

The platformisation of software development: Connective coding and platform vernaculars on GitHub

Liliana Bounegru 
King's College London, UK

Convergence: The International
Journal of Research into
New Media Technologies
2023, Vol. 0(0) 1–21
© The Author(s) 2023



Article reuse guidelines:

sagepub.com/journals-permissions
DOI: 10.1177/13548565231205867
journals.sagepub.com/home/con



Abstract

This article contributes to recent scholarship on platform, software and media studies by critically engaging with the ‘social coding’ platform GitHub, one of the most prominent actors in the online proprietary and F/OSS (free and/or open-source software) code hosting space. It examines the platformisation of software and project development on GitHub by combining institutional and cultural analysis. The institutional analysis focuses on critically examining the platform from a material-economic perspective to understand how it configures contemporary software and project development work. It proposes the concept of ‘connective coding’ to characterise how software intermediaries such as GitHub configure, valorise and capitalise on public repositories, developer and organisation profiles. This institutional perspective is complemented by a case study analysing cultural practices mediated by the platform. The case study examines the platform vernaculars of news media and journalism initiatives highlighted by *Source*, a key publication in the newsroom software development space, and how GitHub modulates visibility in this space. It finds that the high-visibility platform vernacular of this news media and journalism space is dominated by a mix of established actors such as *the New York Times*, *the Guardian* and Bloomberg, as well as more recent actors and initiatives such as ProPublica and Document Cloud. This high-visibility news media and journalism platform vernacular is characterised by multiple F/OSS and F/OSS-inspired practices and styles. Finally, by contrast, low-visibility public repositories in this space may be seen as indicative of GitHub’s role in facilitating various kinds of ‘post-F/OSS’ software development cultures.

Keywords

Connective coding, platform studies, platformization, GitHub, internet studies, digital methods, open-source, journalism, social media, software studies, cultural studies, software development

Corresponding author:

Liliana Bounegru, King’s College London, Strand Campus, Strand, London, WC2R 2LS Room S3.27, UK.
Email: liliana.bounegru@kcl.ac.uk

Introduction

Founded in 2008, GitHub is one of the largest online software development and code hosting services (Gousios et al., 2014). It is also one of the most widely used platforms in the F/OSS (free and/or open-source software) space (Vasudevan, 2021; Otto et al., 2023). The GitHub platform is based on Git, an open-source distributed version control and source code management system created in 2005 by Linus Torvalds to support collaboration on the development of the Linux operating system. GitHub extends Git with cloud-based hosting, browser-based and desktop-based graphical interfaces, and social networking functionalities. Acquired by Microsoft in 2018, GitHub is now part of a group of powerful technology companies on which digital cultural production in the Global North is dependent, known as GAFAM or Google, Apple, Facebook, Amazon and Microsoft (Nieborg and Poell, 2018).

While the platformisation of cultural production has been studied in relation to many online platforms and socio-cultural practices, from games to the web, mobile apps, healthcare, education, news, transport and scholarly communication (Aradau et al., 2019; Blanke and Pybus, 2020; Bogost and Montfort, 2009; Helmond, 2015; Hind et al., 2022; Nieborg and Poell, 2018; Plantin et al., 2016; Srnicek, 2016; Van Dijck et al., 2018), GitHub's platformising mechanisms and how they configure software development deserve further critical examination from platform, software and cultural studies perspectives. The significant contributions made so far in this respect, by e.g., Fuller et al. (2017), Mackenzie (2018), Otto et al. (2023) and Vasudevan (2021) will be discussed in the following sections.

GitHub has become a prominent site for studying how cultural production is platformised. As one of the largest online repositories for proprietary and open-source software development, it plays a central role in the digital economy, particularly as a platform for making infrastructures and platforms (Mackenzie, 2018; Otto et al., 2023; Vasudevan, 2021). It is therefore an important site for studying platform-specific transformations of culture and society (Mackenzie et al. 2015).

User tensions and resistance are also known to surface in this 'hybrid and hybridizing' space (Otto et al. 2023) where both proprietary and F/OSS development practices may be found. The Microsoft takeover was one such moment where tensions within GitHub's F/OSS developer community were amplified and power imbalances within the platform ecosystem were made visible (Vasudevan, 2021).

This article aims to contribute to the growing body of work on platforms, platformisation and cultural production by examining GitHub's platformising mechanisms and how they structure software development, as well as its platform vernaculars (Gibbs et al., 2015). To do so I combine an institutional perspective (Duffy et al., 2019) focused on GitHub's material-economic configuration, with an analysis of platform practices of media and journalism organisations on GitHub.

The article is organised as follows. In the first section I discuss how the relationship between GitHub, F/OSS and proprietary software development has been researched so far. In the second section I review literature on the platformisation of cultural production, focusing particularly on concepts and analytical approaches that inform my research. In the third section I describe the methodology of this study. The fourth and fifth sections discuss the empirical research and its findings, examining GitHub's platformising mechanisms from an institutional perspective before turning to how visibility of media and journalism actors and practices is modulated as a result of metrification and algorithmic sorting and recommendation mechanisms.

GitHub: Networking and hybridising F/OSS, proprietary software and other cultural practices

GitHub as a hybridising space

In relation to contemporary software cultures, GitHub has been conceptualised as a ‘hybrid and hybridizing’ space where F/OSS and proprietary software development practices, logics, values and modes of valuation cohabit and intersect (Otto et al., 2023). This entanglement is conspicuously reflected in Microsoft ranking as top contributor to open-source projects on the platform in 2017, followed by its purchase of GitHub in 2018 (Birkinbine, 2020).

The entanglement of proprietary and open-source software development practices precedes GitHub and has characterised F/OSS practices from their beginning (Kelty, 2008). F/OSS is typically defined by five key components: ‘shared source code, a concept of openness, copyleft licences, forms of coordination, and a movement or ideology’ (Kelty, 2008: 254). According to Kelty, the open-source software narrative was introduced in 1998 in connection to the dotcom boom to signify a ‘pragmatist’ approach to the use of free software in commercial settings, emphasising ‘economic value’, ‘cost savings’ and ‘getting things done’ (2008: 116). The concept represented a departure from the free software narrative, which promoted ‘resistance to proprietary software “hoarding”’ in favour of principles of freedom, autonomy and creativity of individual developers (Kelty, 2008: 99). It is emblematically encapsulated in the release of the source code for the Netscape’s Communicator Web browser.

The expansion of the open-source model of software development has led to coding becoming a more ‘public economic and cultural activity’, attracting interest and participation from a wide range of domains, from business to science, government and art (Mackenzie, 2018: 38). Indeed the open-source ethos no longer applies only to software development but has extended into other cultural practices as well. Kelty (2008) argues for seeing practices such as open access, open science, open content, open data and open government as continuous with and developments of free and open-source software practices.

The hybridisation of proprietary and open-source software development practices has been shown to be intensifying in the digital economy. Birkinbine (2020) sees F/OSS at the intersection between commons-based peer production and capitalist production as corporate involvement in F/OSS increases, both in terms of contributions to F/OSS projects and incorporating these projects and practices in commercial operations and operating principles.

Another aspect of GitHub’s hybridising of software cultures is identified by Fuller et al. (2017). They propose the concept of post-F/OSS to describe a platform culture that co-exists alongside F/OSS and proprietary coding. This comprises widespread practices of publicly posting mundane code projects without a licence and with little to no documentation to help users understand these projects. According to Fuller et al., this ‘general indifference to the discussions of and loyalty to certain kinds of licences and the sense of ethics (GPL) or business models (Open Source) that these drew upon’ (2017: 58), marks a departure from the F/OSS movement’s practices.

GitHub as social networking site

Mackenzie (2018) argues that GitHub and other more recent code sharing platforms are distinct from earlier online code repositories such as SourceForge and Google Code in two ways. Firstly, GitHub configures coding as a ‘social networked practice’ (Mackenzie, 2018). This is materialised by means of ‘layers of social media-style interface’ that ‘adorn code repositories with all the social

media-style apparatus of following, watching, liking, and tagging' (Mackenzie, 2014: 10). In this environment, the social life of users and repositories is organised around vanity and reputational metrics (Rogers, 2017) akin to those found on platforms such as Twitter and Facebook. Indeed in a study of an entire public GitHub archive, Mackenzie finds that 'social' events exceed 'technical' events pertaining to code writing on GitHub (2014).

Secondly, GitHub's social action model based on combining social networking features with the decentralised version control system Git, has extended into cultural practices beyond software development, such as 'how-to guides, metadata on the Tate's art collection, the White House's open data policy, legal documents, recipes, books, and blogs' (Mackenzie et al., 2015: 369). One of the implications of converting software development into a 'social media platform practice or social coding practice' (Mackenzie, 2014: 1), is the facilitation of platform-oriented forms of centralisation and control (Fuller et al., 2017).

Following these studies, this article turns to questions that emerge at the intersections between research on GitHub's platform configuration and the cultural practices which it mediates. Firstly, how exactly does GitHub network and platformise software and project development? While the studies reviewed in this section point towards some of these mechanisms, our understanding of this platform would benefit from a more comprehensive analysis of GitHub's platformisation and social networking mechanisms. Secondly, how do GitHub's hybrid use practices emerge through its material-economic infrastructure? How are attention and visibility modulated by means of GitHub's social networking features and algorithmic ordering mechanisms? What practices are 'elevated into prominence' (Geboers & Van De Wiele, 2020) by means of these mechanisms?

Platforms, platformisation and platform vernaculars

Research from platform, software and infrastructure studies, as well as cultural studies, offers useful approaches and analytical frameworks to examine the platformisation (Helmond, 2015; Plantin et al., 2016; Van Dijck et al., 2018) of various aspects of collective life, that is, how social media platforms shape infrastructures, participation, sociality, culture and other relations and processes (Gillespie, 2010).

Poell et al. (2019) define platformisation as 'the penetration of infrastructures, economic processes and governmental frameworks of digital platforms in different economic sectors and spheres of life, as well as the reorganisation of cultural practices and imaginations around these platforms' (p.1). This definition suggests that understanding platformisation involves attending to both institutional dimensions and cultural practices (Duffy et al., 2019). Similarly, Dolata & Schrape (2023) call for a multifaceted understanding of platforms in terms of platform companies and their business goals, the platform as a market and social action space, and platform governance mechanisms. In the context of studying GitHub, Otto et al. (2023) argue that an exclusive focus on institutional processes risks missing the platform's multivalence and the heterogeneous use practices and valuation regimes associated with it.

Nieborg and Poell (2018), Poell et al. (2019), Duffy et al. (2019) and Nieborg et al. (2020) offer an analytical framework for examining platformising processes from an institutional perspective along three intersecting dimensions: markets, governance and infrastructure. The first calls on researchers to study how platforms reconfigure markets. For instance, platforms may be studied as multi-sided markets (see, e.g. Evans et al., 2006; Rochet and Tirole, 2003). The governance dimension refers to examining how platforms govern the cultural production practices and industries they mediate. The infrastructural dimension calls for examining the infrastructural aspects of platforms and how they shape cultural production. Here attention is paid to 'interfaces, data flows,

and the availability and functions of software development tools and documentation' (Nieborg and Poell, 2018: p. 13). For example, when studying the infrastructural dimensions of platforms, Gerlitz and Rieder (2018) draw on the notion of 'grammars of action' (Agre, 1994) to attend to how platforms structure action and interaction possibilities by means of 'predefined options' and forms, such as linking, retweeting, following and sharing.

Complementary to the institutional perspective, the cultural practices perspective on analysing platformisation is concerned with how platforms shape cultural practices and how user practices shape platforms (Duffy et al., 2019; Nieborg et al., 2020; Poell et al. 2019). In terms of communication practices, the mutual shaping of platforms and user practices has been understood to materialise as 'platform vernaculars' (Gibbs et al., 2015). The term is used to refer to platform-specific communication practices and genres that share 'conventions and grammars of communication, emerging within platforms and populations of users through the interplay of platform affordances and their appropriations' (Meese et al., 2015: 1820). Different platform vernaculars may materialise around different issues or groups of users. Rieder et al. (2018) propose the notion of platform 'issue vernaculars' to refer to communication practices or genres that emerge on a platform around particular topics or communities. The concept is useful in the context of studying platformisation because it is sensitive to the material-economic configuration of platforms and how it shapes collective cultural practices on the platform.

Studying platform vernaculars requires a sensibility towards platform content as digital objects which are treated and networked by the platform in particular ways, that is, 'how platforms and engines serve, format, redistribute and essentially co-produce content' (Niederer, 2019: 18). The platform's ordering practices and 'ranking cultures' (Rieder et al., 2018) modulating visibility of content and users are constitutive of platform vernaculars. They often introduce new visibility regimes which shape social action 'around the pursuit of visibility' (Bucher, 2012a: 1165; see also Geboers & Van De Wiele, 2020).

Following from these two approaches to the study of platformisation, in this article I examine GitHub's material-economic configuration and its platform vernaculars. The latter is done through a case study on how visibility is modulated in the media and journalism space on GitHub.

Research methods

Methodologically this article combines 'technographic inquiry' (Bucher, 2018) and digital methods approaches for social and media research (Marres and Gerlitz, 2015; Rogers, 2013, 2019; Venturini et al., 2018).

Bucher (2018) defines technography as scrutinising the mechanisms and operational logics of platforms and algorithms and how they configure social action similarly to how an ethnographer would examine culture by means of the way people ascribe meanings to worlds. More specifically, she defines it as 'a way of describing and observing the workings of technology in order to examine the interplay between a diverse set of actors (both human and nonhuman)' (Bucher, 2018: 60).

I use technographic inquiry to understand GitHub's material-economic configuration (primarily prior to its purchase by Microsoft), and how it structures software and project development. To do so I examine materials such as GitHub's various interfaces (including its graphical user interfaces and APIs), platform documentation from help pages, the platform's development blog and technology press articles pertaining to the dimensions of platformisation I focus on, primarily business models and technical configuration.

I use digital methods approaches to social and media research to examine high-visibility platform practices and vernaculars on GitHub by means of an analysis of top starred repositories and

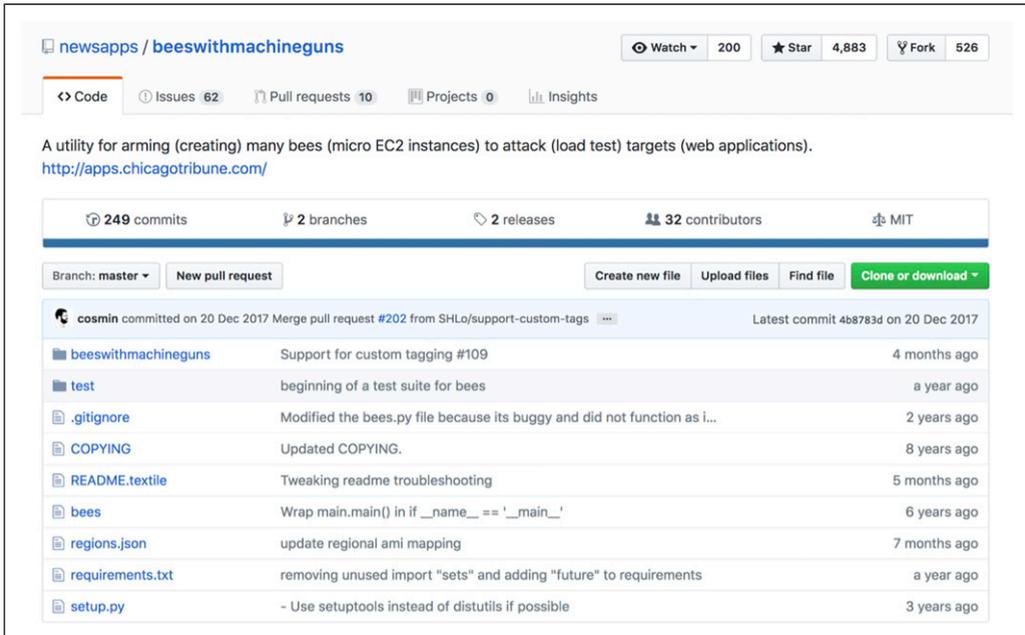


Figure 1. Browser-based user interface view of public GitHub repository.

associated organisations. To examine platform vernacular practices in a situated way as they develop around particular issues, cultural practices or communities, I focus on media and journalism organisations and initiatives on GitHub. I focus on this domain because GitHub is one of the most used code hosting platforms by news organisations (Usher, 2016). I focus on organisational accounts as organisations have been found to drive activity on GitHub (Mackenzie et al., 2014).

For this study I use a collection of 3,665 public repositories created between 2009 and 2016 by 87 media initiatives and organisations with public GitHub accounts (see the Annex for a list of these accounts). An example of one such repository can be seen in Figure 1. The list of organisations is maintained by a key publication in the newsroom software development space, *Source*.¹ I focus on this list because of its association with the OpenNews programme, one of earliest and most prominent programmes supporting open-source digital technological innovation in newsrooms (Lewis and Usher, 2016). The list of organisations was scraped from *Source*'s website in June 2016. Organisations or initiatives that were not primarily news media or journalistic in focus, such as Twitter, were removed from the list and missing GitHub accounts of listed organisations were manually added. The public repositories associated with these accounts and their associated metadata were collected via the repository endpoint of the GitHub API through a number of research software tools that I conceptualised and co-developed with the Digital Methods Initiative.² Besides accessibility reasons, I focus on public repositories because, as I will discuss in the next section, they are subjected to more intensive networking and platformisation processes than private repositories.

Next, I take an emergent coding approach to describe the organisations publishing these repositories in terms of their location, media type (e.g. news organisation, broadcaster and magazine), and whether they are a public or privately owned media entity. To illustrate how platform networking features and algorithmic sorting mechanisms modulate visibility of news and journalism organisations and projects and the kinds of actors that are 'elevated into prominence' (Geboers &



Figure 2. Screenshot from archived version of GitHub's 2008 homepage. <https://web.archive.org/web/20080514210148/http://github.com/>.

Van De Wiele, 2020), I focus on the repository starring feature. Combining digital methods approaches to corpus curation (Rogers, 2019) with qualitative content analysis (Caliandro and Gandini, 2016; Niederer, 2019), I examine 88 repositories that receive at least 100 stars in the collection (see the Annex for a list of these repositories). I analyse the actors who publish these repositories in terms of their public software development styles, and the themes of these repositories. To do so I rely on a number of different empirical materials: the repository and organisation profile on the GitHub platform, the organisations' technology development blogs, as well as other materials about these repositories and organisations published on the web, such as technology articles and press releases.

Making software development connective

One of the earliest and most enduring mottos of the company has been social coding (see Figure 2). GitHub's 'Get Started' guide suggests that being social includes following users, saving and subscribing to repositories and joining discussion forums.³ However, platforms are neither neutral nor stable carriers for social action, but powerful actors that 'participate in shaping the worlds they only purport to represent' (Bucher, 2018: 1). We might therefore ask, what kind of 'sociality' does Github configure? In this section I empirically examine how GitHub configures software development as a social and cultural practice by means of an examination of its platforming processes and mechanisms from an institutional perspective, including its technical infrastructure and market configuration.

GitHub's market configuration

References to GitHub as an open-source code repository co-exist alongside business reports that document GitHub's annual revenue being generated by businesses and individuals maintaining private repositories.¹⁷ The GitHub Octoverse 2022 report states that more than 90% of the largest US-based companies use GitHub services (GitHub, 2022).

The facilitation of these seemingly divergent use cases and groups of end-users is not an anomaly but a constitutive and defining feature of online platforms. Online platforms act as technical

infrastructures that negotiate and enable interactions between two or more different groups of end-users or stakeholders with heterogenous goals and interests (Gillespie, 2010; Srnicek, 2016). In the case of GitHub, these include open-source and individual developers, businesses and third-party application and service developers.

From an economic perspective, platforms such as GitHub are understood as ‘multi-sided markets’ (Evans et al., 2006; Rochet and Tirole, 2003). As a ‘product platform’ (Srnicek, 2016) which generates revenue by providing a service against a subscription fee, GitHub is organised around the freemium model, whereby free repository hosting is provided for public projects and paid hosting for private repositories.

GitHub subsidises individuals and organisations who code publicly and charges businesses and developers who code privately, by providing various paid plans for private repositories. This is because developers and organisations who host their code publicly are not only providing a vast collection of open-source projects that businesses can draw on, but they are also seen as marketing and conversion tools to attract enterprise solutions, as developers who make personal use of the platform may be recruited by businesses.¹⁶ GitHub also charges third-party app developers who build apps to enhance GitHub’s functionalities and advertise them on the GitHub Marketplace. In order to encourage a third-party ecosystem of products and services that use platform data to enhance the platform experience, GitHub does not charge for API usage that complies with its rate limits. It does, however, charge for API usage that goes beyond its rate limits or that results in the marketisation of services that mimic the GitHub service experience.⁴

Making participation economically valuable

An essential platformising process that enables the facilitation of these heterogenous groups and interests into this marketplace configuration, is what Van Dijck (2013) calls ‘turning connectedness into connectivity by means of coding technologies’ (p. 16).

While those engaging in developing software and other projects publicly do not pay fees for their use of the platform, they do contribute value through the publicity, traceability, metrification, analysability and valorisation of their participation on the platform.

Conditions for participation are set through GitHub’s technical infrastructure in alignment with the platform’s economic aims. This includes a front-end which seeks to solicit, intensify and accelerate user engagement and a back-end comprised of servers and data storage, mining and archival capabilities (Gehl, 2014). These features, just as in the case of other social media platforms, are organised around nurturing a platform ecosystem that multiplies valorisation of connectivity around several registers (Gerlitz, 2016; Marres, 2017).

In particular, following the model of other social media platforms, participation in software development is made economically valuable by setting up an infrastructure that pre-defines and standardises possibilities for user action and the forms that these take at the user interface level by means of ‘grammars of action’ (Gerlitz and Rieder, 2018; Gray et al., 2018). In the case of GitHub these grammars would include forms of action such as ‘committing’, submitting ‘pull requests’, ‘forking’, ‘starring’ or ‘watching’ repositories, and following users. Through each of these actions, connections are recorded between users, and between users and objects such as repositories. In addition to this, free repositories are subjected to a regime of transparency (Dabbish et al., 2012), whereby every activity associated with them is attributed and visible to anyone who accesses the platform (see Figure 3).

Standardising possibilities for action enables social media platforms like GitHub to render selected activities, projects and people measurable, calculable and comparable (Gerlitz and Rieder, 2018). This

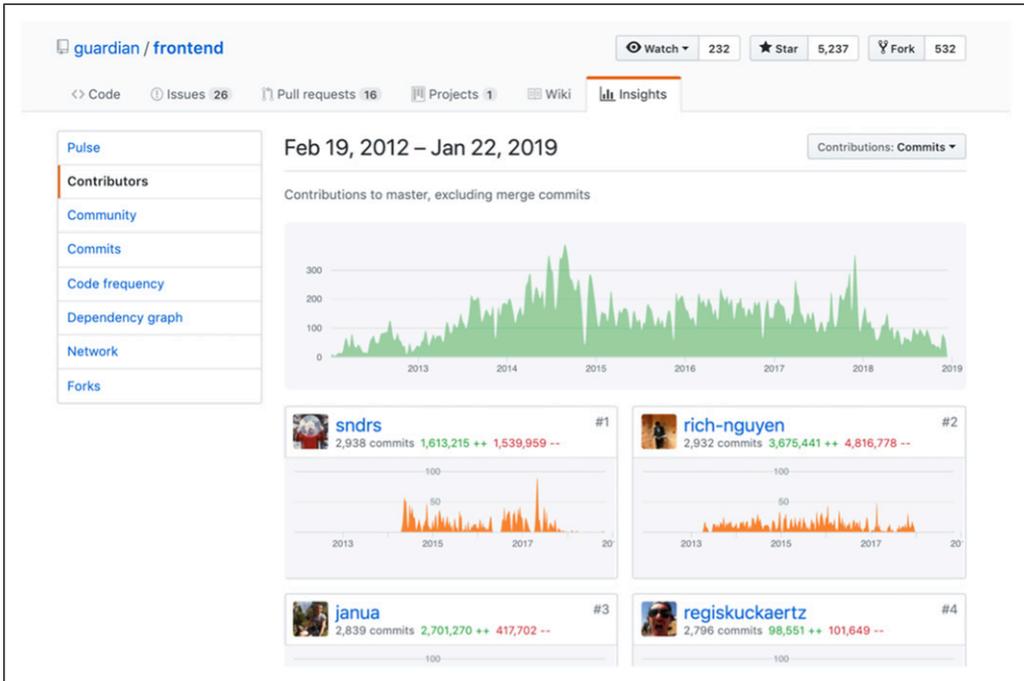


Figure 3. Browser-based user interface view of public GitHub repository showing social counters (top right) and top contributors based on their number of commits.

is done through social metrics and counters such as the ‘forks count’ and ‘stargazers count’ (visible in Figures 2 and 3), rankings such as its ‘trending’ feed, and other calculations and statistics released by the platform, such as through the annual ‘State of the Octoverse’ report.⁵

By making projects and people commensurable through the introduction of common metrics, the platform materialises an auditorial culture (Gane, 2014; Power, 1999; Strathern, 2000) based on quantitative measures, that intensifies evaluation and competition between projects (Rieder, 2017). This environment is generative of reactivity dynamics (Espeland and Sauder, 2007) whereby users modify their behaviours in response to the evaluations that the platform makes available. As platforms are organised around ‘the pursuit of participation’ (Bucher, 2012b: 10), this is a dynamic that platforms welcome and encourage (Gerlitz and Lury, 2014). The ongoing encouragement to gain visibility and perform in ways which are recognised by metrics shapes platform practices and has led to both grassroots responses in the form of how-to guides for increasing visibility by accumulating currency, that is, stars, as well as commercial initiatives that enable the purchase of fake stars and followers.¹⁴

Moreover, the metrification of coding and engagement acts on the platform is accompanied by an intensification and multiplication of social dynamics by means of algorithmic curation, sorting and recommendation (Bucher, 2012b; Gerlitz and Helmond, 2013). For example, GitHub displays user activities on the user’s home page, as well as on the news feeds of her followers. Activity around a repository that a user watches or subscribes to is also displayed in the user’s ‘news feed’. Repositories are recommended for starring and bookmarking based on the user’s own activity on the platform.

All these data streams and recommendation features aim to intensify activity on the platform by inviting further engagement and encouraging certain kinds of actions. In doing so they multiply the collection of economically valuable user data (Gerlitz and Helmond, 2013) and the growth of the platform's connective assets resulting from the conversion of social coding into economically and reputationally valuable connectivity. These connective assets may include platform data and knowledge products, which in turn become part of the platform's offerings to its stakeholders (Gerlitz, 2016; Rieder, 2017). As described above, these may be embedded in various interfaces, for example, as recommendations and trends via the browser-based user interface, as platform data via the application programming interface (API), and annual reports about platform accomplishments geared towards attracting investors and clients.

Connective coding

In light of this analysis of some of the infrastructures and processes by means of which coding is platformised and made economically valuable, GitHub's configuration of software development may be better understood not as social but as connective. Drawing on Van Dijck (2013), the notion of connective coding encompasses not only the connectedness that GitHub enables but also the commodification of software development by 'turning connectedness into connectivity by means of coding technologies' (Van Dijck, 2013: 16). In the context of GitHub, the commodification of software development refers to potential economic capital accumulation by converting public coding activities, developer profiles and behaviours into assets that may attract future revenue and investment to the platform (Mackenzie, 2018), through the platform's technical infrastructures.

These connective assets have the potential to be capitalised by the platform as well as by third parties that make up the platform ecosystem, and to enter various economies and forms of valuation (Gerlitz, 2016). The release of data points via an API enables third parties to derive their own forms of value, be it economic, cultural or social (Gerlitz, 2016). In addition to being used to improve the performance of the platform according to its aims, and to optimise software development processes, due to its large base of public developer profiles and code repositories, GitHub has become a recruitment tool for the technology industry and has come to be seen as a résumé for developers, displaying markers of reputation, productivity and uptake (Fuller et al., 2017; Figuière, 2022). GitHub has also become a key provider of data about the software development sector, supporting a number of startups that provide various kinds of data mining, analytics and recruitment services in this sector.¹⁵

Platformised coding also draws attention to how coding and engagement work also become a form of platform work, in the sense that every coding act also contributes to enacting the platform ensemble. This is illustrative of the asymmetries that characterise online platforms, between the actors who define 'conditions for participation' (Gerlitz, 2016: 19) by setting up a techno-commercial infrastructure (the platform owner), and platform users, as well as between the capacities of datafied users and those of third parties which derive economic value from such datafied user activities (Gerlitz and Helmond, 2013; Lehtiniemi, 2016), such as startups and app developers.

The capture of social dynamics and behaviours around public repositories and the creation of knowledge or information products is not only essential to GitHub's functioning as a marketplace and to its dominant mode of valuation of repositories as connective assets, but also informs platform practices and vernaculars and enables GitHub to function as an ordering mechanism (Rieder, 2017) for public and open-source coding practices. This has social consequences in terms of how visibility of users and repositories is modulated and how attention is guided to repositories and people, what is 'elevated into prominence' (Geboers & Van De Wiele, 2020), and what receives less attention.

The high-visibility platform vernaculars of news and journalism projects

In this section I explore how visibility of media and journalism organisations and projects is modulated by analysing top starred repositories and associated organisations.

At the time of data collection, there were 87 public GitHub accounts from media and journalism organisations and initiatives in the list maintained by *Source*, and 3,665 public repositories associated with these. These repositories were created between 2009 and 2016, with close to 90% of them being created between 2013 and June 2016. The great majority of the organisations owning these repositories are US-based (over 80%), accompanied by a handful of organisations from UK, Canada, Germany, Switzerland, the Czech Republic, Chile, Mexico, Argentina, Kenya and one that self-describes as international.

Close to half of the GitHub accounts are associated with news organisations (local, regional and national ones), such as *The Guardian* and *Baltimore Sun*, followed by magazines (such as *Mother Jones* and *New Yorker*), public radio stations (such as *NPR* or *WNYC*), media groups such as Vox Media and Gannett Digital, television channels and programmes (such as the *Al Jazeera America*, now defunct, and *PBS NewsHour*) and investigative journalism outlets such as *Correctiv* and *ProPublica*. These are complemented by a multiplicity of media entity types appearing only once or twice, including open-source journalism software platforms, data journalism platforms, press foundations, professional associations and networks, arts and media incubators, journalism schools and communities, student news sites and civic technology initiatives (see [Figure 4](#)).

In terms of number of public repositories, the organisations differ greatly, with *The Guardian* having the largest number of repositories at the time of data collection (619), followed by Gannett Digital (252) and *The Texas Tribune* (163), and close to half of the accounts having less than 20 repositories (see [Figure 5](#)). Notable is *The Guardian*'s large volume of public repositories, which is indicative of its software development style. This style is shaped by the newsroom's adoption of openness as an innovation principle applied both to its editorial and to its software products ([Aitamurto and Lewis, 2012](#); [Lewis, 2012](#)). According to the news organisation's Engineering Blog, *The Guardian* develops code in public repositories by default (even for projects that are not open-source), with the exception of sensitive projects which are developed in private repositories.⁶

Not all the repositories published on the organisations' GitHub accounts have been created by them. A feature of public software and project development on GitHub is that through its technical infrastructure based on Git, GitHub enables users to copy code through its fork function. GitHub users can copy repositories by forking them in order to propose revisions to them or to start their own project according to open-source principles.⁷ By automating and simplifying the copying of repositories through the fork button, GitHub institutes a mode of production based on imitation and variation ([Fuller et al., 2017](#)), which is one of the ways through which the platform seeks to intensify relations between its stakeholders.

Close to 80% of the 3,665 repositories studied in this article are created by these 87 organisations (I will call them original repositories) and only 23% of them are copied or forked from other GitHub users and organisation. This is a lower rate than the numbers reported at the level of the entire platform ([Lopes et al., 2017](#); [Mackenzie, 2018](#)). Indeed, several studies have pointed towards the intensity of imitative and duplication work that underpins GitHub code production ([Lopes et al., 2017](#); [Mackenzie, 2018](#)), reporting everything from a third to 70% of code on GitHub to be a duplicate.

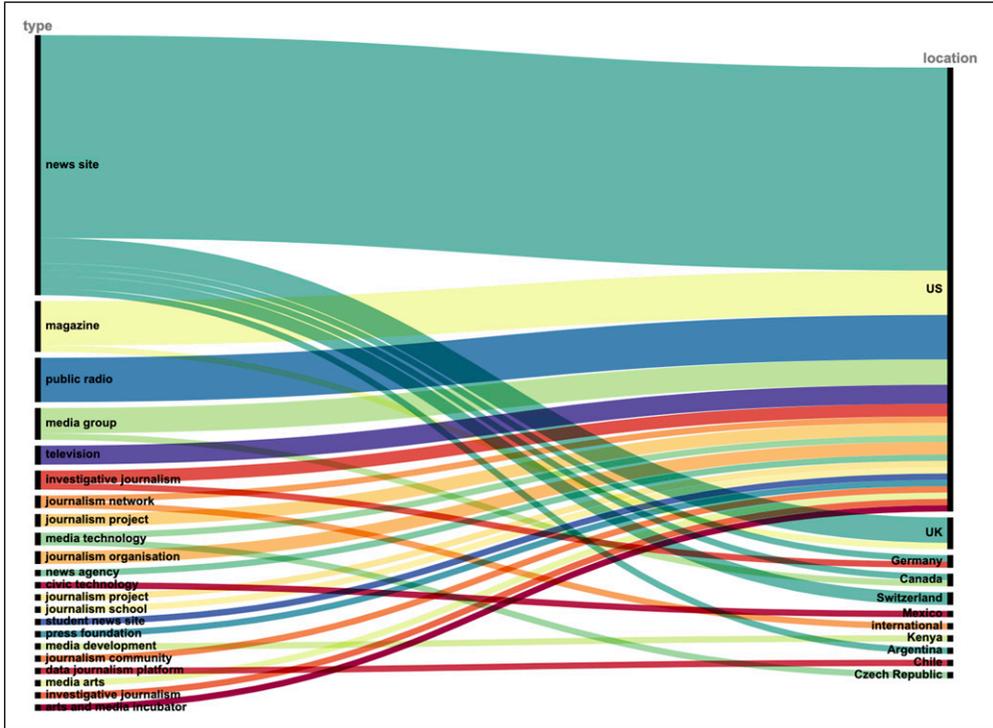


Figure 4. Locations and types of media entities in the list of 87 media and journalism GitHub accounts, sized and ordered by occurrence count.

GitHub account	No. repositories
guardian	619
GannettDigital	252
texastribune	163
seattletimes	145
MinnPost	130
nprapps	119
SCPR	86
statesman	86
datadesk	83
ft-interactive	81

Figure 5. Top 10 news and journalism GitHub accounts by number of repositories at time of data collection.

GitHub account	No. repositories
NYTimes	15
guardian	9
documentcloud	8
bloomberg	7
propublica	7

Figure 6. Top five GitHub accounts owning more than half of repositories with 100+ stars, ranked by number of repositories owned.

The modulation of visibility in the media and journalism repository space

To examine how platform networking features and algorithmic sorting mechanisms modulate the visibility of news and journalism organisations and shape platform vernaculars, I examined a subset of the original repositories which received at least 100 stars at the time of data collection. GitHub users can ‘star’ a repository to bookmark it and generate recommendations, or as an act of public appreciation.⁸

88 repositories created by these organisations receive at least 100 stars. With very few exceptions, these repositories are open-source projects, indicated by the presence of an open-source and/or open content licence. The repositories are generally well-maintained with well-documented ‘README’ sections and links to blog posts, articles or web pages offering more information on the projects. In terms of who gains visibility by means of GitHub’s repository starring practices, it is notable that five organisations own over half of the most starred open-source repositories (see Figure 6).

The five organisations represent a mix of actor types. On the one hand, as with other social media platforms and network structures, large, established and well-resourced organisations (*The New York Times*, *The Guardian* and *Bloomberg*) feature amongst the high-visibility organisations in terms of starred repositories. On the other hand, more recent and smaller actors in the media space, such as *ProPublica* and Document Cloud, also gain visibility in terms of starred repositories. *ProPublica* is an investigative journalism outlet with about 100 journalists on its staff.⁹ Document Cloud is an open-source grant funded project which lists 14 contributors on its GitHub page.

It is not surprising that the high-visibility open-source media and journalism code space features large, established and well-resourced media organisations such as *The New York Times*, *The Guardian* and *Bloomberg*, given that successfully releasing, maintaining and publicising open-source software requires substantial resource and time investments. All five actors have committed to F/OSS in various ways. Several maintain blogs where they document and publicise their open-source work (e.g. *The New York Times*’s Open blog, *The Guardian*’s Engineering blog, *Bloomberg*’s Tech at *Bloomberg* story category, and *ProPublica*’s Nerd blog).¹⁰ *Bloomberg* often publishes press releases to accompany new open-source projects. Private actors such as the *New York Times* and *Bloomberg* not only use and contribute to but also financially support open-source development by means of funds and sponsoring events.¹¹

While they all share a commitment to F/OSS, different ways of doing and engaging with F/OSS can be identified. One style revolves around large private media actors (*New York Times* and *Bloomberg*) engaging with F/OSS practices and communities, which some advocates may view as in tension with the free software movement's community and commons-oriented ethos (Kelty, 2008). Notable is also the tension between these companies' discursive affirmation of open-source values and contributions to open-source projects, and their highly protective practices around their proprietary and profitable products (for more on this point see, e.g. Vasudevan, 2022). Particularly notable here is the tension with *Bloomberg*'s monopolising software and hardware practices in the financial news market sector.

At the same time, while it may be perceived as a tension, engagement in open-source practices is often part of private actors' strategies to build relationships with open-source communities (Otto et al., 2023). Engagement with open-source practices may play several roles for commercial companies and their developers. For both *Bloomberg* and the *New York Times* using and contributing to open-source software are ways to further multiple company goals: firstly, to develop 'good code' and produce software 'smartly' by incorporating already existing software into its products (Otto et al., 2023), and secondly, to invest in the development of software that the company is currently using or may use in the future.¹² In addition to this, releasing open-source code is also a way to enact a sense of being part of a developer community, as well as a reputational marker or what Otto et al. (2023) call 'making a F/OSS-like name' (p. 254).

Another open-source engagement style is visible in the practices of high-visibility non-profit actors in the media and journalism open-source space. Engagement with F/OSS is here shaped by and in alignment with a more general organisational strategy that emphasises openness across multiple organisational practices (Lewis, 2012). For example, *ProPublica* permits their stories to be reused and republished under open content licences.¹³ The *Guardian*, under its 'open journalism' principle, distributes its stories free of charge, encourages reader participation in story production, and enables the development of third-party applications on top of its content by means of public APIs (Aitamurto and Lewis, 2012). Finally, Document Cloud is indicative of a third style of open-source software development style whereby journalists and developers collaborate across organisations to produce software and release it independently of any organisation or select an organisational home based on alignment with the project goals.

When it comes to the high-visibility repositories, these cover a diversity of projects and themes (see Figure 7). More than a third of these repositories are projects pertaining to web development. These include various back-end and front-end software development projects, including the source code for the *Guardian* and the DocumentCloud platform, as well as web analytics libraries and software development conventions. The web development-related repositories are followed by repositories containing datasets, and source code for data collection and analysis tools, making up a quarter of the top starred repositories. For example, *FiveThirtyEight* maintains a repository containing the data and code behind their articles and graphics. There is also a guide to resolving 'bad data' problems, as well as the source code for multiple data visualisation tools.

The third category is repositories sharing source code for tools that support the practice of software development itself. This is source code for various tools and utilities used by developers in the process of creating and maintaining software. Examples of these include testing tools for catching interface bugs, server load testing tools, parallel processing scripts, wrappers and collections of libraries that make up a newsroom's development environment.

The remaining repositories cover a diversity of projects, from various social media tools (including a meme generator), to a whistleblowing platform, virtual reality software, book publishing

Theme	Count
web development	33
data reporting	22
software tools development	11
content management	3
social media tool	3
mobile development	2
API development	2
monitoring system	1
whistleblowing	1
mapping	1
virtual reality development	1
book publishing software	1
anonymous news tip	1
document annotation	1
encryption software	1
quiz making	1
cloud computing development	1
radio software	1
reporting tools	1

Figure 7. Themes in repositories with 100+ stars.

software, anonymous tip submission code, quiz making, encryption and radio station management software.

Across these repositories and themes, GitHub's role as a platform for making news and journalism infrastructures and platforms can be noticed (Mackenzie, 2018), thereby confirming its potential as a site of study for the digitisation and platformisation of news (Nechushtai, 2017). The second category indicates that it is not only open-source practices that gain visibility on the platform but also experiments with open-source practices and principles in the domain of data practices (Kelty, 2008).

Stars are distributed across the 3,665 public repositories following a power law model, whereby a small number of repositories attract the largest amount of stars. In this case close to half of these repositories do not receive any stars and are thus public but receive low visibility as the absence of events leaves them out of the recursive recommendation streams (e.g. trending repositories) in which high event repositories are included.

Lower visibility repositories provide a different view on the public media and journalism space on GitHub. While some of these repositories follow open-source practices of licencing and maintaining clear documentation, others lack a licence, and have little or no documentation or description (see Figure 8). For example, they lack a README page that helps to make the code knowable and usable or this page is empty, and no links to external project documentation are

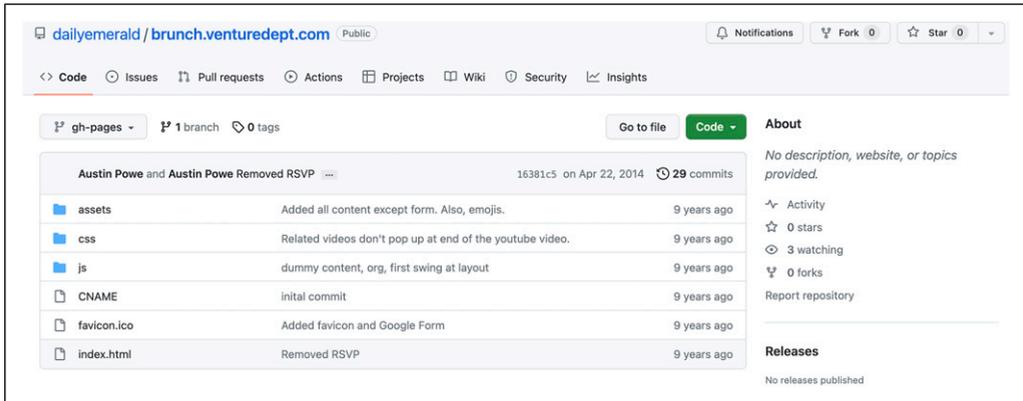


Figure 8. Example of repository illustrative of post-F/OSS culture.

provided. Some repositories are in progress and some have been deleted. These practices are resonant with another aspect of contemporary software culture which co-exists in the GitHub public repository space alongside high-visibility open-source repositories and which Fuller (2017) describes as post-F/OSS. Post-F/OSS is marked by ‘a general indifference to the discussions of and loyalty to certain kinds of licences and the sense of ethics (GPL) or business models (Open-source) that these drew upon’ (Fuller et al., 2017: 58). Alongside well-packaged and well-documented reusable software components one also encounters organisational and infrastructural by-product, remnants, detritus.

Conclusion

GitHub’s dominance in the online software development and F/OSS space and its re-centralisation of software development work (and ongoing expansion into other societal spaces and practices), as well as its acquisition by the technology giant Microsoft, call for more sustained critical engagement with the platform’s impact on F/OSS, public coding and other cultural practices. This article combined an institutional perspective with a platform practices perspective to examine the platformisation processes that underpin GitHub and how the interplays between technical infrastructure and economic imperatives shape platform vernaculars and contemporary programming practices.

The article contributed to software and platform studies by bringing the platformisation of software development and cultural production on GitHub under critical investigation. It examined GitHub’s configuration of social action from a material-economic perspective and proposed the concept of connective coding and software development to characterise GitHub’s dominant modes of configuration and capitalisation of public repositories and profiles and the power relations that underpin it, whereby public software and project development work becomes assets in the platform economy that have the potential to be variously capitalised by the platform and its associated third-party ecosystem.

The article contributed to media and cultural studies by complementing the institutional material-economic analysis with a cultural practices perspective to examine the high-visibility platform vernaculars associated with a set of media and journalism organisations on GitHub. It found that the high-visibility platform vernacular of news and journalism is dominated by a mix of large and established media actors such as the *New York Times*, the *Guardian* and *Bloomberg*, as well as

smaller and more recent actors and initiatives such as *ProPublica* and *Document Cloud*. It also found it to be characterised by multiple styles of participation and engagement with F/OSS, from the incorporation of F/OSS into media commodity production by large companies such as in the case of *Bloomberg* and the *New York Times*, to the extension of an organisational strategy of openness in non-profit media production to the software development space such as in the case of the *Guardian* and *ProPublica*, as well as cross-organisational collaborations to create an independent open-source product. The high-visibility vernacular was dominated by F/OSS projects focused on web development and software tools development, as well as by experiments with open-source principles and practices in areas such as data work. Finally, by contrast, low-visibility public repositories in this space may be seen as indicative of GitHub's role in facilitating what Fuller et al. (2017) have described as a post-F/OSS culture in contemporary software development.

While this article focused on GitHub's platformisation mechanisms that existed prior to its purchase by Microsoft, future research on connective software development may explore post-acquisition platform changes and responses. On the one hand this includes the expansion and intensification of platform metrics and the addition of platform features (e.g. the introduction of AI, machine learning and automation into developer workflows). On the other hand further studies may examine post-acquisition counter-mobilisations and movements to alternative code hosting platforms – and other ways of organising, networking and socialising the development of the software that we live with.

Acknowledgments

I am most grateful to Erik Borra, Emile den Tex and the Digital Methods Initiative for their support in co-developing the DMI GitHub software tools, as well as to Sam Leon for developing custom web scrapers used in research for this article. I would also like to thank Marcel Broersma, Mark Deuze, Carolin Gerlitz, Jonathan Gray, Sarah Van Leuven, Jean-Christophe Plantin, Karin Raeymaeckers, Richard Rogers, Tamara Witschge, Tommaso Venturini and Esther Weltevrede for discussion and valuable comments on earlier versions of this piece. Furthermore, I would also like to acknowledge the contributions of my collaborators in an earlier phase of this project, a pilot study I co-led with Jonathan Gray and Stefania Milan at the Digital Methods Summer School 2015 at the University of Amsterdam on GitHub as transparency device in data journalism, open data and data activism. The project team included: Jonathan Albright, Matteo Azzi, Stefan Baack, Stefano Bandera, Rishabh Dara, Rebeca Diez, Sylvain Firer-Blaess, Ivo Furman, Robert Gutounig, Janna Joceli, Cristel Kopolaking, Lisa Krieg, Lisa Langenkamp, Sam Leon, Sjoukje van der Meulen, Mariola Pagán, Tamara Pinos, Ana Pop Stefanija, Tim Riley, Richard Rogers and Savaş Yıldırım. I would also like to express appreciation to Tommaso Venturini, my collaborator on a related study, "Journalism as a data public: Mapping code ecologies for data valuation on GitHub," presented at Data Publics: Investigating the Formation and Representation of Crowds, Groups, and Clusters in Digital Economies, University of Lancaster, 31 March - 2 April 2017. Finally, I would like to extend my gratitude to the two anonymous reviewers for their valuable suggestions on this work.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

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

ORCID iD

Liliana Bounegru  <https://orcid.org/0000-0003-0198-5158>

Supplemental Material

Supplemental material for this article is available online.

Notes

1. <https://source.opennews.org/about/>
2. <https://wiki.digitalmethods.net/Dmi/ToolDatabase?cat=DeviceCentric&subcat=Github>
3. <https://docs.github.com/en/get-started/quickstart/be-social>
4. <https://help.github.com/articles/github-terms-of-service/#h-api-terms>
5. <https://octoverse.GitHub.com/>
6. <https://www.theguardian.com/info/developer-blog/2014/nov/28/developing-in-the-open>
7. <https://help.github.com/articles/fork-a-repo/>
8. <https://help.github.com/articles/be-social/>
9. <https://www.ProPublica.org/about>
10. <https://open.nytimes.com/>, <https://www.Bloomberg.com/company/values/tech-at-Bloomberg/open-source/>, <https://www.theguardian.com/info/series/engineering-blog>, <https://www.ProPublica.org/nerds>
11. <https://web.archive.org/web/20071030152959/https://open.nytimes.com/>, <https://www.Bloomberg.com/company/values/tech-at-Bloomberg/open-source/>
12. <https://www.Bloomberg.com/company/values/tech-at-Bloomberg/open-source/>, <https://web.archive.org/web/20071030152959/https://open.nytimes.com/>, <https://www.linuxfoundation.org/blog/blog/the-future-of-the-Bloomberg-terminal-is-open-source#>
13. <https://www.ProPublica.org/steal-our-stories/>
14. See, e.g. <https://www.freecodecamp.org/news/how-to-get-more-engagement-with-your-open-sourceproject/> and <https://dagster.io/blog/fake-stars>.
15. See, e.g. <https://techcrunch.com/2023/03/03/prog-ai-wants-to-help-recruiters-find-technical-talent-by-inferring-skills-from-github-code/>.
16. See, e.g., <https://www.silicon.co.uk/software/open-source/GitHub-enterprise-coding-open-source-193037>.
17. See, e.g. <http://uk.businessinsider.com/GitHubs-enterprise-service-brings-in-half-its-200-million-in-sales-2017-10>.

References

- Agre PE (1994) Surveillance and capture: two models of privacy. *The Information Society* 10(2): 101–127. DOI: [10.1080/01972243.1994.9960162](https://doi.org/10.1080/01972243.1994.9960162).
- Aitamurto T and Lewis SC (2012) Open innovation in digital journalism: examining the impact of open APIs at four news organizations. *New Media & Society* 15(2). DOI: [10.1177/1461444812450682](https://doi.org/10.1177/1461444812450682).
- Aradau C, Blanke T, and Greenway G (2019) Acts of digital parasitism: Hacking, humanitarian apps and platformisation. *New Media & Society* 21(12). DOI: [10.1177/1461444819852589](https://doi.org/10.1177/1461444819852589).
- Birkinbine BJ (2020) *Incorporating the Digital Commons*. London, UK: University of Westminster Press.
- Blanke T and Pybus J (2020) The material conditions of platforms: monopolization through decentralization. *Social Media + Society* 6(4). DOI: [10.1177/2056305120971632](https://doi.org/10.1177/2056305120971632).
- Bogost I and Montfort N (2009) Platform studies: frequently questioned answers. Available at: http://nickm.com/if/bogost_montfort_dac_2009.pdf.
- Bucher T (2012a) Want to be on the top? Algorithmic power and the threat of invisibility on Facebook. *New Media & Society* 14(7): 1164–1180. DOI: [10.1177/1461444812440159](https://doi.org/10.1177/1461444812440159).

- Bucher T (2012b) *Programmed Sociality: A Software Studies Perspective on Social Networking Sites*. Oslo, Norway: University of Oslo.
- Bucher T (2018) *If...Then: Algorithmic Power and Politics* New York, NY: Oxford University Press.
- Caliandro A and Gandini A (2016) *Qualitative Research in Digital Environments: A Research Toolkit*. Routledge.
- Dabbish L, Stuart C, Tsay J, et al. (2012) Social coding in GitHub: transparency and collaboration in an open software repository. In: *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work*. Seattle, WA, 2012 February (p. 1277). ACM Press.
- Dolata U and Schrape J-F (2023) Platform companies on the internet as a new organizational form. A sociological perspective. *Innovation: The European Journal of Social Science Research* 0: 1–20. DOI: [10.1080/13511610.2023.2182217](https://doi.org/10.1080/13511610.2023.2182217).
- Duffy BE, Poell T, and Nieborg DB (2019) Platform practices in the cultural industries: creativity, labor, and citizenship. *Social Media + Society* 5(4). DOI: [10.1177/2056305119879672](https://doi.org/10.1177/2056305119879672).
- Espeland WN and Sauder M (2007) Rankings and reactivity: how public measures recreate social worlds. *American Journal of Sociology* 113(1): 1–40.
- Evans DS, Hagiú A, and Schmalensee R (2006) *Invisible Engines: How Software Platforms Drive Innovation and Transform Industries*. Cambridge, UK: MIT Press.
- Figuière N (2022) How to Use GitHub to Find & Hire Developers in 2022. *CodinGame*. Available at: <https://www.codingame.com/work/blog/find-developers/how-to-use-github-to-find-hire-developers/>.
- Fuller M, Goffey A, Mackenzie A, et al. (2017) Big diff, granularity, incoherence, and production in the GitHub software repository. In *How to Be a Geek: Essays on the Culture of Software*. Cambridge, UK: Polity Press.
- Gane N (2014) The emergence of neoliberalism: thinking through and beyond Michel Foucault's lectures on biopolitics. *Theory, Culture & Society* 31(4): 3–27.
- Geboers MA and Van De Wiele CT (2020) Regimes of visibility and the affective affordances of Twitter. *International Journal of Cultural Studies* 23(5): 745–765. DOI: [10.1177/1367877920923676](https://doi.org/10.1177/1367877920923676).
- Gehl RW (2014) *Reverse Engineering Social Media: Software, Culture, and Political Economy in New Media Capitalism*. Philadelphia, PA: Temple University Press.
- Gerlitz C (2016) What counts? Reflections on the multivalence of social media data. *Digital Culture & Society* 2(2): 19–38.
- Gerlitz C and Helmond A (2013) The like economy: social buttons and the data-intensive web. *New Media & Society* 15(8), 1348–1365. DOI: [10.1177/1461444812472322](https://doi.org/10.1177/1461444812472322).
- Gerlitz C and Lury C (2014) Social media and self-evaluating assemblages: on numbers, orderings and values. *Distinktion: Journal of Social Theory* 15(2): 174–188.
- Gerlitz C and Rieder B (2018) Tweets are not created equal: investigating twitter's client ecosystem. *International Journal of Communication* 12: 528–547.
- Gibbs M, Meese J, Arnold M, et al. (2015) #Funeral and Instagram: death, social media, and platform vernacular. *Information, Communication & Society* 18(3): 255–268. DOI: [10.1080/1369118X.2014.987152](https://doi.org/10.1080/1369118X.2014.987152).
- Gillespie T (2010) The politics of 'platforms'. *New Media & Society* 12(3): 347–364.
- GitHub (2022) *Octoverse 2022: The state of open source software*. Retrieved 10 July 2023, from <https://octoverse.github.com/>.
- Gousios G, Vasilescu B, Serebrenik A, et al. (2014) *Lean GHTorrent: GitHub Data on Demand*. New York, NY: ACM Press, pp. 384–387.
- Gray J, Gerlitz C, and Bounegru L (2018) Data infrastructure literacy. *Big Data & Society* 5(2). DOI: [10.1177/205395171878631](https://doi.org/10.1177/205395171878631).

- Helmond A (2015) The platformisation of the web: making web data platform ready. *Social Media + Society* 1(2). doi: [10.1177/205630511560308](https://doi.org/10.1177/205630511560308)
- Hind S, Kanderske M, and van der Vlist F (2022) Making the car ‘platform ready’: how big Tech is driving the platformisation of automobility. *Social Media + Society* 8(2). doi: [10.1177/20563051221098697](https://doi.org/10.1177/20563051221098697)
- Kelty CM (2008) *Two Bits: The Cultural Significance of Free Software*. Durham, UK: Duke University Press.
- Lehtiniemi T (2016) Reorienting datafication? New roles for users in online platform markets. In: *Internet, Policy and Politics Conference*, Oxford, UK: University of Oxford, p. 16.
- Lewis SC (2012) The tension between professional control and open participation: journalism and its boundaries. *Information, Communication & Society* 15(6): 836–866. DOI: [10.1080/1369118X.2012.674150](https://doi.org/10.1080/1369118X.2012.674150).
- Lewis SC and Usher N (2016) Trading zones, boundary objects, and the pursuit of news innovation: a case study of journalists and programmers. *Convergence: The International Journal of Research Into New Media Technologies* 22(5): 543–560.
- Lopes CV, Maj P, Martins P, Saini V, Yang D, Zitny J, Sajani H, and Vitek J (2017) DéjàVu: a map of code duplicates on GitHub. *Proceedings of the ACM on Programming Languages* 1(24): 1–28. DOI: [10.1145/3133908](https://doi.org/10.1145/3133908).
- Mackenzie A (2014) *Where Code Acts: Passionate Imitation, Infrastructures and Making Programming Work*. Stirling, UK: University of Stirling. Available at: https://github.com/metacommunities/metacommunities/tree/master/publications/code_acts_stirling_sept2014.
- Mackenzie A (2018) 48 million configurations and counting: platform numbers and their capitalization. *Journal of Cultural Economy* 11(1): 36–53.
- Mackenzie A, Fuller M, Goffey A, et al. (2014) Metacommunities of practice in the code-sharing commons. Available at: <http://metacommunities.github.io/metacommunities/>.
- Mackenzie A, Mills R, Sharples S, et al. (2015) Digital sociology in the field of devices. In: L Hanquinet and M Savage (eds), *Routledge International Handbook of the Sociology of Art and Culture*. Abingdon, UK: Taylor & Francis Group.
- Marres N (2017) *Digital Sociology: The Reinvention of Social Research*. Cambridge, UK: Polity Press.
- Marres N and Gerlitz C (2016) Interface methods: renegotiating relations between digital social research, STS and sociology. *The Sociological Review* 64(1): 21–46. DOI: [10.1111/1467-954X.12314](https://doi.org/10.1111/1467-954X.12314).
- Meese J, Gibbs M, Carter M, et al. (2015) Selfies|selfies at funerals: mourning and presencing on social media platforms. *International Journal of Communication* 9(0), 14.
- Nechushtai E (2017) Could digital platforms capture the media through infrastructure. *Journalism* 19(8), 1043–1058. DOI: [10.1177/1464884917725163](https://doi.org/10.1177/1464884917725163).
- Nieborg DB and Poell T (2018) The platformisation of cultural production: theorizing the contingent cultural commodity. *New Media & Society* 20(11): 4275–4292. doi: [10.1177/1461444818769694](https://doi.org/10.1177/1461444818769694)
- Nieborg DB, Duffy BE, and Poell T (2020) Studying platforms and cultural production: methods, institutions, and practices. *Social Media + Society*, 6(3). DOI: [10.1177/2056305120943273](https://doi.org/10.1177/2056305120943273).
- Niederer S (2019) *Networked Content Analysis: The Case of Climate Change*. Amsterdam, Netherlands: Institute of Network Cultures.
- Otto EL, Salka JH, and Blok A (2023) How app companies use GitHub: on modes of valuation in the digital attention economy. *Journal of Cultural Economy* 16(2): 242–259. DOI: [10.1080/17530350.2023.2186916](https://doi.org/10.1080/17530350.2023.2186916).
- Plantin J-C, Lagoze C, Edwards PN, et al. (2016) Infrastructure studies meet platform studies in the age of Google and Facebook. *New Media & Society* 20(1): 293–310.
- Poell T, Nieborg D, and van Dijck J (2019) Platformisation. *Internet Policy Review* 8(4). Available at: <https://policyreview.info/concepts/platformisation>
- Power M (1999) *The Audit Society: Rituals of Verification*. New Edition. Oxford, MS: Oxford University Press.

- Rieder B (2017) Beyond surveillance: how do markets and algorithms ‘think’. *Le foucauldien* 3(1): 8. DOI: [10.16995/lefou.30](https://doi.org/10.16995/lefou.30).
- Rieder B, Matamoros-Fernández A, and Coromina Ò (2018) From ranking algorithms to ‘ranking cultures’: investigating the modulation of visibility in YouTube search results. *Convergence: The International Journal of Research Into New Media Technologies* 24(1): 50–68. DOI: [10.1177/1354856517736982](https://doi.org/10.1177/1354856517736982).
- Rochet J-C and Tirole J (2003) Platform competition in two-sided markets. *Journal of the European Economic Association* 1(4): 990–1029.
- Rogers R (2013) *Digital Methods*. Cambridge, UK: MIT Press.
- Rogers R (2017) Otherwise engaged: from vanity metrics to critical analytics. *International Journal of Communication* 12: 450–472.
- Rogers R (2019) *Doing Digital Methods*. SAGE.
- Srnicek N (2016) *Platform Capitalism*. Cambridge Malden, MA: Polity Press.
- Strathern M (ed), (2000) *Audit Cultures: Anthropological Studies in Accountability, Ethics and the Academy*. 1st Edition. London, UK: Routledge.
- Usher N (2016) *Interactive Journalism: Hackers, Data, and Code*. Champaign, IL: University of Illinois Press.
- van Dijck J (2013) *The Culture of Connectivity: A Critical History of Social Media*. London, UK: Oxford University Press.
- van Dijck J, Poell T, and Waal MD (2018) *The Platform Society: Public Values in a Connective World*. New York, NY: Oxford University Press.
- Vasudevan R (2021) Share and share unlike: reciprocity, corporate power, and the open-source ethos. *New Media & Society* 25(11). DOI: [10.1177/14614448211038350](https://doi.org/10.1177/14614448211038350).
- Venturini T, Bounegru L, Gray J, et al. (2018) A reality check(list) for digital methods. *New Media & Society* 23. DOI: [10.1177/1461444818769236](https://doi.org/10.1177/1461444818769236).

Author biography

Liliana Bounegru is Lecturer in Digital Methods at the Department of Digital Humanities, King’s College London. She is a research associate at the Digital Methods Initiative (University of Amsterdam) and Sciences Po Paris médialab. She co-founded the Public Data Lab, a network of research labs which seeks to facilitate research, democratic engagement and public debate around the future of the data society. Previously she was a postdoctoral research fellow at the Oxford Internet Institute (University of Oxford) and obtained her double PhD degree (cum laude) from the University of Groningen and Ghent University.