

From the beginning, the technical infrastructure for the paid membership tiers complicated the company's ability to centrally manage its data. A crucial factor was that Apple required apps to utilize its in-app purchase mechanism and allow users to pay for membership without being logged into an account. Therefore, iOS app users were able to pay for *HuffPost* membership without creating a *HuffPost* account, bypassing the membership program's implicit datawall. Data about members who joined *via* in-app purchase was therefore controlled by Apple, limiting *HuffPost's* assembly of a central, comprehensive membership database. Thus, the launch and challenges of *HuffPost's* membership program were situated in the infrastructural power of Big Tech platforms.

## Capturing first-party data

In this context, executives from Verizon Media began to take greater interest in growing the free tier rather than the paid tiers of *HuffPost's* membership program. The conglomerate's executives were nervous about the

expected loss of third-party cookies in Google Chrome. This was particularly important to Verizon Media because its most valuable properties were not just its well-known utility and media services such as Yahoo Mail, Finance, and Sports, but also included its advertising business, which relied on third-party cookies to collect data and target users with digital ads across the internet. According to a *HuffPost* senior engineer, "Verizon Media was the publisher and it placed the ad network, which creates a kind of interesting conflict of interest, in terms of what is the business's goal." Thus, the free tier – though envisioned as a precursor to a financial paywall and an engagement opportunity – was repurposed as a vehicle to collect first-party data to prepare for the impending loss of third-party cookies.[5]

In early 2020, *HuffPost* moved its commenting platform behind its free tier of membership – behind the datawall. The platform was powered by OpenWeb (then called Spot.IM), a third-party tool that was integrated into the article page. Previously, users would sign into the platform *via* single sign-on (SSO) using their Facebook or Google accounts, and OpenWeb managed the user data. After the migration of the commenting platform behind *HuffPost's* datawall, users needed to create new accounts with *HuffPost* using the company's membership registration system. *HuffPost*'s registration page required users to agree to terms and services that meant that Verizon Media owned data previously held by Facebook, Google, and Spot.IM.

Despite significant growth from the commenting platform migration, members in the free tier still comprised a minority of *HuffPost* users, so Verizon Media executives asked for more dramatic growth in user registrations. However, the team found that large numbers of readers were unlikely to complete an extensive registration process – which included agreeing to Verizon Media's terms of service, signing up either with Yahoo SSO or by creating a new username and password (since Facebook, Google, and Apple SSO were not permitted by Verizon Media anymore), and verifying their email address – if the primary value of registration was access to product features targeted to only the most loyal users. Thus, the editorial and product teams regularly discussed the possibility of moving news content behind the datawall.

This momentum was halted when *HuffPost* was sold to BuzzFeed in early 2021. Upon announcing the sale, all paid memberships were canceled, partially elapsed subscriptions were refunded, and the entire paid tier was shut down. The free tier – the datawall – was permitted to continue once *HuffPost* migrated the backend technical infrastructure from Verizon Media's Yahoo-based stack to *BuzzFeed's* stack.

### *The value of first-party data*

One interviewee from *HuffPost's* data team expressed that *HuffPost* was in a unique situation because of its relationship with Yahoo, a less commonly considered – albeit still massive – platform company:

> I don't know if the experience at Verizon Media is consistent with the experience across all other places. It's different from most publishers because it is a platform in a lot of ways, the same way that it is a publisher. That's where a lot of the desire to get people to log in comes from, because then there is extra value in an email account. That's really what it boils down to for Verizon Media.

A senior editorial team member from *HuffPost* shared how the value of first-party data had been framed to them in a similar way:

> The theory was that Verizon [Media] is really big and has a lot of access to a lot of people who could be members or, in some way or another, loyal customers of all different brands. And we could create potentially a set of original data – first-party data –that could compete with Google and Facebook.

A product team member also recounted that *HuffPost's* membership program eventually became "less focused on as a revenue model and became a little bit more focused on getting data from people." This was because "once we [*HuffPost*] have that first-party data – you've created a login – we have your identity and we can track you, follow you, and target you." The value of first-party data was further discussed by several employees from *HuffPost* in terms of "put[ting] [*HuffPost*] in a position to have a better profile for ad targeting and for any kind of monetization [through] targeting," and even "boost[ing] the value of *HuffPost*" when anticipating the sale of *HuffPost* to new companies. Thus, the value of first-party data was widely recognized across teams at *HuffPost*.

However, the specific plan for monetizing first-party data remained ambiguous. For example, since *HuffPost* operated on a different host (huffpost.com), its cookies would be third-party to Yahoo's platforms (hosted on yahoo.com) and vice versa. Thus, an engineer from *HuffPost* described one potential scenario:

> I was always very nervous about that at *HuffPost* because, if third-party cookies go away, basically all of our revenue vanishes. Because we were connected with Yahoo and the broader Yahoo network and we had our own ad network, I think the plan was, if Verizon Media still existed today, that there would

> have been a point where *HuffPost* would have been moved to yahoo.com/huffpost. Because then third-party cookies don't matter. We have a first-party cookie which is Yahoo and that's profitable and strong and has brand recognition. That is absolutely how they would have solved that problem if it came to it.

These quotes show how various *HuffPost* employees were concerned that the loss of third-party cookies would undermine their ability to accumulate data. *HuffPost's* status as a subsidiary within a platform company complicated its ability to navigate the challenges of platformization since Verizon Media offered in-house technologies that could not be easily adopted.

### *Negotiating logics*

Decisions about how to grow the membership program also surfaced tensions about balancing the multiple goals and audiences of *HuffPost*'s journalism. For example, a senior editor from *HuffPost* reflected on the rebranding initiative:

> I really thought it would be wrong to take a subscription model [for news content], considering the whole way we had branded, identified, and been trying to sketch out what *HuffPost* was going to be in this iteration. It was about democratizing news – news for non-elites. The idea of charging people would make it impossible to actually achieve our mission.

This attitude was not limited to the newsroom. A *HuffPost* engineer added:

> Engineering was philosophically very much against a paywall [for news content]. One of the reasons I loved *HuffPost* so much is everyone – it's not just the editors – we all had sort of the same ethos, the same view of our users, and really bought into the brand and aligned with it. And charging users for content was like a kind of antithesis to *HuffPost's* DNA.

The tension over whether news content should be behind a financial paywall is not particularly new, considering that news organizations have long grappled with appeasing both audiences and advertisers. Indeed, prior research has highlighted that even news organizations that implement financial paywalls sometimes "drop" them for particular goals, such as to "inform the public during crises and emergencies" or to "provide wider access to nonemergency content seen as publicly valuable" (Ananny and Bighash 2016). These decisions reveal how news organizations continually balance diverse motivations, including "sociotechnical negotiations between democratic and commercial logics" (Ananny and Bighash 2016, 3360). Therefore, *HuffPost's* continued decision to sustain free access to its news reveals its employees' consensus about the primary role of their journalism.

Even though *HuffPost* did not end up limiting access to news content, the discussion about moving news content behind a wall recurred occasionally as *HuffPost* struggled to remain profitable. The *HuffPost* case highlights how the pressures of platformization intersect with the longstanding tension between editorial priorities and revenue in journalism. For example, *HuffPost* was able to serve content quickly by leveraging VZM's content delivery network. But if a financial paywall was introduced, it would have slowed down and complicated the process of loading news in users' browsers because it would have to be implemented differently across different platforms. An engineer from *HuffPost* described this challenge:

> The method to put up the [financial] paywall is different on every platform. It would have been different [from the web] in the app, and we also had off-site distribution. So Google AMP (Accelerated Mobile Pages), Apple News… we were in Yahoo! News as a vertical. If you're behind a [financial] paywall, do we show [our articles] on Yahoo or not? Because if you can get it on Yahoo for free, why go to *HuffPost* at all? So it raises all these really complicated questions for how we're going to distribute our content… It [also] would have hurt search engine rankings because the search engines have to be able to see it, but then users can't. And sometimes search engines may penalize us for that. It gets very, very, very messy.

This predicament illustrates the ways decisions about how and whether to limit access to news content surfaced questions about *HuffPost's* role and its responsibilities to the public and advertisers. Negotiations existed not only among editorial leaders in terms of the civic responsibility of news, but also among technical and business staff given their distribution strategy's dependency on partnerships with platforms whose content curation practices and algorithms drove significant traffic and therefore ad revenue. This might be why more news outlets are considering limiting access to product features rather than news content behind datawalls, as illustrated in the landscape analysis earlier.

## Discussion

### *A spectrum of paywalls*

In our study, we found that datawalls do exist, but on a spectrum with financial paywalls with many hybrid models in between. There is no clear binary, as outlets pursue different access strategies for news content and

product features – with product features more likely to be restricted (behind either datawalls or financial paywalls) than news content. For example, outlets such as *Fox News, Daily Mail, BBC, NBC News,* and *HuffPost* keep major features – most often access to comments – behind datawalls while providing free, unlimited access to news content. On the other hand, the *Guardian* only limits access to content – albeit through a soft, porous datawall – without restricting access to features. This indicates that publishers may be more likely to require data as payment from readers in order to access product features compared to news content.

Another way outlets pursue diverse access strategies is by modulating access using different mechanisms and thresholds. Prior scholarship has addressed different mechanisms such as hard, soft-metered, and soft-freemium paywalls (Myllylahti 2014) and different financial paywall thresholds for news content (Olsen, Kalsnes, and Barland 2021). Our findings build on this research by also examining datawalls and investigating thresholds for both news content and product features. Financial paywall strategies examined in our study include meters, freemium models that restrict access for either individual articles only or niche content, and hard financial paywalls, whereas datawall strategies include meters and hard datawalls. Sometimes, publishers pursue both at the same time. For example, the *New York Times* keeps access to essentially all content and features behind a datawall and also further restricts news content with a metered financial paywall; *Insider* keeps all content behind a metered datawall and further restricts access to selected articles to users who subscribe to the financial paywall; and *Forbes*, *Slate*, and the *Washington Post* keep content behind a metered financial paywall while keeping product features behind a datawall. In general, the landscape analysis demonstrates that, among the outlets we studied, it is more common for outlets with product features behind datawalls to offer free access to news content rather than keeping news content behind a financial paywall. In addition, nearly every outlet in our landscape analysis restricts access to *something* – news content and/or product features – by requiring registration and login, whether in the form of a datawall or a financial paywall.

Finally, we have found that datawalls can serve as stepping stones to financial paywalls in two ways. First, in terms of the user journey, datawalls are used to increase a user's likelihood to make a financial payment in the future. This is supported by strategies such as email marketing, to which users are often subscribed automatically once they create an account. Second, in terms of business strategy, news outlets can use datawalls to test technical infrastructure, collect data to test assumptions and change strategy, and register readers to reduce friction so that users only need to process a payment in the future to become paying subscribers. Thus, our evidence suggests there may not be a transition from financial paywalls to datawalls, as implied by prior literature, but rather a parallel emergence of *both* financial paywalls and datawalls.

Collectively, these findings seem to validate our conceptualization of both financial paywalls and datawalls as distinct forms of paywalls. We defined these terms earlier in response to ambiguity about what constitutes "payment" in a paywall. However, a simple binary division between financial paywalls and datawalls is likely insufficient to account for the spectrum of paywall strategies across the digital news industry. After all, in the cases of *HuffPost* and *Reuters,* datawalls have served less as specific strategies but rather as evidence of an evolving process of experimentation as digital news organizations grapple with a changing sociotechnical and economic landscape. Elucidating what this spectrum of paywall strategies looks like and how news organizations navigate them requires more detailed case studies and ethnographic research in the future.

In such future research, it will be important to consider how news organizations and their paywalls navigate a host of infrastructural and strategic considerations beyond what form of payment to demand in exchange for access. For example, while our study sought to empirically research datawalls in the field and thus focused primarily on web experiences, news organizations do not necessarily pursue identical strategies across platforms. *HuffPost*, in fact, had challenges accessing data from its iOS mobile app audiences because users were able to purchase membership through Apple's infrastructure, bypassing *HuffPost's* account registration system. This complicates whether companies can collect first-party data on websites compared to mobile apps—and raises questions about other platforms. It also illustrates questions digital news outlets grapple with related to how to maximize the value of audience data by preferring internal infrastructure (e.g., *HuffPost*'s registration page only offered Yahoo as an SSO option) and avoiding platform companies. Thus, decisions to implement various forms of paywalls take into account not only financial considerations, but also sociotechnical dynamics and political economy of platformization, and thus require further analysis.

## Platforms as templates?

### The role of platforms

Our findings also suggest that platforms have played an important role in the emergence of datawalls in

a way that exceeds the definition of platformization cited earlier: the "penetration of the infrastructures, economic processes, and governmental frameworks of platforms in different economic sectors and spheres of life" (Poell, Nieborg, and Van Dijck 2019, 2). We found little specific evidence of platforms "penetrating" into the datawall model with their processes and infrastructures.

Instead, we suggest that datawalls may represent digital news organizations' attempts to *adopt platform logics* themselves. In other words, datawalls could constitute a form of platformization in which platforms serve not as direct infrastructure or architecture, but rather as *templates* for other sectors to emulate. According to our study, datawalls are likely not simply short-term monetization strategies but can rather be investments in a broader reconfiguration of digital news websites as data platforms – depending on how and to what extent the data get captured, processed, and used within and beyond the core functions of journalism. Indeed, a recent report by a trade association for digital publishers prescribed "a profound, historic shift from platform-controlled unknown user data to publisher-controlled known customer data" (Williams 2022, 3). Researchers have arrived at similar conclusions about more nuanced dynamics of platformization where platform pressures on digital news result in "forms of mimetic and normative isomorphism" (Laaksonen, Koivula, and Villi 2022).

Specifically, we have found evidence of digital news organizations adopting platform logics as a template in the case of *HuffPost*. The evidence of platformization is clear, beginning with Verizon's decision to create a media division with the purchase of Yahoo and AOL (which included *HuffPost*). We can understand the emergence of Verizon Media and the pressures it subsequently imposed on its subsidiaries as symptoms of broader trends of platformization. Specifically, these pressures – and the programs they produced, such as HuffPost Plus – represent attempts to grapple with Google's and Facebook's command over revenues in the digital news industry by imitating their practices.

In this context, our findings suggest that *HuffPost's* specific actions in building its membership program likely reveal a mimetic power of platform logics. Building and capitalizing on user loyalty is not a new strategy for digital publishers. However, the decision to leverage internal technical architecture (e.g., limiting access to features rather than news content, changes in SSOs) to collect and ultimately monetize first-party data resembles the business logic of a platform, where personalized interactions are organized among users through the systematic collection and monetization of data (Poell, Nieborg, and Van Dijck 2019). Thus, we see evidence that the digital news industry adopts *platforms as templates* and is responding by imitating the logic and infrastructure of platforms themselves. We argue that our findings on datawalls illustrate this dynamic by representing a shift in business strategies from collecting financial payment to harvesting data as a way to enhance economic value.

We do acknowledge that our case study on *HuffPost* cannot be generalized to the context of every digital news media outlet. However, the findings still signal emerging dynamics of adopting platform logics to capture, control, and profit from first-party audience data at play. This insight can be a guidepost for future research to explore how this might be the case for other news outlets. Our analysis suggests a latent and relational form of sociotechnical platformization, as we may not find direct traces of Facebook or Google in datawalls, and news or technical staff may not refer explicitly to such companies as models. Rather, this form of platformization – platforms as templates – could be better understood by systematically examining business strategies and trade associations.

### *Data capital in digital news*

Our exploratory study lends preliminary insights on the stakes of datafication in this evolving logic of platformization – platforms as templates – with the case of datawalls. While news-audience relationships have always been shaped by technological and economic conditions (Napoli 2011), including increasingly granular quantification of audience, datawalls do not seem to fit neatly in the platformization framework's key mechanisms of datafication and commodification. This was most evident in interviewees' inability to articulate specific financial value or concrete end uses of audience data collected through datawalls. Instead, interviewees alluded to ambiguous relationships between data collection and revenue opportunities. This shows that the goal of data collection through datawalls is likely to accumulate data whenever possible to be ready for multiple uses in the future, even though the data may not be immediately translatable to profit as a clear revenue source.

Thus, we argue that the emergence of datawalls illustrates how digital news organizations might be interested in data not for its value as a commodity but rather as *capital*. Sadowski (2019) has distinguished between these two forms of data to understand political economic models and practices today.

As capital, data represents a vehicle for investment toward the goal of endless accumulation, often demonstrated in the prevalent mantra of "collecting data first and figuring it out later" (Sadowski 2019, 4). The logic of *data as capital* is being adopted by various sectors, and this might be the case for the digital news industry as well through datawalls, while platforms contribute to the capitalization of data by reconfiguring markets with data-driven advertising networks.

Why does this logic of data as capital matter? It helps us understand digital news organizations' responses to a shifting technological and socio-economic environment in a more dynamic and relational way. For example, the Cookiepocalypse, described earlier, may be expected to reduce the power of individual platforms by limiting their ability to collect data from third-party cookies across many websites. However, at the same time, *the platformed logic of data as capital* is in fact reinforced when new actors such as digital news outlets seek to accumulate first-party data as well. This logic can be helpful to understand digital news organizations' unexpected responses to changes in data markets. For example, our interviewees alluded to the value of first-party data for increasing their newsroom's attractiveness to potential investors or new parent companies, suggesting that readers are not only commodifying their data in exchange for access to content and/or features, but also financializing themselves by contributing investment capital to news organizations seeking to grow their own economic value.

Digital news outlets' embrace of data as capital amid platformization may not be a surprising trajectory in the contemporary era of "surveillance capitalism" (Zuboff 2018). For example, news websites have for years exchanged data with many third-party applications, especially Google (Kammer 2021), thus transforming audiences into data points (Willig 2021) while subjecting them to ethically complex privacy policies (Adams 2020). Therefore, the logic of data as capital in digital journalism carries implications for user privacy that need to be reconciled with journalistic values.

## Conclusion

This study examined datawalls in digital news, which were conceptualized in prior literature but not yet empirically examined. First, we found that datawalls exist on a spectrum that is different on two dimensions – product features and news content, and that access is modulated through different mechanisms and thresholds, such as meters, porous walls, and hard walls. These findings suggest that the digital news industry is likely in an experimental phase, launching datawalls to test hypotheses and lay stepping stones to financial paywalls. Thus, datawalls bear a complicated relationship to financial paywalls and provide evidence of evolving data monetization and platform pressures.

We also explored how well these dynamics relate to platformization and datafication. Based on our findings, we argue that platforms play an important role in the emergence of datawalls in a way that exceeds coercive "penetration" since datawalls seem to represent digital news organizations' attempts to *adopt platform logics* themselves. In other words, datawalls can represent a form of platformization in which platforms serve not (only) as direct infrastructure or architecture, but rather as templates for other sectors.

Our study is limited by its methodology and level of analysis. First, the sample for the landscape analysis was drawn from the top digital news outlets among US audiences in the year 2022. This represents a particular moment in time with heightened platformization and datafication dynamics led by a small number of Big Tech platform companies based in the US. Therefore, future research should draw from more diverse samples, both temporally and spatially. On one hand, future research could look into longitudinal changes, building on our findings, as datafication pressures may intensify with complex interactions with artificial intelligence systems. On the other hand, future research could also analyze global markets and smaller, more nimble publishers such as nonprofit news or public service media. Lastly, as *HuffPost* is a digital-born news outlet, the insights from our exploratory case study cannot be generalized to every type of news organization; the goal of our study was simply to explore possible explanations of the fast-evolving dynamics of platformization and datafication on the ground.

Ultimately, our study contributes to scholarship on digital journalism, platforms, and datafication by providing analytic clarity about emerging platformization and datafication processes in digital news, and thereby also opening up new questions about similar dynamics in other sectors. Our study points to the need for further research about alternative spheres and sites in which platforms exert active pressure not only by penetrating existing technological, economic, and governmental systems but also – or perhaps instead – by emanating mimetic pressure that induces industry actors to adopt platform logics. While our study was

empirically focused on datawalls in digital news, we believe that examining *platforms as templates* can provoke similar analyses of alternative forms and implications of this logic in other sectors.

## Notes

1. See Nerone and Barnhurst (2003) for a more detailed discussion of how access, attention, and revenue have been recurring problems in the historical contexts wherein some news is paid and some is free.
2. The outlets included in the rank are based on SimilarWeb's classification of news and media publishers, refined by *Press Gazette*. The list encompasses outlets such as *Breitbart* and *Daily Mail*, whose lack of news values is highly contested (see Heft et al. 2020; Roberts and Wahl-Jorgensen 2022). We clarify that the goal of our study is not to uphold the legitimacy of all included outlets as news organizations; rather we decided to include them in our analysis in order to map how such diverse media outlets that are frequented by US readers as an il(legitimate) source of information are responding to and adopting the logics of platformization.
3. *HuffPost* has multiple country-specific editions around the world. We examined the United States edition (huffpost.com) since this study was focused on digital news outlets frequented by US audiences.
4. The funnel is a marketing metaphor to visualize a pipeline of activities to attract, engage, and convert an audience into consumers. The "top of the funnel" refers to activities intended to create brand awareness and attract new leads, while the "bottom of the funnel" refers to activities intended to convert leads into consumers through engagement, purchases, and brand loyalty.
5. The value of first-party cookies was echoed by interviewees from other outlets as well. For example, a *Reuters* employee suggested that journalism will "need[s] to own the relationship with readers or customers" and be a "trustworthy" actor as third-party cookies go away.

## Acknowledgments

The authors are grateful to Dr. Mike Ananny, members of the Center for Information, Technology, and Public Life (CITAP), and the anonymous reviewers for their generative feedback that greatly improved this article.

## Disclosure statement

The authors report there are no competing interests to declare.

## Funding

This work was supported by the Association for Education in Journalism and Mass Communication, the USC Graduate School, and the Annenberg Doctoral Student Summer Research Fellowship.

## ORCID

Rohan Grover http://orcid.org/0000-0002-9042-8259
Jeeyun (Sophia) Baik http://orcid.org/0000-0001-5057-1432

## References

Adams, P. C. 2020. Agreeing to surveillance: Digital news privacy policies. *Journalism & Mass Communication Quarterly* 97 (4):868–89. doi:10.1177/1077699020934197.

Ananny, M., and L. Bighash. 2016. Why drop a paywall? Mapping industry accounts of online news decommodification. *International Journal of Communication* 10:3359–80.

Bechmann, A., K. Bilgrav-Nielsen, and A.-L. Korsgaard Jensen. 2016. Data as a revenue model : Sharewall as a payment method and editorial algorithm in the news industry. *Nordicom Information* 38 (1):76–82.

Bell, E., and T. Owen. 2017. The platform press: How Silicon Valley reengineered journalism. *Columbia Journalism Review*, March 29. Accessed April 15, 2024. https://www.cjr.org/tow_center_reports/platform-press-how-silicon-valley-reengineered-journalism.php/.

Bodó, B. 2019. Selling news to audiences: A qualitative inquiry into the emerging logics of algorithmic news personalization in European quality news media. *Digital Journalism* 7 (8):1054–75. doi:10.1080/21670811.2019.1624185.

Braun, J., and T. Gillespie. 2011. Hosting the public discourse, hosting the public: When online news and social media converge. *Journalism Practice* 5 (4):383–98. doi:10.1080/17512786.2011.557560.

Carpes da Silva, G., and G. G. Sanseverino. 2020. Business model innovation in news media: Fostering new relationships to stimulate support from readers. *Media and Communication* 8 (2):28–39. doi:10.17645/mac.v8i2.2709.

Christin, A. 2020. *Metrics at work: Journalism and the contested meaning of algorithms*. Princeton, NJ: Princeton University Press.

Chyi, H. I., and Y. M. M. Ng. 2020. Still unwilling to pay: An empirical analysis of 50 U.S. newspapers' digital subscription results. *Digital Journalism* 8 (4):526–47. doi:10.1080/21670811.2020.1732831.

El Hana, N., M. Mercanti-Guérin, and O. Sabri-Zaaraoui. 2022. Cookiepocalypse: What are the most effective strategies for advertisers to reshape the future of online display advertising? Accessed April 19, 2024. https://ideas.repec.org//p/hal/journl/hal-03779006.html

Evens, T., and K. Van Damme. 2016. Consumers' willingness to share personal data: Implications for newspapers' business models. *International Journal on Media Management* 18 (1):25–41. doi:10.1080/14241277.2016.1166429.

Flensburg, S., and S. Lomborg. 2023. Datafication research: Mapping the field for a future agenda. *New Media & Society* 25 (6):1451–69. doi:10.1177/14614448211046616.

Heft, A., E. Mayerhöffer, S. Reinhardt, and C. Knüpfer. 2020. Beyond Breitbart: Comparing right-wing digital news infrastructures in six Western democracies. *Policy & Internet* 12 (1):20–45. doi:10.1002/poi3.219.

Helmond, A. 2015. The platformization of the web: Making web data platform ready. *Social Media + Society* 1 (2):205630511560308. doi:10.1177/2056305115603080.

Kammer, A. 2021. Resource exchanges between mobile news apps and third-parties. *Digital Journalism* 11 (6):1154–72. doi:10.1080/21670811.2021.2000455.

Laaksonen, S.-M., M. Koivula, and M. Villi. 2022. Mediated by the giants: Tracing practices, discourses, and mediators of platform isomorphism in a media organization. *New Media & Society*:146144482211222. doi:10.1177/14614448221122220.

Li, K. 2021. Reuters postpones website paywall amid Refinitiv dispute. *Reuters*, May 27. Accessed April 19, 2024. https://www.reuters.com/business/media-telecom/reuters-postpones-website-paywall-amid-refinitiv-dispute-2021-05-27/.

Majid, A. 2022. Top 50 most popular news sites in the US: Growth at NYT, Mail Online and New York Post in July. *Press Gazette,* August 22. Accessed April 19, 2024. https://pressgazette.co.uk/most-popular-websites-news-us-monthly/.

Mayer-Schönberger, V., and K. Cukier. 2013. *Big data: A revolution that will transform how we live, work, and think*. New York: Houghton Mifflin Harcourt.

Meese, J., and E. Hurcombe. 2021. Facebook, news media and platform dependency: The institutional impacts of news distribution on social platforms. *New Media & Society* 23 (8):2367–84. doi:10.1177/1461444820926472.

Mersmann, O., B. Bischl, H. Trautmann, M. Preuss, C. Weihs, and G. Rudolph. 2011. Exploratory landscape analysis. In *GECCO '11: Proceedings of the 13th Annual Conference on Genetic and Evolutionary Computation*, 829–36. New York: Association for Computing Machinery. doi:10.1145/2001576.2001690.

Myllylahti, M. 2014. Newspaper paywalls – The hype and the reality: A study of how paid news content impacts on media corporation revenues. *Digital Journalism* 2 (2):179–94. doi:10.1080/21670811.2013.813214.

Napoli, P. M. 2011. *Audience evolution: New technologies and the transformation of media audiences*. New York: Columbia University Press.

Neilson, T., and T. A. Gibson. 2022. Social media editors and the audience funnel: Tensions between commercial pressures and professional norms in the data-saturated newsroom. *Digital Journalism* 10 (4):556–78. doi:10.1080/21670811.2021.2004553.

Nerone, J., and K. G. Barnhurst. 2003. US newspaper types, the newsroom, and the division of labor, 1750–2000. *Journalism Studies* 4 (4):435–49. doi:10.1080/1461670032000136541.

Nieborg, D. B., and T. Poell. 2018. The platformization of cultural production: Theorizing the contingent cultural commodity. *New Media & Society* 20 (11):4275–92. doi:10.1177/1461444818769694.

Nielsen, R. K., and S. A. Ganter. 2018. Dealing with digital intermediaries: A case study of the relations between publishers and platforms. *New Media & Society* 20 (4):1600–17. doi:10.1177/1461444817701318.

Olsen, R. K., and M. K. Solvoll. 2018. Bouncing off the paywall: Understanding misalignments between local newspaper value propositions and audience responses. *International Journal on Media Management* 20 (3):174–92. doi:10.1080/14241277.2018.1529672.

Olsen, R. K., B. Kalsnes, and J. Barland. 2021. Do small streams make a big river? Detailing the diversification of revenue streams in newspapers' transition to digital journalism businesses. *Digital Journalism* :1–22. doi:10.1080/21670811.2021.1973905.

Petre, C. 2021. *All the news that's fit to click: How metrics are transforming the work of journalists*. Princeton, NJ: Princeton University Press.

Poell, T., D. Nieborg, and J. Van Dijck. 2019. Platformisation. *Internet Policy Review* 8 (4):1–13. doi:10.14763/2019.4.1425.

Portilla, I. 2018. Privacy concerns about information sharing as trade-off for personalized news. *El Profesional de la Información* 27 (1):19–26. doi:10.3145/epi.2018.ene.02.

Roberts, J., and K. Wahl-Jorgensen. 2022. Reporting the news: How Breitbart derives legitimacy from recontextualised news. *Discourse & Society* 33 (6):833–46. doi:10.1177/09579265221095422.

Sadowski, J. 2019. When data is capital: Datafication, accumulation, and extraction. *Big Data & Society* 6 (1):205395171882054. doi:10.1177/2053951718820549.

Shields, R. 2019. Google mulls third-party ad-targeting restrictions. *Adweek,* March 22. Accessed May 5, 2024. https://www.adweek.com/programmatic/google-mulls-third-party-ad-targeting-restrictions/.

Simon, F. M. 2022. Uneasy bedfellows: AI in the news, platform companies and the issue of journalistic autonomy. *Digital Journalism* 10 (10):1832–54. doi:10.1080/21670811.2022.2063150.

Škvorc, U., T. Eftimov, and P. Korošec. 2021. The effect of sampling methods on the invariance to function transformations when using exploratory landscape analysis. In *2021 IEEE congress on evolutionary computation*, 1139–46. New York: IEEE.

Stake, R. E. 2005. Qualitative case studies. In *The Sage handbook of qualitative research*, ed. N. K. Denzin and Y. S. Lincoln, 443–66. London: SAGE.

The Bridgespan Group. 2016. Market mapping and landscape analysis. Accessed April 18, 2024. https://www.bridgespan.org/insights/market-mapping-and-landscape-analysis

van Dijck, J., T. Poell, and M. de Waal. 2018. *The platform society: Public values in a connective world*. New York: Oxford University Press.

Walters, P. 2021. Reclaiming control: How journalists embrace social media logics while defending journalistic values. *Digital Journalism* 10 (9):1482–501. doi:10.1080/21670811.2021.1942113.

Williams, M. L. 2022. First-party data playbook for publishers: The quintessential guide for jump-starting your strategy. Accessed April 19, 2024. https://www.fipp.com/wp-content/uploads/2022/06/FIPP-First-Party-Data-Playbook.pdf.

Willig, I. 2021. From audiences to data points: The role of media agencies in the platformization of the news media industry. *Media, Culture & Society* 44 (1):56–71. doi:10.1177/01634437211029861.

Zuboff, S. 2018. *The age of surveillance capitalism: The fight for a human future at the new frontier of power*. New York: PublicAffairs.

## Appendix A: Interview participants

| Outlet | Team |
|---|---|
| HuffPost | Product |
| Reuters | Audience |
| Anonymous | Data |
| HuffPost | Editorial |
| HuffPost | Engineering |
| HuffPost | Editorial |
| HuffPost | Product |
| Anonymous | Data |
| Reuters | Editorial |
| HuffPost | Engineering |