Anne Helmond and Fernando N. van der Vlist

Social Media and Platform Historiography: Challenges and Opportunities

Abstract

In this article, we propose a methodological outlook for historical platform studies to increase the prominence of platform historiography in the field and practice of web history and archiving. We discuss the challenges of social media archiving and the research opportunities for 'platform' historiography by focusing on the distinctive characteristics of web-based social media 'platforms'. Based on our review of the literature, we argue that it is critical to foreground how contemporary platforms serve multiple user groups beyond end users (e.g. developers, business, investors) and how they operate on multiple levels (e.g. interface, architecture, ecosystem). By attending to the multiple sides and layers of social media beyond their end users only, we can reconstruct histories of platforms, not only as social networks, but also as technical artefacts, business organisations, and more. A focus on the materiality of platforms introduces numerous underutilised archived web sources that present significant entry points and research opportunities for platform historiography. We assess the availability of these archived sources across the leading web archives. Our results show that despite the challenges of social media archiving, platforms' resources are in fact well-preserved if we look beyond their end-user interfaces. Drawing on illustrative examples, we discuss two sets of entry points and materials for platform historiography at length: first, for writing developer-side histories, and second, for business-side histories. We conclude with recommendations for platform historians and archiving practitioners and reflections on the future of platform historiography.

Keywords: social media, platforms, web historiography, web history, web archiving

Introduction: Writing the Histories of 'Platforms'

In March 2018, news broke that Cambridge Analytica had gained access to the personal data and profiles of millions of Facebook users that were harvested without their consent and used for political campaigns that allegedly swayed elections. The scandal fuelled mounting criticism against the social media platform regarding its data-sharing practices. Only nine months later,

it was revealed that Facebook had provided access to users' personal data to some of the world's largest technology companies through special data-sharing arrangements and partnerships.[1] These recent revelations and scandals underscore the need for a better understanding of social media today through a historical awareness of how they emerged and evolved over time. That is, these issues require an understanding of social networks' evolution into social media *platforms* and of their evolving conditions for third-party development and data access.

Although socio-technical histories of social media exist,[2] most histories privilege cultural memory and digital heritage perspectives to capture the 'important cultural moments' of the past.[3] Since social networks are a relatively recent phenomenon – Facebook only turned fifteen this year – they are not always considered of historical significance. Their contemporary forms and operations are deemed more important than their evolutionary trajectories. Additionally, since social media are persistently updated, they pose severe challenges for historians and encourage certain forms of historical research at the expense of others.[4] In a nutshell, since the mid-2000s, social media have evolved from social networking services developed 'on top of' the open web's infrastructure to closed platforms for social media and then to closed platform-based ecosystems with large-scale networks of connected apps and integrations.[5] Meanwhile, their explosive growth has led to the prevalence of new kinds of socio-technical systems and infrastructures that now underpin the provision of social networking services and the larger ecosystem of 'connective' media, but which remain hidden.[6] These developments raise the need for critical enquiries into these hidden aspects of social media, which are not always immediately observable from the graphical end-user interface, to complement the existing social and cultural histories of social media platforms and their visible aspects.[7] In particular, such critical histories can contextualise – denaturalise and reveal the contingencies of – the explosive growth of social media platforms and their present dominance and infrastructural presence in all corners of society.

In this article, we present a methodological outlook for historical social media and platform studies. Histories of social media should reckon with their distinctive characteristics as 'platforms' – or what we refer to as their *platformness*. Platforms have mainly been studied from three perspectives. From an information systems perspective, they have been studied as the extensible codebases of software-based systems that provide functionality to connected apps and services.[8] From an economic perspective, they have been studied as multi-sided markets that mediate the interactions between multiple stakeholder groups.[9] From a media studies perspective, they have been studied critically as non-neutral intermediaries.[10] In this article, we argue why social media and platform historians should incorporate these multiple perspectives to capture the interplay between a platform's evolving programmability and its expanding use cultures and presence in society. As we propose, social media are characterised by *multi-sidedness*, or by how they cater to multiple stakeholder groups, as much as by *multi-layeredness*, or how they operate on multiple levels for these groups.[11] Due to these distinctive characteristics, social media platforms have extended far beyond their original boundaries and have become embedded in other websites, apps, and societal domains.

The current stage in the evolution of social media has been theorised as an interplay between processes of 'platformisation' of infrastructure and 'infrastructuralisation' of platforms,[12] which further stresses the need to take the multiple sides and layers of social media

platforms into account. This also suggests that social media platforms are inherently distinct from websites in general, even if most of them are web-based platforms. Accordingly, there is an urgent need for methodological reflections on writing the histories of social media platforms in ways that acknowledge the evolving nature of social media as platforms and thereby complement existing web and website historiography.[13] As we demonstrate, the multiple sides and layers of social media platforms provide many new and relevant entry points that represent significant research opportunities for web history and historical platform studies.

Building on existing reflections on web historiography and digital platforms,[14] we discuss the methodological opportunities and challenges of social media *'platform' historiography*. In developing this methodological outlook, we focus on the distinctive characteristics of social media as platforms and discuss their consequences and opportunities for web and platform historians. We encourage explorations of, and methodological reflections on, the many available yet underutilised archived web sources and their historiographical value. Platform historiography, like historiography in general, foregrounds the methodological considerations and reflections associated with the use of multiple sources to interpret platforms' pasts. Although large-scale web archiving initiatives such as the Internet Archive have existed since the mid-1990s, before the rise of social media, most web archives remain largely unknown and underused as a primary source for researchers.[15] We, therefore, explore the value and utility of numerous archived web sources as entry points for social media platform historiography. However, to trace the emergence and evolution of social media platforms, historians need to first 'reconstruct' and piece together the platform as an object of study using the archived material traces,[16] including the reconstruction of their multiple user groups and structural embeddings.

In what follows, we first discuss the specific issues and challenges of social media archiving and web historiography in general based on our review of the literature. Next, we discuss the materiality of social media 'platforms' and take stock of the material traces left behind. We also detail a method to examine the availability of these materials in the leading international web archives to derive platform histories. Then, we explore two sets of entry points for 'platform' historiography: first, from the perspective of developers; second, from the perspective of businesses and business developers. In both cases, we discuss significant research opportunities and reflect on the afforded histories. We conclude this article with historiographical reflections and recommendations for future platform historiography.

Social Media and the Archived Web

The histories of social media that have been written previously and that can be written in the future are tightly connected to the sources that have been archived, how they have been archived, for which purposes, and how they have been made available to researchers.[17] A core challenge of web archiving, Brügger contends, is that the archived website as an object of historical study is formed in the archiving process and is a reconstruction, and not a representation, of its

original manifestation online.[18] Therefore, web archiving strategies determine the affordances for present and future historical research.[19] Consequently, web archives come with particular 'research affordances' for conducting historical platform studies as they enable distinct modes of research.[20] In this section, we discuss the specific issues and challenges of social media archiving and their implications for platform historiography based on our review of the literature.

Issues and Challenges of Social Media Archiving

Social media archiving is often understood as an extension of 'traditional' web archiving.[21] Web archiving is conducted in different ways, including capturing screenshots and recordings, downloading specific web pages, crawling websites, and collecting web data via application programming interfaces (APIs), each coming with its own specific challenges and limitations.[22] Traditionally, web archiving has employed web crawlers – automated agents that systematically crawl, request, and retrieve collections of websites or web pages by following their hyperlinks, which are then made available in a web archive. Such web crawlers face challenges since some websites are not created in an 'archive-friendly' way because they contain dynamic content or streaming media, are password-protected or database-driven, or block access to web crawlers.[23] Built on top of the open web, social media have inherited and intensified some of these challenges as their end-user interfaces and content have become increasingly dynamic and personalised. User content and data are stored in platform databases, which require users and crawlers alike to log in, and most popular social networks actively block web crawlers from accessing their sites.[24] Additionally, the volume and velocity of social media content generation and the distributed nature of social media platforms pose severe archival challenges.[25] Therefore, while it is technically feasible to archive social media like other kinds of websites using web crawlers, Thomson contends that 'the networked infrastructure of social media pages, as well as the underlying database systems that store user data, differentiate social media substantially from the traditional World Wide Web', thereby foregrounding the unique properties and archiving challenges of social media platforms.[26]

Thomson further argues that platforms are distinct from other websites as they provide 'large quantities of machine-readable data', which requires researchers to come up with new approaches to capture these data.[27] API-based archiving is more attuned to the architectures of social media platforms and enables such programmatic retrieval of structured data forms directly from social media's underlying databases. As a result, while web crawling remains a popular approach to web archiving, social media archiving typically employs APIs to retrieve and archive data, when these interfaces are available.[28] Thus, the source for traditional web archiving is the website, whereas the source for social media archiving is the API. However, API-based archiving comes with its own challenges as platforms determine and govern which data are available, at which access rate, and at which price point.[29] In this sense, the turn to 'API-based research' in contemporary social media research has been replicated in social media

archiving and has inherited the consequences of the increasing research limitations on APIs.[30] But even more fundamentally, by conceiving of platforms as databases, API-based archiving privileges data-driven and interaction-oriented forms of historical research in the terms specified by the platform itself, including histories of public user profiles, pages and groups, and specific hashtags. Moreover, by extracting data directly from databases, the content and interactions in question are detached from their original contexts, including the end-user interface and third-party clients.[31] Finally, these data only provide a partial view of the platform, and its operations as platforms are more than their data.

Table A-1 (see Appendix) summarises the main issues of social media archiving, their associated challenges, and the kinds of histories that are privileged, which inform our proposed entry points for historical platform studies.

Multi-Sided and Multi-Layered Histories

Despite the challenges of social media archiving, there are also ample opportunities for platform historiography, which we derive from various available, yet underutilised, archived web sources. Building on insights from a decade of interdisciplinary research on digital platforms,[32] social media's use cultures are, on the one hand, distributed across the multiple sides of platforms (multi-sidedness), and on the other hand, manifested across the different architectural levels of platforms (multi-layeredness). As multi-sided platforms, social media offer technologies, products, and services that create value by mediating the interactions between different user groups ('sides').[33] The leading social media platforms are inhabited by multiple user groups, including end users ('consumers'), businesses, advertisers and marketers, developers, media and publishers, content creators, politicians and governments, investors, non-profits, and more. These user groups are offered distinct user interfaces, information, and other resources, and each reveals different aspects of a platform's history. As such, a multi-sided perspective on platforms moves beyond the study of end users, user-generated content, and user interactions as the dominant analytical focus of historical social media studies. Instead, it foregrounds the other user groups that play an important role on the platform and influence platform evolution.[34]

Additionally, we should consider how platforms operate on different levels. There have been several proposals for multi-layered approaches to web history in general. Brügger has distinguished five interrelated analytical web strata: web elements, web pages, websites, web spheres, and the web as a whole.[35] As Brügger details, it is possible to analyse one or multiple strata, either in visible (e.g. rendered website) or invisible form (e.g. source code, web infrastructure).[36] Such a focus enables histories that move beyond the front-end of the web (e.g. presentation, interaction, use) to also include the back-end (e.g. data, infrastructure, hardware). Correspondingly, Paloque-Berges has suggested a (infra)structural approach to internet history by considering the formats, codes, and infrastructural aspects of Usenet newsgroup messages instead of their content.[37] We contend that social media platforms can also be

analysed at different levels to foreground anything from their technical standards to their data, databases, architectures, infrastructures, markets, and the app ecosystems that emerge around them, highlighting how platforms operate on different levels of scale. We argue that multi-sided and multi-layered analyses of platforms are of critical importance to derive situated accounts of the political economy of platforms and their operations.[38]

The multi-sidedness and multi-layeredness of platforms suggest that there are many kinds of histories, about different aspects of platforms, that have yet to be written. Therefore, we propose platform historiography as a form of web historiography that seriously accounts for the characteristics of platforms to consider how social media have emerged and operated as platforms in the past, and how their platformness may have changed over the years. Such histories can account for the evolving nature and relations between social media and the stakeholder groups they have accommodated and thwarted.

The Materiality of 'Platforms'

In this section, we introduce a material perspective on social media as platforms that leave various material traces behind. Many of these materials have been preserved in web archives and offer unique research opportunities for platform historiography. We also detail a method to examine the availability of these archived sources before discussing some of the research opportunities they provide.

Material Circumstances and Traces

In his call for historical software studies, Kirschenbaum outlines that the material circumstances of software – including software platforms such as social media – leave material traces that we may use to recover their histories.[39] While these traces are sometimes lost, they are mostly just underutilised and forgotten. Therefore, Kirschenbaum argues that one task of researchers is the fashioning of 'documentary methods' for recognising and recovering the material traces of software.[40] The main challenge is finding the materials that can be used to understand the evolutionary processes of social media in the first place.[41] Fortunately, social media, as digital platforms and as (publicly traded) companies, routinely leave such material traces behind in their operations, revealing aspects of their production, performance, embedding, reception, and business goals amongst others. These traces include APIs, software development kits (SDKs), developer pages, reference documentation, changelogs, version histories, app development guides, best practices, debugging tools, and ad targeting fields and parameters. They also include business product pages, ad management and insights tools, webinars, partner programmes and directories, certifications and awards, training courses and learning resources, app review guidelines, and help centres. Finally, they include blog archives, technical reports, research publications, patent applications, developer conferences and meetings, partner summits, earnings releases, U.S. Securities and Exchange Commission (SEC) filings, court

documents and filings, GitHub repositories, Twitter posts, public statements, and technology blogs. Platforms provide these diverse materials to their user groups on different public pages and subdomains, such as developer and business subdomains. Thus, while social media content and use pose severe archiving challenges, we contend that many of these material traces are in fact archived and awaiting scholarly use.

A subset of these materials are of particular interest for platform historiography. In the field of information systems, the concept of 'platform boundary resources' describes the subset of materials that support, govern, and control third-party development.[42] Dal Bianco et al. distinguish two kinds of platform boundary resources (PBRs): 'technical PBRs', which expose and extend a platform's architecture to enable developers to build apps and services on top of platforms, and 'social PBRs', which serve to transfer knowledge about and coordinate this app development.[43] In a broader sense, PBRs serve as interfaces that mediate the technical and social relationships between platforms and third parties. For us, the concept provides a framework to determine how user groups other than end users become part of a platform in the first place and to understand the conditions and dynamics of a platform's expansion and embedding into other domains.[44] Therefore, these technical boundary resources serve as a particularly relevant entry point to determine a platform's users and to indicate how it operates on multiple levels.

Unlike a platform's sides, which can be determined from the names and structure of a platform's boundary resources (e.g. 'for Developers', 'for Business'), its layers cannot always be elicited directly from the platform's materials. For this purpose, we draw on existing concepts and frameworks developed in the same field (i.e. information systems) to conceive the multiple layers of digital platforms from a technical as well as socio-technical perspective.[45] From a technical perspective, 'platforms' are the extensible codebases of software-based systems that provide core functionality to a set of interoperable 'modules'. These 'modules' are add-ons that connect to a platform and add functionality, such as specific tools and third-party apps. The 'ecosystem' is the entire set of modules related to the platform, and the 'architecture' is the design or conceptual blueprint of the ecosystem that describes the relations between a platform and its connected apps and services. Finally, 'interfaces' are the specifications and rules of interaction and exchange between platforms and modules, such as for data access and user interaction. Yet, as we argue in the following sections, these concepts are both technical and social as there are socio-technical processes and organisational structures associated with them. When platforms attract business developers to build complimentary apps and services, they form a technical ecosystem of connected apps and an organisational ecosystem of connected organisations.[46]

Source Availability in Web Archives

In Table A-2, the key user groups of the twenty most popular social media platforms worldwide are listed.[47] We took inventory of the main materials provided to these groups and considered the many kinds of histories afforded by these materials, only some of which we elaborate on within this article. Platform historians can use any (combination) of these

materials to write histories about specific aspects of platforms, provided that the relevant materials are preserved.

To determine whether or not these materials are preserved, and where to locate them, we conducted an exploratory analysis of the availability and accessibility of relevant archived web sources. We first collected the URLs pointing to a sample of relevant resources for each specific user group (e.g. facebook.com, developers.facebook.com, facebook.com/business, investor.fb.com). We then used the Memento Time Travel service to check the availability of each URL in our corpus against a list of international web archives that support the Memento protocol.[48] This protocol 'is the first web archiving API, enabling aggregation of access to disparate web archives.'[49] The service returns a structured list of available 'Mementos', an archived version of an original resource, for all archives that actually hold the URL.[50] To programmatically retrieve the availability of all URLs in our corpus, we employed the open source tool MemGator.[51]

We analysed the source availability along three criteria or dimensions: the volume of availability (the total number of Mementos held), the depth of availability (the number of days, months, or years between the first and last Mementos), and the breadth of availability (the number of archives holding Mementos). The first two criteria determine the amount of available material and the possible granularity or resolution of historical analysis. The third criterion, critically, enables researchers to triangulate historical sources and 'reconstruct' past states of platforms when certain pieces are missing, for example. Figure 1 presents an overview of the source availability for social media platforms overall (cumulative, 1a) as well as for specific user groups (small multiples, 1b). Each scatter point in the diagram represents one source type, and the larger the point, the larger its volume of availability. Sources positioned closer to the right have been archived more widely, and sources closer to the top have been archived for longer time periods.

Although all of the leading social media platforms are archived on multiple 'sides' and contain archived materials for their different user groups, we found significant differences in terms of their presence in web archives. First, some archives are more suitable for platform historiography than others due to their social media archiving strategies, including source selection and crawler depth. The Internet Archive Wayback Machine is by far the most inclusive web archive (holding 64.70% of all captures; 5.89% uniques), followed by Archive-It, Bibliotheca Alexandrina Web Archive, Stanford Web Archive, archive.today, UK Web Archive, Icelandic Web Archive, Arquivo.pt, Library of Congress, Canadian Archive, and perma.cc. For reconstruction purposes, however, every capture counts, and roughly two-thirds of the archives hold relevant sources from our corpus. Second, some social media platforms are more comprehensively archived than others, which means that not all of them are equally suitable for historical platform studies. While some platforms certainly resist archiving more than others, their limited availability is primarily due to a lack of archiving inclusion rather than of resistance, especially in the case of non-Western platforms. Overall, Twitter, Facebook, YouTube, Pinterest, Reddit, and Instagram are the most suitable for historical platform research, since they are most broadly archived on average. The Chinese Douyin (Tik Tok) and WeChat are the least suitable in terms of overall source availability (less than 5,000 captures). Finally, some aspects (and by implication, user

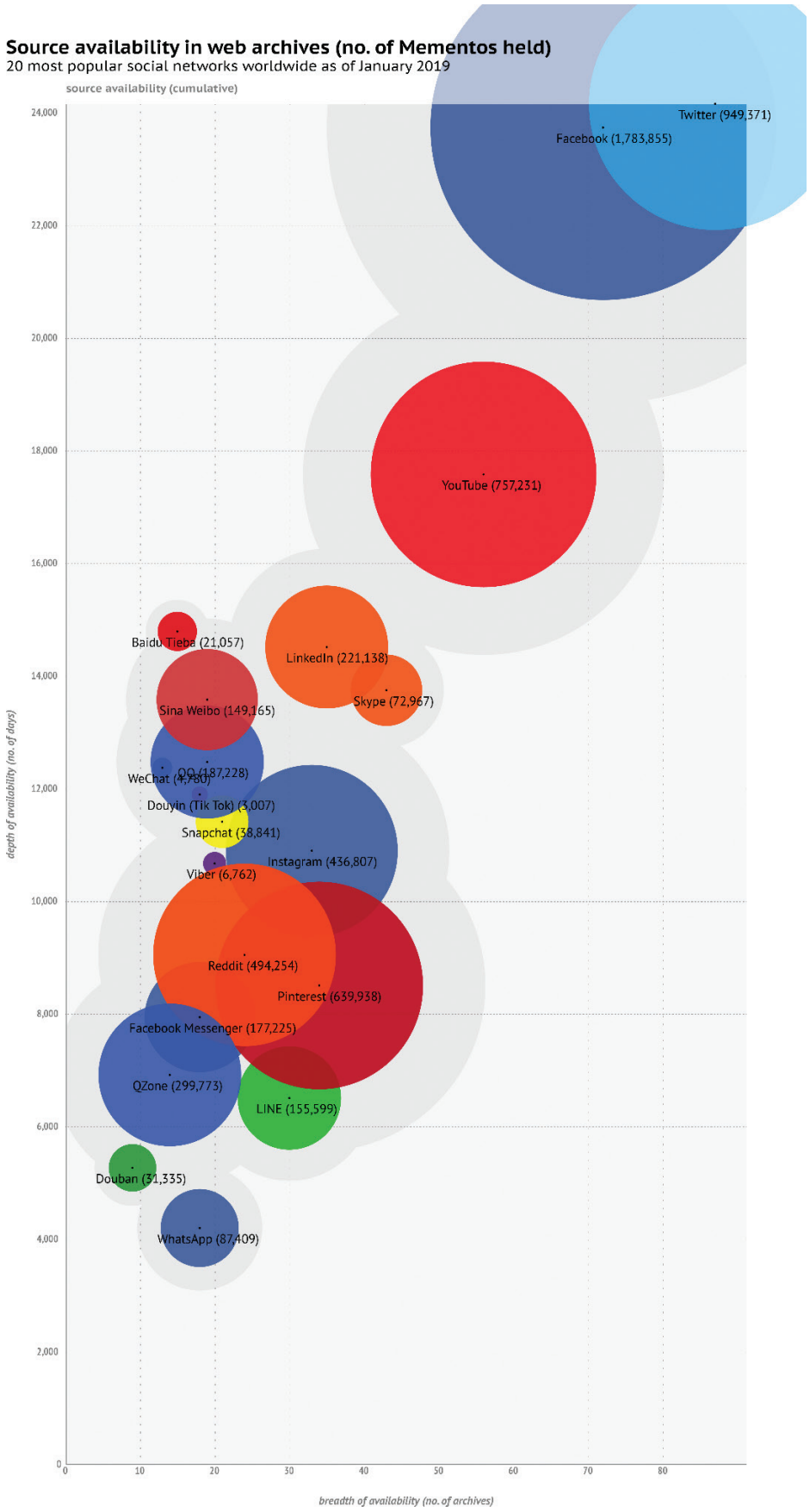

Figure 1a. Availability of archived web sources (N = 110 URLs) in Memento-compatible web archives (N = 21): cumulative. X-axis: breadth of availability (no. of archives); Y-axis: depth of availability (no. of days); size: by no. of Mementos held; colour-coding: by platform owner (brand colour).

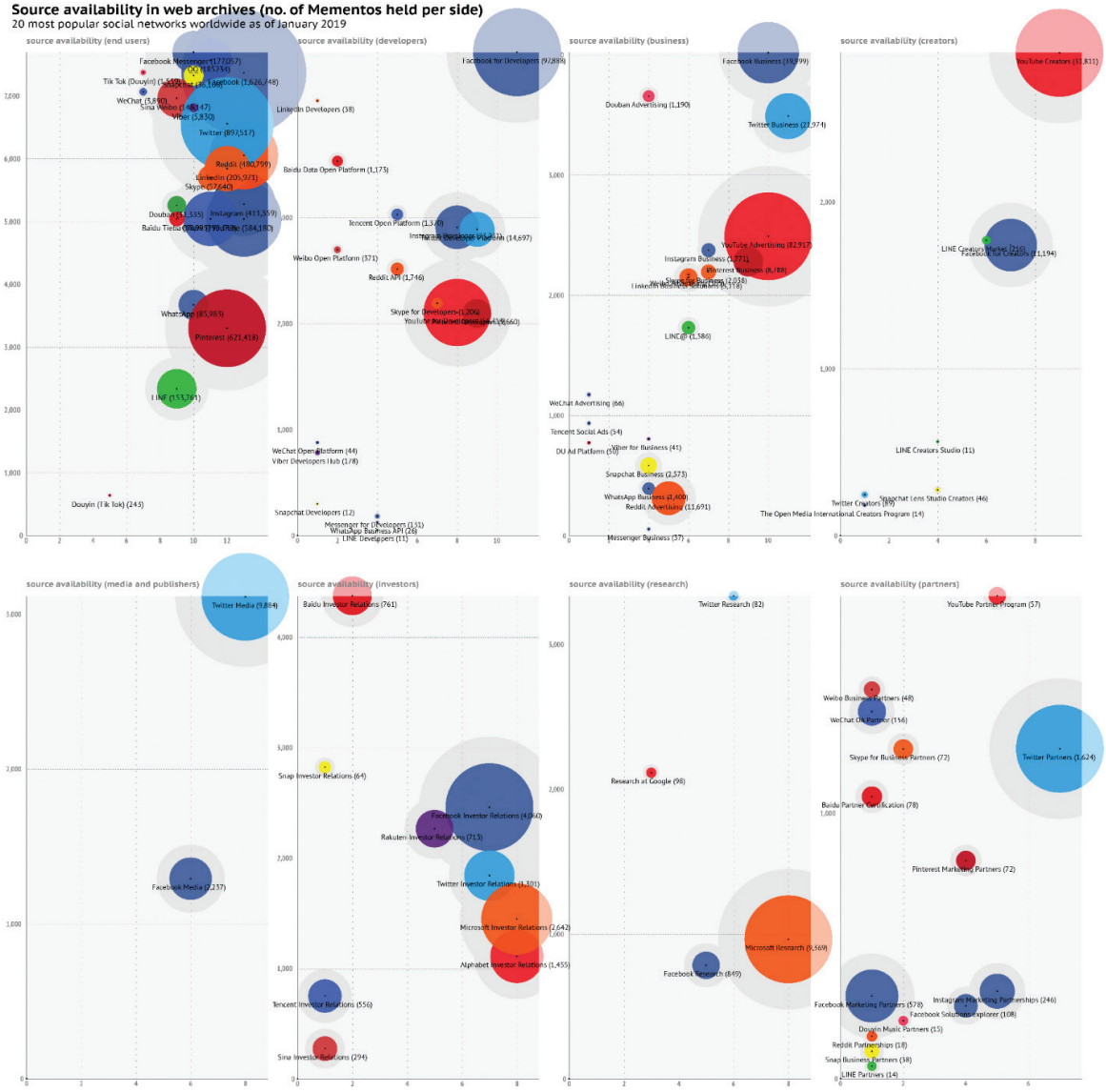

Figure 1b. Availability of archived web sources (N = 110 URLs) in Memento-compatible web archives (N = 21): by user group (1b). X-axis: breadth of availability (no. of archives); Y-axis: depth of availability (no. of days); size: by no. of Mementos held; colour-coding: by platform owner (brand colour).

groups or 'sides') of social media platforms are better represented in these web archives than others. Resources for end users are by far the best represented (92.75% of all captures) because they are hosted on the main domains (e.g. facebook.com), followed by resources for developers and business.

In the following sections, we discuss two sets of entry points for 'platform' historiography based on the available archived web sources: first, from the perspective of developers; second, from the perspectives of businesses and developers. Sources for both sides are very well-archived in web archives. To illustrate this, we draw on examples and case studies from our own previous (collaborative) historical research on Facebook and Twitter, which demonstrated the value of the proposed entry points and available archived web sources.

Entry Points for Developer-Side Histories

As we have noted, there are many research opportunities for platform historiography, depending on the researcher's entry point to a platform.[52] The first set of entry points that we discuss originates from social media platforms' resources for developer users. These include a platform's technical and social boundary resources, such as APIs and reference documentation, and other materials (Table A-2). These materials afford a wide range of platform histories, including histories of API-based data sharing, data strategy and interoperability, programmability, data-structuring practices and techniques, platform governance and control, and platform-developer relations more generally. We focus our discussion of research opportunities on the levels of architecture, interface, module, and standard.

Architecture-Level Histories

On the level of platform architecture, we can take stock of the breadth of official and unofficial APIs and SDKs as documented on developer pages. Most social media platforms provide a set of core platform APIs and SDKs that expose a platform's design and which can be used to interact with the platform and retrieve data (e.g. Facebook 'Graph API', Twitter 'Search API'). These core elements are usually versioned and remain available and unchanged for longer periods of time, while extended APIs and SDKs are subject to more rapid changes and can expire much sooner. As we previously demonstrated, by tracing platforms' core and extended elements, we gain insights into which elements platforms consider key aspects of their programmability as well as the pace of their development.[53] Specific APIs and SDKs and their documentation are commonly packaged into 'developer products', which foreground the use of specific data and functionality to developers (e.g. Facebook 'Pages API' and 'Live Video API'). These materials provide insights into when and how platforms turn towards new markets such as advertising, mobile, virtual reality, gaming, video, automated messaging, and artificial intelligence, which provide the necessary means to historicise the dominant market positions of platforms and their infrastructural presence in these different domains. Furthermore, SDKs can show which application development practices for specific platforms or software packages have been supported over time, such as Nokia, Blackberry, Android, and iOS (mobile platforms), PHP and JavaScript (programming languages), and Unity (game engine). Additional developer sources such as platform policies and app review requirements can be used to trace the evolving conditions for app development and the terms of use for APIs and SDKs, enabling histories of how platforms govern their developer communities.[54]

We can also trace the evolution of each API and SDK as they were launched, changed, and deprecated. In our previous research on Facebook, the volume and depth of the archived sources enabled us to trace how and when the platform expanded from a social networking site into a development platform and then into an expansive advertising and marketing development platform. We documented the platform's incremental changes and observed how Facebook's business platform arose from its development platform within a short period of time and how it

advanced its advertising and marketing technology tools and products with the help of its partners.[55] To further contextualise a platform's motivations and strategies behind particular technical and business developments, we can complement the use of developer pages and reference documentation with blog posts and archives, press releases, and annual reports.

Moreover, we can use these materials to examine how the contemporary infrastructural presence of social media platforms has evolved. Platforms regularly launch new features, products, and services, change and deprecate existing ones, and eventually remove them. Most platforms document versioned changes to their core elements in 'changelog' archives, including breaking changes that may affect the functioning of services and apps built on top of platforms if developers do not implement announced changes. Due to such changes, social media platforms and their extended infrastructural presence can become 'legacy systems' – or systems of the past that still exist or partially operate in un-updated apps and services. Therefore, this perspective foregrounds a platform's continued efforts and practices of maintenance and repair to provide insights into the legacies of past states of social media and their use cultures.

Tracing the evolution of specific APIs and SDKs further affords histories of a platform's changing programmability, which remains an essential characteristic of platforms as extensible codebases. Programmability is what enables a platform to be extended and embedded into other domains. As such, we have traced how the conditions for platform extensions and embeddings have changed.[56] Additionally, we can determine a platform's level of support – or the lack thereof – for developing with certain programming languages and in specific areas like artificial intelligence and machine learning. Beyond histories of specific computational objects and architectures, these materials enable us to trace processes of platformisation and infrastructuralisation.[57] That is, we can observe how changes related to programmability – as an essential characteristic of platforms – in turn change the nature and roles of social media platforms.

Interface, Module, and Standard-Level Histories

On the level of APIs as a specific type of interface between platforms and developers, we can trace their evolving specifications using API reference documentation. This documentation describes individual objects (e.g. a user, photo, or comment), the relations between them, the available data fields and parameters for each object (e.g. a user's birthday or hometown), and how to access each object using API endpoints. These details mark the technical conditions for programmatic data access, data exchange, and interaction with a platform. Moreover, since the reference documentation describes a platform's available objects, relations, and fields, we can employ these materials to write histories of a platform's changing data strategies, as demonstrated by media scholars.[58] We can use them to reconstruct the interplay between recent data scandals around social media and the changes made to their APIs and development platforms. More generally, we can examine how platforms have changed their data governance strategies in response to issues of data abuse (e.g. Cambridge Analytics), evolving legal and policy requirements (e.g. GDPR, CCPR), and data-based revenue models (e.g. API data access tiers).

For example, changing revenue models led Twitter to introduce premium and enterprise-level APIs, which turned historical access to Twitter data into a paid-for service. The new APIs target businesses who use historical data for social media-based prediction. While a seven-day 'look back window' is still free, anything else comes at a premium, thus resulting in 'premium histories'.

On the level of platform modules, we can trace how platforms allow themselves to be embedded in external websites and apps through various widgets, social plugins, and web 'embeds'. Developers can use these resources to implement platform data functionality into their own websites and apps, enabling users to authenticate and log into an app, import personal details and settings, and to like, share, and save content to a platform. By tracing how these popular modules have evolved, we can examine their roles in longer-term processes such as the platformisation of the web and mobile ecosystem.[59]

On the level of protocols and standards, which tie in with platform architecture, we can trace the protocols and standards launched by leading platforms over the past decades. In the case of Facebook, we traced four standards development efforts in an exploratory case study. We noted a gradual change from platform-specific protocols and standards, such as Facebook Query Language (FQL) and Facebook Markup Language (FBML), which had both been derived from open web standards (i.e. SQL, HTML), to the introduction of industry-wide protocols and standards that extend beyond Facebook, such as its Open Graph protocol and GraphQL, a new query language for APIs. These efforts tell two interrelated stories about the platform's evolution: about its growth as a development and business platform and about its crucial pivot from web-based to web-and-app-based development.

In particular, in our exploratory case study we noted that since Facebook introduced the idea of a social graph to map the relations between users in their database, its standards development efforts have become increasingly graph-based with the introduction of the Graph API, the Open Graph protocol, GraphQL, and Graph Academy. These technical developments reflect Facebook's vision to model the world in graphs: 'It's Graphs All the Way Down'. As others have noted, such data-structuring practices impact how information is structured and organised to become 'algorithmically recognisable' and 'platform ready'.[60] For example, FQL and GraphQL handle data transactions differently due to their unique architectural styles: GraphQL not only offers a more computationally efficient data-fetching method, but also represents a richer sociological model of a data structure centred around (social) graphs. That is, the new data model reflects both technological as well as business needs.[61] As such, it demonstrates how we can trace changes in specifications of protocols and standards to characterise the applications and logics of data models and data ontologies, which relate to issues of data justice.[62] Complementary sources such as open source software repositories, research publications, and technical reports may provide further insights into the design rationales of these architecture styles and standards.[63]

In conclusion, the developer pages, and especially reference documentation, provide significant research opportunities for historical platform studies from a technical perspective and beyond, as they provide important entry points to examine concerns around data privacy and

data justice. Additionally, they highlight the interplay between the multiple sides and layers of platforms, providing insights into their operative scales and scopes: from the level of data types and formats to API specifications and design rules to architecture design to the technical conditions of large-scale platform ecosystems that have been gradually woven into the technical fabric of the web and mobile ecosystem through APIs and SDKs and their modules.

Entry Points for Business-Side Histories

The second set of entry points originates from social media platform's resources for businesses and business developers. Their resources include business and advertising tools, products, and services, news and help pages, advertising and marketing APIs and their reference documentation, training courses, and partner pages (Table A-2). These materials similarly afford a wide range of platform histories, including histories of specific tools and products, ads and targeting audiences, platform growth and market embeddings, and platform-business relations more generally. Here, we focus on the levels of architecture, interface, ecosystem, and module.

Product Architecture-Level Histories

From the perspective of product architecture, we can trace how a platform's business tools, products, and services have evolved, as well as social media-based business more generally. At first, platforms' business pages generally provided basic information about a platform's value and contained best practices, guides, and case studies (see Figure 2), followed by offering advertising and analytics tools, broader suites of marketing tools and products, and finally technical specifications related to API-based marketing products to enable large programmatic ad campaigns. By tracing the evolution of core platform business tools and products in our previous work, we examined how different social media-based business practices have evolved and how social data turned into a dominant revenue model for platforms.[64] In particular, core tools and products can provide insights into the evolution of the perceived value and utility of social media for businesses and of social data.

Since advertising is the main revenue source for leading social media platforms, the specifications of their ad formats are relevant too. Using business pages, we can determine supported ad formats, measurements, and audience targeting capabilities as they have evolved. Such accounts can contribute to political economic histories of online advertising.[65] We can trace how Twitter and Facebook's advertising products initially revolved around basic platform components, such as promoted accounts, tweets, trends, pages, and sponsored groups and how they evolved into more complex products around GIFs, video, live video, mobile apps, while also introducing new platform-specific content formats, such as Facebook 'Stories Ads', Twitter 'Conversational Ads', and Snapchat 'Sponsored Lens'. Accordingly, evolving ad formats reflect the specific kinds of content and practices supported by platforms.

Furthermore, some platforms have evolved into full-fledged, market-leading ad business platforms. For example, Facebook and Twitter were both key drivers behind the rise and

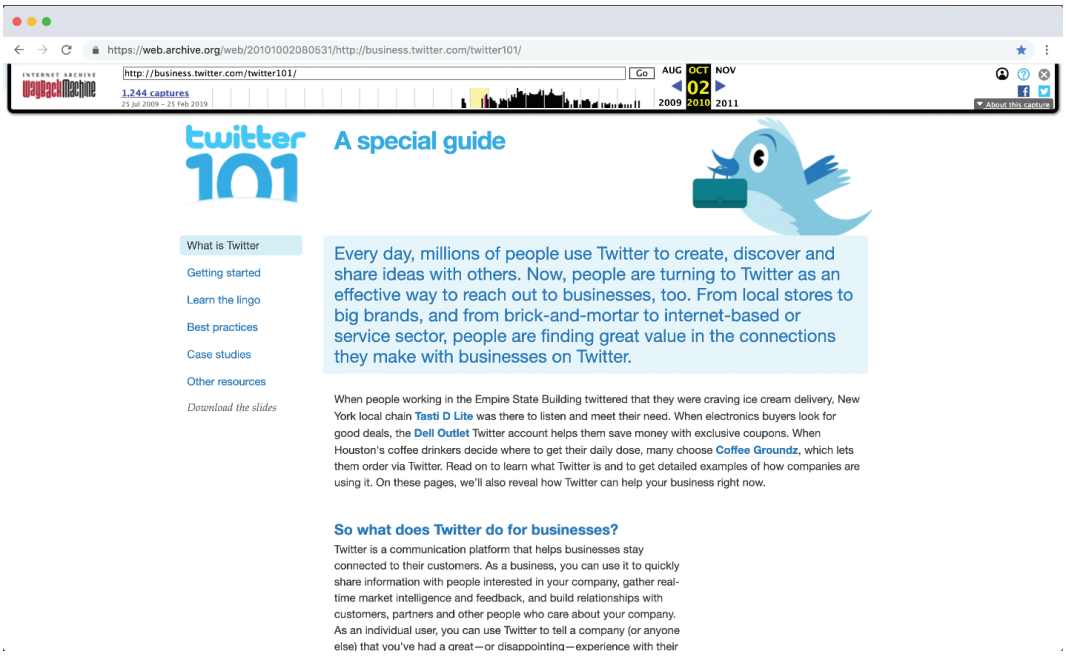

Figure 2. Screen capture of business.twitter.com (October 2, 2010). Source: Internet Archive Wayback Machine.

legitimation of social and mobile advertising and marketing, which have become increasingly data-driven and automated over the past decade. At the same time, Instagram and Snapchat drove influencer marketing. By now, social media's advertising and marketing technologies are no longer confined to their platforms, with rapid expansion and consolidation into a worldwide technology infrastructure that supports targeting people both on and off their platforms.[66] We can trace the emergence of Facebook's Audience Network for targeting people on third-party websites and apps using Facebook data or Twitter's MoPub, which has become a key mobile advertising platform that connects buyers and sellers of mobile ads. Using their documentation, we can examine how they have become connected to the larger marketing and advertising ecosystem, who they are connected with (to examine their overall reach), and what their targeting capabilities are outside of their own platforms. In short, by tracing the evolution of these products and their materials, we have previously demonstrated how they can be used to reconstruct both platform-specific and industry-wide advertising practices and developments.[67]

Interface-Levels Histories

From the perspective of the interface, we can trace the evolving structure and specifications of business and advertising-related interfaces.[68] Most social media platforms offer two kinds of interfaces to their business and advertising users: the self-service tools aimed at small and medium-sized businesses (SMBs) and the APIs aimed at large agencies and software companies for running multiple big ad campaigns. The self-service advertising tools allow users to create and run ad campaigns and track their performance. Users can create an audience, choose where to run the ads, set a budget, and pick an ad format. Audiences can be specified in different ways, such as by age, location, device, gender, demographics, interests, behaviours, or by

uploading their own customer information such as email addresses or phone numbers. Since all of these options and settings are documented in detail on business product and help pages, we can trace the evolution of social advertising and targeting capabilities available to small business users. The Yugoslavian-based research and data investigation lab SHARE LAB previously mapped the targeting options and techniques available to Facebook advertisers and visualised the data fields underpinning the platform's complex profiling and targeting processes.[69] Repeating such studies over time would provide further insights into how these processes have evolved.

We can also trace evolving specifications of business APIs and SDKs that allow businesses to programmatically create and manage large-scale ad campaigns from within their own software systems. Such integrations have enabled the rise of massive, interconnected digital advertising infrastructures. These programmatic interfaces are similar to the self-serving interfaces for creating, running, and managing ads but are aimed at business developers to automate and scale these processes. We can zoom into individual advertising and marketing APIs to examine evolving options and data fields for building and targeting audiences and bidding strategies. Because social media platforms have collected large amounts of data about their users, they can target very specific audiences. However, since audience segmentation is largely an automated process, this has controversially led to automatically generated targeting categories such as 'jew haters' and people 'interested in Nazis'.[70] Additionally, large platforms have often partnered with third-party data providers to expand their targeting capabilities by offering 'partner categories' to target specific audiences, such as 'homeowners' or 'car buyers'. In response to ad targeting misuse and privacy concerns, some platforms have terminated such audience data partnerships and have eliminated specific audiences from their audience segmentation, including more than 5,000 in the case of Facebook alone.[71] Using archived business developer pages, we previously traced the dynamics of the removal of specific audience targeting fields, including the 'ethnic_affinity' and 'interested_in' fields in Facebook's Marketing API (see Figure 3), which were both disabled due to algorithmic discrimination concerns. Additionally, we can consult complementary sources such as research publications, technical reports, and public patent applications to examine the technical specifications and implementations of algorithmic classification techniques.[72] For example, a patent assigned to Facebook describes a technique 'to predict the socio-economic group of users'.[73] We can use such materials to trace the evolving techniques and practices of classification at the level of the interface. Drawing on existing social and critical theory, this enables us to expand our knowledge about how such classification practices become encoded in databases and data fields, as well as the consequences of classification.[74]

Platform Architecture, Ecosystem, and Module-Level Histories

From the perspective of platform architecture, we can take stock of the APIs and SDKs that are aimed specifically at business developers, providing insights into core business tools and products. In our previous research on Facebook's APIs, we found that these are increasingly geared towards the management of programmatic ad campaigns (e.g. Facebook's 'Business

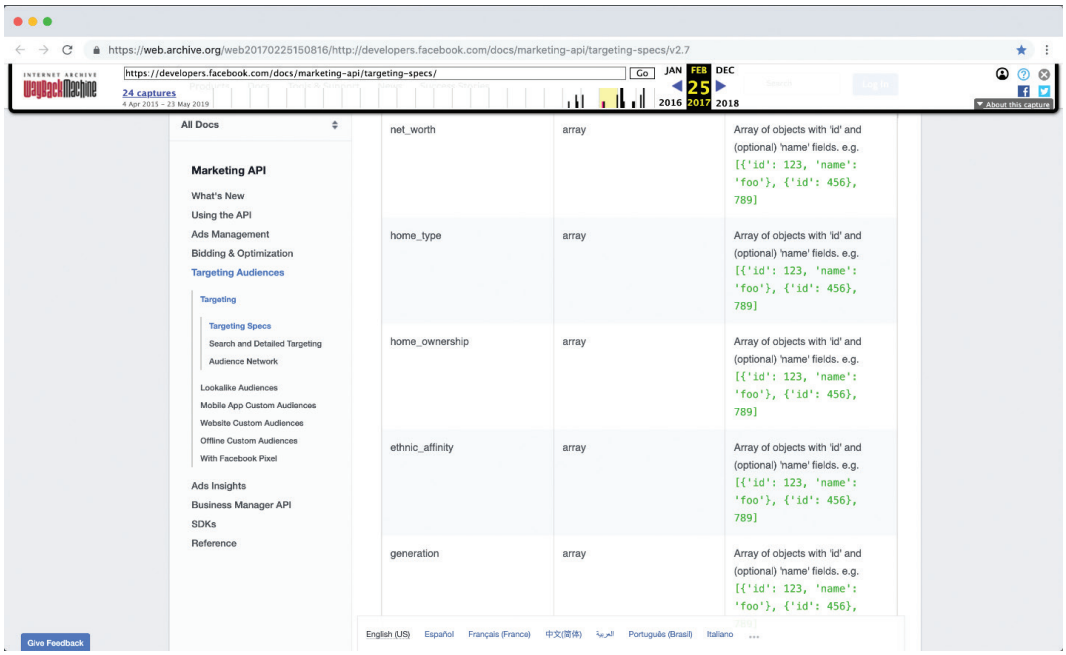

Figure 3. Screen capture of Facebook Marketing API audience targeting specifications at developers.facebook.com/docs/marketing-api/targeting-specs/ (February 25, 2017), including the 'ethnic_affinity' field (now disabled). Source: Internet Archive Wayback Machine.

Manager API', 'Pages API', 'Marketing API'). Although access to some of these resources requires approved businesses, we can nonetheless employ their publicly accessible documentation to inventory these tools and products as they evolved. Moreover, we can examine how these tools and products are integrated with the platform's overall architecture. In recent years, Facebook has consolidated its disparate resources – including the Marketing APIs, Instagram API, and Messenger Platform – into a single unified Graph API. This streamlining connects the platform's different products at their core, thereby facilitating cross-product interactions, such as targeting Facebook users on Instagram and Messenger.[75]

From the perspectives of platform and business ecosystems, we previously traced how social media platforms engage in partnerships with other business users.[76] Platforms' key business products such as advertising and marketing APIs are typically restricted to a select group of certified marketing partners. These official partners help other business users with their advertising and marketing needs by providing additional tools and services that are built on top of social media platforms. They represent a worldwide community of influential firms and organisations – many of them are Fortune 500 holding giants – and cover industries around the world including financial services, retail and e-commerce, gaming, telecommunications, technology, automotive, and restaurants. These partners have been crucial in embedding social media in these industries and the digital marketing industry at large, which has accelerated their growth and influence to power. In our previous work, we demonstrated the value of these archived partner materials to trace the evolution of specific business partnerships, their details, the solutions offered by partners, the rise and fall of partnerships, and evolving partner communities.[77] Thus, we can employ these materials to examine the interplay between processes of platformisation and infrastructuralisation to examine a platform's expansion and embedding

into other domains and environments (e.g. organisations, markets, industries, jurisdictions) through API-based software integrations and business partnerships.[78] As such, these materials can reveal how technical and organisational ecosystems emerge around platforms. Tracing these developments provides insights into the evolving nature of the digital marketing and data economy more generally.

From the perspective of platform modules, we can trace core business products such as analytics tools. These include tracking technologies such as Facebook Pixel or Twitter Pixel, which business users can implement into their websites to collect data for their ad campaigns. These technologies are usually integrated with other tools and products to enable business users to measure, optimise, and retarget ads across the platform's properties, as well as their own websites and even external websites and apps.

In conclusion, business pages and business developer pages similarly provide significant research opportunities for historical platform studies with insights into the technical, social, commercial, and organisational dimensions of a platform's infrastructural presence in society. As such, these examples further stress the need to consider the multiple sides and layers of platforms.

Conclusion: The Future of Platform Historiography

We proposed a methodological outlook for historical platform studies based on prior research that demonstrated the value of archived web sources. Building on recent theoretical reflections on web historiography and digital platforms,[79] we discussed the methodological opportunities and challenges of social media *'platform' historiography*. As we have argued, there are many relevant entry points and significant research opportunities for platform historiography. These entry points – and the unique materials they offer on their multiple sides and layers – provide insights into the histories of social media as platforms, their evolving users and uses, platform architectures, data strategies, embeddedness in other domains, infrastructural presence, and even industry-wide practices and trends. In particular, these critical histories can contextualise the international media coverage and criticism of social media related to surveillance, data breaches, data and privacy scandals, third-party app development, media manipulation, political advertising and targeting, automation and bots, and more.

Social networking, and in particular the platform model, has gained a dominant presence since the mid-2000s, accruing billions of end users and millions of developers, advertisers, and businesses worldwide. In less than two decades, social media has become an integral part of social, cultural, and economic life. Although it is important and necessary to understand the historical impact of social media in terms of cultural memory and digital heritage by focusing on end users (e.g. user-generated content, user interactions), this is not sufficient to capture the infrastructural presence and impact of social media on past, present, and future societies (e.g. technological and business impact, datafication, surveillance, privacy practices). By shifting the focus from end users to platforms' multiple user groups, we are able to study the platform not

only as a social network but also as a technical artefact and business organisation.[80] In Table A-2, we have outlined the many kinds of platform histories such a focus affords.

Core to our proposal for platform historiography is our call to web and platform historians as well as to archiving practitioners to foreground the unique characteristics of social media as digital platforms that actively mediate interactions between different user groups, rather than considering them as mere social networking or user-generated content services. In particular, foregrounding the multiple sides and layers of platforms is crucial to understand their present operational scale and scope as well as their infrastructural presence and their influence and power in various domains. Thus, through this article, we hope to increase the prominence of platform historiography in the field as well as the practice of web history and archiving.

Platform historiography, perhaps more than web historiography in general, requires its object of study to be 'reconstructed' by the historian. This is because platforms are distributed entities with multiple sides and layers that cannot be preserved in their entirety or original form. As such, both a platform's multi-sidedness and multi-layeredness must be reconstituted. Yet, as Bruns and Weller have argued, a main challenge for platform history is to find the right materials.[81] Consequently, social media history and archiving have typically been discussed in terms of their specific issues and challenges as opposed to their research affordances. Here, we introduced a material perspective on social media as 'platforms' to illustrate how they inevitably leave behind a variety of material traces – as digital platforms and as companies – that should be of interest to current and future platform historians.

In our exploration of entry points for platform historiography, we uncovered many under-utilised archived web sources that have been preserved despite the challenges of social media archiving. We found that most of the leading social media platforms are very well archived beyond their end-user sides, although we noted significant differences in the volume, breadth, and depth of source availability in web archives. In particular, some web archives, social media platforms, and user groups are much better suited for historical platform studies than others due to their archiving, or lack thereof. Therefore, we recommend web archiving practitioners to not only archive the 'front pages' of social media and end-users' content but also the full breadth of platform-related subdomains and pages, including those that accommodate other user groups. Our inventory in Table A-2 could serve as a starting point for their crawlers.

The historiographical contributions of this article relate to three main fields of study. First, to the broader field of historical social media and platform studies, we offer a methodological outlook and entry points for developing historical studies of social media platforms with a variety of archived web sources. Our proposal to move beyond platform end-user and content histories, enabled by 'API-based research', has become increasingly relevant with growing restrictions on API access for researchers, raising the need to explore new sources and approaches that may be closer to digital fieldwork.[82] Our exploration of two sets of entry points demonstrates the platform histories that await to be written with readily available sources. We encourage platform scholars and historians to further explore these – and uncover other – relevant entry points that we scouted and to develop case studies to further advance platform historiography. Additionally, we provide an exploratory approach to systematically retrieve the availability of such sources

across international web archives. This enables web and platform historians to 'triangulate' and verify specific sources or source sets and to 'reconstruct' past states of platforms.[83] At the same time, it enables archiving practitioners to assess source redundancy, or the existence of additional source copies available from other web archives, as well as to identify and prevent gaps or absences in these archives. Our outlook on 'platform' historiography could also extend beyond social media platforms alone. The web-based software platform is a prolific software form that can be found across various digital industries. These other kinds of digital platforms, which are generally less well-known and not part of academic debates in platform studies, include anything from demand-side and supply-side advertising platforms, data management platforms, customer data platforms, marketing clouds, analytics and engagement platforms, and more. Future ecosystem-level histories could focus on this entire ecosystem of connective platforms beyond social media platforms.[84]

Second, we offer contributions to the field and practice of web history, and especially web historiography. In contrast to website historiography, which has largely focused on the media content and networked structure of websites and pages,[85] platform historiography focuses on the multiple sides and layers of web-based platforms, which come with their own empirical historical opportunities. As we detailed elsewhere, the available archived web sources allow us to reconstruct the minute, incremental evolutionary trajectories of platforms in great detail as opposed to only in broad strokes.[86] Overall, the granularity of archived web sources is sufficient to operationalise empirical histories of platforms' evolutionary trajectories and dynamics. Moreover, we encourage web and platform historians to further explore the theoretical and practical implications of the differences between web and platform history and to develop comparative historical platform studies.

Finally, our contribution offers lessons for historical app studies. Nearly all of the most popular social networks allow third parties to develop apps and integrations on top of their platforms. These apps are modules that have been built with platform boundary resources to interact or exchange data with platforms. In short, these apps sit on top of the platform architecture. However, in some cases, the social networks themselves only exist – or exist primarily – as apps and not as platforms, such as with Douyin (Tik Tok) and WeChat. Although app archiving and app historiography come with their own specific issues and challenges[87] – some of which we argue are more aligned with software preservation and emulation challenges than with web archiving and reconstruction challenges – our multi-sided and multi-layered methodological outlook provides several unique entry points and opportunities for historical app studies. Similar to digital platforms, apps have material circumstances that leave material traces that could be of use to app historians. In other words, while these apps may not always be archived or preserved, some of their material traces may still be archived and preserved in one or multiple web archives. These materials enable app historians to recount stories of the development and functionality of apps or their data collection and sharing practices, for example. We, therefore, encourage web archiving practitioners and institutions to increase their focus on the systematic archiving of the breadth of materials of social media platforms and apps beyond their front pages and user-generated content. In other words, they should focus their archiving efforts

not only on user-generated content but also on the material traces of the medium, which then let us reconstruct those media forms and understand them within their contexts. Our hope is that this creates the conditions for reconstructing and writing histories of social media platforms and apps and for understanding their significance in the past, present, and future.

Acknowledgements

All authors contributed equally to this work. Thanks to the Memento community and Sawood Alam for their technical support. Thanks also to Andy Jackson for repairing the UK Web Archive's Memento endpoints.

This work was supported by the Netherlands Organisation for Scientific Research (NWO) under Grant 275-45-009; and the German Research Foundation (DFG) under Grant DFG-SFB-1187.

Notes

1. Gabriel J. X. Dance, Michael LaForgia, and Nicholas Confessore, "As Facebook Raised a Privacy Wall, It Carved an Opening for Tech Giants," *The New York Times*, December 18, 2018, sec. Technology, https://www.nytimes.com/2018/12/18/technology/facebook-privacy.html.
2. José van Dijck, *The Culture of Connectivity: A Critical History of Social Media* (New York, NY: Oxford University Press, 2013).
3. Sara Day Thomson, "Preserving Social Media," *DPC Technology Watch Reports* (United Kingdom: Digital Preservation Coalition, 2016), https://doi.org/10.7207/twr16-01.
4. Niels Brügger, *The Archived Web: Doing History in the Digital Age* (Cambridge, MA: MIT Press, 2018); Richard Rogers, "Periodizing Web Archiving: Biographical, Event-Based, National and Autobiographical Traditions," in *The SAGE Handbook of Web History*, ed. Niels Brügger and Ian Milligan (London: SAGE, 2018), 42-56.
5. Brügger, *The Archived Web*, 208-9; Anne Helmond, "The Platformization of the Web: Making Web Data Platform Ready," *Social Media + Society* 1, no. 2 (2015), doi:10.1177/2056305115603080.
6. Carolin Gerlitz and Anne Helmond, "The Like Economy: Social Buttons and the Data-Intensive Web," *New Media & Society* 15, no. 8 (2013): 1348-65, doi:10.1177/1461444812472322; Van Dijck, *The Culture of Connectivity*.
7. Brügger, *The Archived Web*; Niels Brügger, "Web Historiography and Internet Studies: Challenges and Perspectives," *New Media & Society* 15, no. 5 (2013): 752-764, doi:10.1177/1461444812462852.
8. Mark de Reuver, Carsten Sørensen, and Rahul C. Basole, "The Digital Platform: A Research Agenda," *Journal of Information Technology* 33, no. 2 (2018): 124-135, doi:10.1057/s41265-016-0033-3.
9. Annabelle Gawer, "Bridging Differing Perspectives on Technological Platforms: Toward an Integrative Framework," *Research Policy* 43, no. 7 (September 2014): 1239-49, doi:10.1016/j.respol.2014.03.006.
10. Tarleton Gillespie, "The Politics of 'Platforms,'" *New Media & Society* 12, no. 3 (2010): 347-64, doi:10.1177/1461444809342738; Van Dijck, *The Culture of Connectivity*.
11. Michael Dieter et al., "Multi-Situated App Studies: Methods and Propositions," *Social Media + Society*, n.d., doi:10.1177/2056305119846486.
12. Jean-Christophe Plantin et al., "Infrastructure Studies Meet Platform Studies in the Age of Google and Facebook," *New Media & Society* 20, no. 1 (2018): 293-310, doi:10.1177/1461444816661553; Helmond, "The Platformization of the Web"; Anne Helmond, David B. Nieborg, and Fernando N. van der Vlist, "Facebook's Evolution: Development of a Platform-as-Infrastructure," *Internet Histories: Digital Technology, Culture and Society* 3, no. 2 (2019): 123-46, doi:10.1080/24701475.2019.1593667; Panos Constantinides, Ola Henfridsson, and Geoffrey G. Parker, "Introduction – Platforms and Infrastructures in the Digital Age," *Information Systems Research* 29, no. 2 (May 21, 2018): 381-400, doi:10.1287/isre.2018.0794.

13. Brügger, "Web Historiography and Internet Studies"; Richard Rogers, "Doing Web History with the Internet Archive: Screencast Documentaries," *Internet Histories* 1, no. 2 (2017): 160-72, doi:10.1080/24701475.2017.1307542.
14. Brügger, *The Archived Web*; De Reuver, Sørensen, and Basole, "The Digital Platform"; Helmond, Nieborg, and Van der Vlist, "Facebook's Evolution."
15. Jane Winters, "Coda: Web Archives for Humanities Research – Some Reflections," in *The Web as History: Using Web Archives to Understand the Past and the Present*, ed. Niels Brügger and Ralph Schroeder (London: UCL Press, 2017), 238; Rogers, "Doing Web History."
16. Brügger, *The Archived Web*.
17. Ibid.
18. Ibid.
19. Brügger, "Web Historiography and Internet Studies"; Brügger, *The Archived Web*; Richard Rogers, *Digital Methods* (Cambridge, MA: MIT Press, 2013); Rogers, "Periodizing Web Archiving."
20. Esther J. T. Weltevrede, "Repurposing Digital Methods: The Research Affordances of Platforms and Engines" (Ph.D., University of Amsterdam, 2016), https://wiki.digitalmethods.net/Dmi/RepurposingDigitalMethods.
21. Brügger, *The Archived Web*; Thomson, "Preserving Social Media"; Justin Littman et al., "API-Based Social Media Collecting as a Form of Web Archiving," *International Journal on Digital Libraries* 19, no. 1 (2018): 21-38, doi:10.1007/s00799-016-0201-7; Rogers, "Periodizing Web Archiving."
22. Brügger, *The Archived Web*, 80-81.
23. Archive-It, "Known Web Archiving Challenges," *Archive-It Help Center*, August 21, 2018, http://support.archive-it.org/hc/en-us/articles/209637043-Known-Web-Archiving-Challenges.
24. Brügger, *The Archived Web*.
25. Ibid.; Thomson, "Preserving Social Media"; Littman et al., "API-Based Social Media Collecting"; Stine Lomborg, "Researching Communicative Practice: Web Archiving in Qualitative Social Media Research," *Journal of Technology in Human Services* 30, no. 3-4 (2012): 219-231.
26. Thomson, "Preserving Social Media," 4.
27. Ibid., 2.
28. Thomson, "Preserving Social Media"; Littman et al., "API-Based Social Media Collecting."
29. Ganaele Langlois and Greg Elmer, "The Research Politics of Social Media Platforms," *Culture Machine* 14 (2013), http://culturemachine.net/index.php/cm/article/view/505/531.
30. Tommaso Venturini and Richard Rogers, "'API-Based Research' or How Can Digital Sociology and Journalism Studies Learn from the Facebook and Cambridge Analytica Data Breach," *Digital Journalism*, 2019, doi:10.1080/21670811.2019.1591927.
31. Brügger, *The Archived Web*; Axel Bruns and Katrin Weller, "Twitter As a First Draft of the Present: And the Challenges of Preserving It for the Future," in *Proceedings of the 8th ACM Conference on Web Science*, WebSci '16 (New York, NY: ACM, 2016), 183-189, doi:10.1145/2908131.2908174; Lomborg, "Researching Communicative Practice."
32. Ian Bogost and Nick Montfort, "Platform Studies: Frequently Questioned Answers," in *Proceedings of the Digital Arts and Culture Conference* (DAC09: After Media Embodiment and Context, University of California, Irvine, 2009), http://escholarship.org/uc/item/01r0k9br.pdf; Gillespie, "The Politics of 'Platforms'"; Helmond, "The Platformization of the Web"; Adrian Mackenzie, "From API to AI: Platforms and Their Opacities," *Information, Communication & Society*, June 11, 2018, 1-18, doi:10.1080/1369118X.2018.1476569; De Reuver, Sørensen, and Basole, "The Digital Platform."
33. De Reuver, Sørensen, and Basole, "The Digital Platform"; Jean-Charles Rochet and Jean Tirole, "Platform Competition in Two-Sided Markets," *Journal of the European Economic Association* 1, no. 4 (2003): 990-1029, doi:10.1162/154247603322493212.
34. Helmond, Nieborg, and Van der Vlist, "Facebook's Evolution."
35. Brügger, "Web Historiography and Internet Studies"; Brügger, *The Archived Web*.
36. Brügger, *The Archived Web*, 40.
37. Camille Paloque-Berges, "Usenet as a Web Archive: Multi-Layered Archives of Computer-Mediated-Communication," in *Web 25: Histories from the First 25 Years of the World Wide Web*, ed. Niels Brügger (New York, NY: Peter Lang, 2017), 241.
38. Dieter et al., "Multi-Situated App Studies: Methods and Propositions"; Helmond, Nieborg, and Van der Vlist, "Facebook's Evolution."

39. Matthew G. Kirschenbaum, "Virtuality and VRML: Software Studies after Manovich," *Electronic Book Review*, 2003, http://www.electronicbookreview.com/thread/technocapitalism/morememory.
40. Kirschenbaum, "Virtuality and VRML."
41. Bruns and Weller, "Twitter As a First Draft of the Present," 186.
42. Ahmad Ghazawneh and Ola Henfridsson, "Balancing Platform Control and External Contribution in Third-Party Development: The Boundary Resources Model," *Information Systems Journal* 23, no. 2 (2013): 173-92, doi:10.1111/j.1365-2575.2012.00406.x.
43. Vittorio Dal Bianco et al., "The Role of Platform Boundary Resources in Software Ecosystems: A Case Study," in *2014 IEEE/IFIP Conference on Software Architecture* (New York, NY: IEEE, 2014), 11-20, doi:10.1109/WICSA.2014.41.
44. Helmond, Nieborg, and Van der Vlist, "Facebook's Evolution"; David B. Nieborg and Anne Helmond, "The Political Economy of Facebook's Platformization in the Mobile Ecosystem: Facebook Messenger as a Platform Instance," *Media, Culture & Society* 41, no. 2 (2019): 196-218, doi:10.1177/0163443718818384.
45. De Reuver, Sørensen, and Basole, "The Digital Platform"; Amrit Tiwana, Benn Konsynski, and Ashley A. Bush, "Research Commentary – Platform Evolution: Coevolution of Platform Architecture, Governance, and Environmental Dynamics," *Information Systems Research* 21, no. 4 (November 18, 2010): 675-87, doi:10.1287/isre.1100.0323.
46. De Reuver, Sørensen, and Basole, "The Digital Platform."
47. Statista, "Global Social Media Ranking 2018 | Statistic," *Statista*, February 1, 2019, https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/.
48. Memento, "Time Travel," accessed February 22, 2019, http://timetravel.mementoweb.org/.
49. Michael L. Nelson and Herbert van de Sompel, "Adding the Dimension of Time to HTTP," in *The SAGE Handbook of Web History*, ed. Niels Brügger and Ian Milligan (London: SAGE, 2018), 189-214.
50. Memento, "About Time Travel," September 22, 2016, http://timetravel.mementoweb.org/about/; Nelson and van de Sompel, "Adding the Dimension of Time to HTTP."
51. Sawood Alam and Michael L. Nelson, "MemGator-A Portable Concurrent Memento Aggregator: Cross-Platform CLI and Server Binaries in Go," in *Proceedings of the 16th ACM/IEEE-CS on Joint Conference on Digital Libraries* (ACM, 2016), 243-244.
52. Dieter et al., "Multi-Situated App Studies."
53. Helmond, Nieborg, and Van der Vlist, "Facebook's Evolution."
54. Anne Helmond et al., "Apps and Their Stories: Volatility, Diversity, Policy," *Digital Methods Initiative Wiki*, February 4, 2019, https://wiki.digitalmethods.net/Dmi/WinterSchool2019AppsStories.
55. Helmond, Nieborg, and Van der Vlist, "Facebook's Evolution."
56. Nieborg and Helmond, "The Political Economy of Facebook's Platformization in the Mobile Ecosystem"; Helmond, Nieborg, and Van der Vlist, "Facebook's Evolution."
57. Mackenzie, "From API to AI."
58. Robert Bodle, "Regimes of Sharing," *Information, Communication & Society* 14, no. 3 (2011): 320-37, doi:10.1080/1369118X.2010.542825; Anja Bechmann, "Internet Profiling: The Economy of Data Intraoperability on Facebook and Google," *MedieKultur* 29, no. 55 (2013): 72-91, doi:10.7146/mediekultur.v29i55.8070; Taina Bucher, "Objects of Intense Feeling: The Case of the Twitter API," *Computational Culture* 3 (2013), http://computationalculture.net/article/objects-of-intense-feeling-the-case-of-the-twitter-api.
59. Nieborg and Helmond, "The Political Economy"; Helmond, "The Platformization of the Web"; Gerlitz and Helmond, "The Like Economy."
60. Mikkel Flyverbom and John Murray, "Datastructuring – Organizing and Curating Digital Traces into Action," *Big Data & Society* 5, no. 2 (July 1, 2018): 2053951718799114, doi:10.1177/2053951718799114; Helmond, "The Platformization of the Web"; Tarleton Gillespie, "Algorithmically Recognizable: Santorum's Google Problem, and Google's Santorum Problem," *Information, Communication & Society* 20, no. 1 (2017): 63-80, doi:10.1080/1369118X.2016.1199721.
61. Niels Kerssens, "'The Database 'Revolution': The Technological and Cultural Origins of the Big-Data-Based Mindset in American Management, 1970s–1980s," *Tijdschrift Voor Mediageschiedenis* 21, no. 2 (2018), doi:10.18146/tmg21364.
62. Kerssens, "'The Database 'Revolution'"; Andrew Iliadis, "Algorithms, Ontology, and Social Progress," *Global Media and Communication* 14, no. 2 (2018): 219-30, doi:10.1177/1742766518776688.
63. Fernando N. van der Vlist, "Accounting for the Social: Investigating Commensuration and Big Data Practices at Facebook," *Big Data & Society* 3, no. 1 (2016): 1-16, doi:10.1177/2053951716631365.

64. Helmond, Nieborg, and Van der Vlist, "Facebook's Evolution."
65. Matthew Crain, "A Critical Political Economy of Web Advertising History," in *The SAGE Handbook of Web History*, ed. Niels Brügger and Ian Milligan (London: SAGE, 2019).
66. Fernando N. van der Vlist and Anne Helmond, "Infrastructuring Social Data: Social Media as Ad Platforms" January 5, 2019; Helmond, Nieborg, and Van der Vlist, "Facebook's Evolution."
67. Helmond, Nieborg, and Van der Vlist, "Facebook's Evolution: Development of a Platform-as-Infrastructure."
68. Bruns and Weller, "Twitter As a First Draft of the Present," 186.
69. SHARE LAB, "Quantified Lives on Discount," *SHARE LAB*, August 19, 2016, https://labs.rs/en/quantified-lives/.
70. Varner Angwin, Julia Madeleine, "Facebook Enabled Advertisers to Reach 'Jew Haters,'" *ProPublica*, September 14, 2017, https://www.propublica.org/article/facebook-enabled-advertisers-to-reach-jew-haters.
71. Facebook Business, "Keeping Advertising Safe and Civil," Blog post, *Facebook Business*, (August 21, 2018), https://www.facebook.com/business/news/keeping-advertising-safe-and-civil.
72. Van der Vlist, "Accounting for the Social."
73. Brendan M. Sullivan et al., Socioeconomic group classification based on user features, United States Patent and Trademark Office US20180032883A1, filed July 27, 2016, and issued February 1, 2018, https://patents.google.com/patent/US20180032883A1/.
74. Geoffrey C. Bowker and Susan. L. Star, *Sorting Things Out: Classification and Its Consequences* (Cambridge, MA: MIT Press, 1999); van der Vlist, "Accounting for the Social."
75. Nieborg and Helmond, "The Political Economy"; Helmond, Nieborg, and Van der Vlist, "Facebook's Evolution."
76. Helmond, Nieborg, and Van der Vlist, "Facebook's Evolution."
77. Ibid.
78. Ibid.; Van der Vlist and Helmond, "Infrastructuring Social Data."
79. Brügger, *The Archived Web*; De Reuver, Sørensen, and Basole, "The Digital Platform."
80. Helmond, Nieborg, and Van der Vlist, "Facebook's Evolution."
81. Bruns and Weller, "Twitter As a First Draft of the Present."
82. Venturini and Rogers, "'API-Based Research.'"
83. Anat Ben-David and Adam Amram, "Computational Methods for Web History," in *The SAGE Handbook of Web History*, ed. Niels Brügger and Ian Milligan (London: SAGE, 2018), 153-67.
84. Cf. Van Dijck, *The Culture of Connectivity*.
85. Brügger, "Web Historiography and Internet Studies"; Rogers, "Doing Web History."
86. Helmond, Nieborg, and Van der Vlist, "Facebook's Evolution."
87. Katherine Boss and Meredith Broussard, "Challenges of Archiving and Preserving Born-Digital News Applications," *IFLA Journal* 43, no. 2 (2017): 150-57, doi:10.1177/0340035216686355; Dieter et al., "Multi-Situated App Studies: Methods and Propositions."
88. Brügger, *The Archived Web*, 202-4.
89. Justin F. Brunelle et al., "The Impact of JavaScript on Archivability," *International Journal on Digital Libraries* 17, no. 2 (2016): 95-117, doi:10.1007/s00799-015-0140-8.
90. Brügger, *The Archived Web*; Thomson, "Preserving Social Media."
91. Thomson, "Preserving Social Media."
92. Ibid.; Littman et al., "API-Based Social Media Collecting."
93. Brügger, *The Archived Web*; Ben-David and Amram, "Computational Methods for Web History."
94. Bruns and Weller, "Twitter As a First Draft of the Present."
95. Brügger, *The Archived Web*; Thomson, "Preserving Social Media"; Littman et al., "API-Based Social Media Collecting"; Rogers, "Periodizing Web Archiving."
96. Robert Gehl, *Reverse Engineering Social Media: Software, Culture, and Political Economy in New Media Capitalism* (Philadelphia, PA: Temple University Press, 2014); Niels Brügger, "A Brief History of Facebook as a Media Text: The Development of an Empty Structure," *First Monday* 20, no. 5 (2015), doi:10.5210/fm.v20i5.5423.
97. Thomson, "Preserving Social Media"; Brügger, *The Archived Web*.
98. Brügger, *The Archived Web*; Thomson, "Preserving Social Media."
99. Winters, "Coda."
100. Rogers, "Doing Web History."
101. Lomborg, "Researching Communicative Practice"; Brügger, *The Archived Web*.

102. Anne Helmond, "Historical Website Ecology: Analyzing Past States of the Web Using Archived Source Code," in *Web 25: Histories from the First 25 Years of the World Wide Web*, ed. Niels Brügger (New York, NY: Peter Lang, 2017), 139-55.
103. Helmond, Nieborg, and Van der Vlist, "Facebook's Evolution."
104. Thorsten Busch and Tamara Shepherd, "Doing Well by Doing Good? Normative Tensions Underlying Twitter's Corporate Social Responsibility Ethos," *Convergence* 20, no. 3 (2014): 293-315, doi:10.1177/1354856514531533; Richard Rogers, "Debanalizing Twitter: The Transformation of an Object of Study," in *Proceedings of the 5th Annual ACM Web Science Conference* (New York, NY: ACM, 2013), 356-365; Anna Lauren Hoffmann, Nicholas Proferes, and Michael Zimmer, "'Making the World More Open and Connected': Mark Zuckerberg and the Discursive Construction of Facebook and Its Users," *New Media & Society* 20, no. 1 (2018): 199-218, doi:10.1177/1461444816660784.
105. Brügger, "A Brief History of Facebook"; Søren Vigild Poulsen, "Becoming a Semiotic Technology – a Historical Study of Instagram's Tools for Making and Sharing Photos and Videos," *Internet Histories* 2, no. 1-2 (2018): 121-39, doi:10.1080/24701475.2018.1459350; José van Dijck, "Tracing Twitter: The Rise of a Microblogging Platform," *International Journal of Media & Cultural Politics* 7, no. 3 (2011): 333-348, doi:10.1386/macp.7.3.333_1.
106. Busch and Shepherd, "Doing Well by Doing Good?"; Carolin Gerlitz et al., "Regramming the Platform: Infrastructural Relations between Apps and Social Media," *Computational Culture*, n.d.
107. Van Dijck, "Tracing Twitter."
108. Cornelius Puschmann and Jean Burgess, "The Politics of Twitter Data," in *Twitter and Society*, ed. Katrin Weller et al. (New York, NY: Peter Lang, 2014), 43-54.
109. Robert Bodle, "Regimes of Sharing," *Information, Communication & Society* 14, no. 3 (2011): 320-37, doi:10.1080/1369118X.2010.542825; Puschmann and Burgess, "The Politics of Twitter Data"; Bruns and Weller, "Twitter As a First Draft of the Present."
110. Bechmann, "Internet Profiling"; Puschmann and Burgess, "The Politics of Twitter Data."
111. Mackenzie, "From API to AI"; Nieborg and Helmond, "The Political Economy"; Helmond, Nieborg, and Van der Vlist, "Facebook's Evolution"; Carolin Gerlitz et al., "Regramming the Platform."
112. Flyverbom and Murray, "Datastructuring"; Van der Vlist, "Accounting for the Social"; Helmond, "The Platformization of the Web."
113. Gehl, *Reverse Engineering Social Media*; van der Vlist, "Accounting for the Social"; Ghazawneh and Henfridsson, "Balancing Platform Control."
114. Gerlitz and Helmond, "The Like Economy"; Helmond, "The Platformization of the Web"; Helmond, Nieborg, and Van der Vlist, "Facebook's Evolution"; Nieborg and Helmond, "The Political Economy"; Plantin et al., "Infrastructure Studies Meet Platform Studies"; Mackenzie, "From API to AI."
115. Bucher, "Objects of Intense Feeling"; Anne Helmond et al., "Apps and Their Stories."
116. Nieborg and Helmond, "The Political Economy"; Helmond, Nieborg, and Van der Vlist, "Facebook's Evolution."
117. Van Dijck, *The Culture of Connectivity*.
118. Helmond, Nieborg, and Van der Vlist, "Facebook's Evolution"; Van der Vlist and Helmond, "Infrastructuring Social Data."
119. Helmond, Nieborg, and Van der Vlist, "Facebook's Evolution."
120. Yuval Dror, "'We Are Not Here for the Money': Founders' Manifestos," *New Media & Society* 17, no. 4 (2015): 540-55, doi:10.1177/1461444813506974.
121. Nieborg and Helmond, "The Political Economy"; Helmond, Nieborg, and Van der Vlist, "Facebook's Evolution."
122. Ronan Ó. Fathaigh, Joris van Hoboken, and Nico A. N. M. van Eijk, "Mobile Privacy and Business-to-Platform Dependencies: An Analysis of SEC Disclosures," SSRN Scholarly Paper (Rochester, NY: Social Science Research Network, March 16, 2018), https://papers.ssrn.com/abstract=3142246.
123. Greg Elmer, "Precorporation: Or What Financialisation Can Tell Us about the Histories of the Internet," *Internet Histories* 1, no. 1-2 (2017): 90-96, doi:10.1080/24701475.2017.1308197.
124. Michael A. DeVito, "From Editors to Algorithms," *Digital Journalism* 5, no. 6 (2017): 753-773, doi:10.1080/21670811.2016.1178592; Esther J. T. Weltevrede, "Repurposing Digital Methods: The Research Affordances of Platforms and Engines" (Ph.D., University of Amsterdam, 2016), https://wiki.digitalmethods.net/Dmi/RepurposingDigitalMethods.
125. Van der Vlist, "Accounting for the Social."
126. Mackenzie, "From API to AI."

Biographies

Anne Helmond is an Assistant Professor of New Media and Digital Culture at the University of Amsterdam. Her research interests include software studies, platform studies, app studies, digital methods, and web history.

Fernando van der Vlist is a PhD candidate at Utrecht University and a research associate with the Collaborative Research Centre "Media of Cooperation" at the University of Siegen. His research interests include software studies, digital methods, social media and platform studies, app studies, and critical data studies.

Appendix

Table A-1. Issues and challenges of social media archiving.

Issues[88]	challenges	approaches (sample)	privileged histories (sample)
'high-speed updating and APIs'	*general:* • database-driven and dynamic content (e.g., JavaScript)[89] • customisation and personalisation (e.g., personalised content and interfaces) *API-related:* • data volume and velocity[90] • presentism due to content and data recency biases[91] • data access restrictions and terms of service[92] • data completeness, reliability, and sampling[93] • privacy and ethical concerns[94]	*general:* • alternative web crawling technologies (e.g., Brozzler) • de-personalised (or defaulted) archiver profiles to enable web crawling beyond logins (e.g., IA's Facebook user, Charlie Archivist) *API-related:* • API-based archiving[95] • re-using previously collected data sets (often limited due to data re-sharing restrictions), including donated data sets (e.g., Library of Congress Twitter Archive) • collecting data sets individually (often limited due to API specifications and data storage limitations) • purchasing data sets commercially, either directly or through resellers (often limited due to pricing and recency bias)	*general:* • histories of 'empty structures'[96] *API-related:* • (big) data-driven histories of user content and interactions, public figures, media events, and large organisations (e.g., thematic collections of public pages, groups, hashtags, and user profiles) and of selectively tracked topics or accounts[97]
'inaccessibility' and fenced off content	*general:* • robots.txt exclusion standards • private user profiles, logins, and password-protected sites • *API-related:* • data access restrictions (e.g., blocking)	*general:* • subscription-based solutions (e.g., Archive-It) • micro-archiving (e.g., downloading personal content and data archives, data access requests under GDPR)[98] • archiving screen captures and recordings • creating or purchasing data sets commercially	*general:* • histories of selectively tracked personal accounts or topics and histories of openly accessible or published data[99] • auto-biographical histories and 'screencast documentaries'[100] • premium histories with paid-for content and data
'integrated digital media environments'	*general:* • syndicated and embedded user content[101]	*general:* • historical source code analysis[102] • historical partnership analysis[103]	*general:* • histories of relationship networks (e.g., hyperlink networks, ecosystem evolution)

Table A-2. User groups of social media platforms.

user group	archived resources (sample)	afforded histories (sample)
end users	• graphical user interfaces and 'empty' frames • sign up and registration procedures • about and feature pages • data and privacy policies • terms of service and use • end-user license agreements (EULA) • account security pages • help pages • language support	• self-description histories[104] • feature and practice histories[105] • 'stakeholder politics' histories[106] • revenue model histories[107] • data and privacy policy histories • terms of service and use histories[108]
developers	• tools and product pages • application programming interfaces (APIs) and endpoints • software development kits (SDKs) • integrated development environments (IDEs) • software and developer tools and frameworks • guides for app development, best practices, app review, and privacy and consent • online training courses for developers • developer support, help pages, and frequently asked questions (FAQs) • API reference documentation, version histories, and changelogs • developer news, blog posts, and blog archives • open source projects and code repositories • programming, query, and markup languages • bug reports • platform status • annual developer conferences • developers community groups, meetups and local developer communities • startup accelerator programmes • platform policies • careers • platform and privacy policies • cookies • terms of service and use	• API-based data sharing histories[109] • data strategy and 'intraoperability' histories[110] • 'programmability' and app development histories[111] • app ecosystem histories • tracking technology histories • standards and protocol histories • 'datastructuring' histories[112] • platform architecture design and governance and control histories[113] • platform growth and embedding histories[114] • platform status, maintenance, repair, and 'issue' histories • platform and privacy policy histories[115]
business	• ad creation and management tools and interfaces • tools and product pages • business news, blog pages, and archives • case studies, success stories, and insights • guides for marketers, creative tips, and inspiration • advertising and marketing tools, products, and services (e.g., ads and marketing APIs and SDKs) • webinars and online training courses for digital buyers, publishers, and marketers • industry insights and reports and series • insights newsletters • business certification programmes • common questions, advertiser help centres, and support • privacy policies • brand resources	• tool and product histories • ad creation and targeting fields histories • platform growth and embedding histories[116] • revenue model histories[117]

Table A-2. (Continued)

user group	archived resources (sample)	afforded histories (sample)
partners	• partner solutions and integrations • partner programmes and directories • about pages • partner news, blog posts, and archives • case studies and success stories • partner sign-up and registration procedures • partner certification programmes and badges • partner help centres and support • partner newsletters • partner summits and events	• partnerships, partner solutions, and partner ecosystem histories[118] • strategic partnerships and strategic growth histories • platform growth and embedding histories[119]
creators	• creator tools and products (e.g., for creating, managing, and monetizing content and communities) • webinars and online training courses for creators • guides for content production and monetisation • guides for community building and fan engagement • copyright guidelines	• content and media format histories • content and media monetisation histories • platform and algorithm optimisation strategy histories • brand histories
media and publishers	• media and publisher solutions and integrations • media and publisher partner programmes and directories • media and publisher help centres and support • guides for content production and monetisation • guides for community building and fan engagement • press releases, news, blog posts, and archives • analytics and measurement tools • webinars and online training courses for producing and discovering content and managing audiences	• content and media format histories • content and media monetisation histories • platform and algorithm optimisation strategy histories • brand histories
investors	• company information and company histories • financial information (e.g., quarterly earnings, annual reports, US SEC filings; webcasts, balance sheets, income statements, earnings call transcripts and slides) • stock information • education centres, success stories, business guides, and frequently asked questions (FAQs) • press releases, blog posts, and archives • annual meetings and investor events • corporate governance and responsibilities • investor newsletters	• investor relations histories • corporate strategy and financial performance histories • founder manifestos, mission statements, and business discourse histories[120] • stock trading, stock price development, and analyst coverage histories • mobile platform strategy histories[121] • data privacy and 'business-to-platform' dependency histories[122] • 'financialised histories'[123] • corporate governance structure histories
research	• research areas • publications (e.g., conference proceedings, journal articles, technical reports, patent filings, white papers) • downloads • press releases, blog posts, and archives • academic programmes and scholarships • careers	• intellectual and research area histories • technology design and innovation histories • algorithm design and maintenance histories[124] • platform architecture design and infrastructure histories[125] • 'programmability' design tendency and experimentation histories[126] • professional expertise, talent acquisition, and recruitment strategy histories