Open-Source Repositories as Trust-Building Journalism Infrastructure: Examining the Use of GitHub by News Outlets to Promote Transparency, Innovation, and Collaboration

Rodrigo Zamith, Journalism Department
University of Massachusetts Amherst
E-mail: rzamith@umass.edu. Twitter: @rodzam
ORCID: 0000-0001-8114-1734

Abstract

This study incorporates the concepts of transparency, innovation, and collaboration within a broader analytic lens of trust-building infrastructure and applies that lens to an examination of the use of GitHub by 124 prominent news outlets over more than a decade. It finds that (a) their use of GitHub is not widespread but several outlets do actively use it; (b) they use GitHub to open-source a mixture of technologies and journalistic materials; (c) their introductory project documentation routinely includes at least partial amounts of both ambient and disclosure forms of transparency, but rarely exhibits participatory transparency; (d) collaboration is almost non-existent in the vast majority of their repositories; and (e) there has been a decline in their use of GitHub and the collaboration affordances within their repositories in recent years. The study extends the transparency literature by adapting key concepts to journalism-adjacent infrastructure and offers empirical evidence about the innovativeness of open-source technologies originating from prominent news organizations and the amount of collaboration that occurs around them. This builds to an intervention that raises some questions about the direct impact of open-source repositories as trust-building infrastructure while drawing attention to less-considered but nevertheless useful performative functions that such infrastructure enables.

Keywords

GitHub, open-source, infrastructure, transparency, innovation, collaboration, trust, journalism

Acknowledgements

The author would like to thank Michael Karlsson, the guest editors of the special issue, and the two anonymous reviewers for their insightful critiques and recommendations.

Funding Details

This work did not receive funding from a grant-awarding body.

Disclosure Statement

No potential competing interest was reported by the authors.

OPEN-SOURCE REPOSITORIES AS TRUST-BUILDING JOURNALISM INFRASTRUCTURE

Scholars and practitioners alike have devoted considerable attention to the notion of trust in recent years, especially in light of a sustained decline in trust in journalism among large segments of the population around the world. This has led to examinations of concepts like transparency (e.g., Karlsson, 2021; Vos & Craft, 2017), innovation (e.g., Belair-Gagnon & Steinke, 2020; Lowrey, Sherrill, & Broussard, 2019), and collaboration (e.g., Bunce, Wright, & Scott, 2018; Müller & Wiik, 2023) in relation to the fostering of trust and mutually beneficial relationships. Moran and Nuchushtai (2023) in particular have called for examining trust through the lens of infrastructure by evaluating how trust is embedded throughout the systems and technologies that enable and constrain journalistic practices and performances.

Such inquiries tend to focus on traditional infrastructure (e.g., news websites) and infrastructure that has been integral to journalism for some time now (e.g., social media), as well as on the outputs of journalistic labor produced through said infrastructure (e.g., online stories). Examinations of infrastructure that exist at the margins of journalism are far less common, even though such infrastructure may enable new practices and performances and can be highly relevant to certain groups (e.g., peers within a subfield).

Open-source repositories have received scant attention within journalism studies. This includes GitHub, the world’s largest platform for hosting open-source repositories and organizing open-source labor. Scholars have found that GitHub has been used by at least some news outlets to explain journalistic practices, share innovations, and invite collaboration from different kinds of actors (Boyles, 2020a; Haim & Zamith, 2019). GitHub, and open-source practices more broadly, have been especially connected to data journalism, a form of journalism that stresses openness and transparency as mechanisms for increasing journalistic rigor, accountability, and public trust (Camaj, Martin, & Lanosga, 2022; Zamith, 2019). Such infrastructure can be used to overcome obstacles in obtaining public data, facilitate collaborative reporting projects, and even enable more advocacy-minded journalistic role enactments (Martin, Camaj, & Lanosga, 2022). However, scholarly examinations have not systematically evaluated how industry-leading actors in journalism are using such sociotechnical systems and spaces, or directly related them to trust-building infrastructure.

The present study thus incorporates the concepts of transparency, innovation, and collaboration within a broader analytic lens of trust-building infrastructure. It applies that lens to a part-computational, part-human analysis of the use of GitHub by 124 prominent news organizations that are of particular interest to audiences in the United States. The analysis tackles four main research objectives. First, it identifies *which* and *how many* of those outlets use GitHub in a public-facing manner. Second, it examines which *types of innovations* and *forms of transparency* manifest themselves in those organizations' use of GitHub, both in terms of the kinds of projects they publish and how they introduce those projects. Third, it evaluates *the extent of collaboration* involved in those projects. Finally, it assesses *how the use of GitHub by news organizations and their collaborators has changed over time*, both in terms of publishing new repositories and the amount of collaboration they generate.

In addressing those objectives, this study adds to the literature on transparency by translating Karlsson's (2010, 2020) conceptualizations to fit a new context and introducing the Performative Transparency Model (Karlsson, 2021) to infrastructure that is generally seen to be situated outside of journalism. It also contributes to the literature on innovation and collaboration by identifying GitHub as a potential permanent 'trading zone' (see also Haim & Zamith, 2019; Weber & Kosterich, 2018) while offering empirical evidence about the innovativeness of open-source technologies originating from prominent news organizations and the amount of collaboration that occurs around them. This builds to an intervention that raises some questions about the *direct* impact of open-source repositories as trust-building infrastructure while drawing attention to less-considered but nevertheless useful performative functions that such infrastructure enables for signaling values and bonding social groups.

Literature Review

Trust and Infrastructure

Moran and Nechushtai (2023) critique the reception-oriented paradigm of examining trust in journalism for its failure to interrogate questions pertaining to how journalistic processes, norms, and valuations are implicated in the sustenance (or loss) of trust, especially during periods of instability and change (see also Usher, 2018). To address this shortcoming, they recommend conceptualizing trust as an important part of the sociotechnical and physical infrastructure that makes journalism possible and broadly accessible. They argue that "trust is embedded into every activity, personnel, and object associated with news media" (Moran & Nechushtai, 2023, p. 458).

An infrastructure lens is helpful for relocating the scholarly interrogation of trust toward the embedded practices, technologies, and labor involved in doing journalism. It examines both internetworked sites of activity as well as the interacting dependencies among the many sociotechnical actors, actants, and activities involved within and across those sites (Plantin, Lagoze, Edwards, & Sandvig, 2018). An infrastructure lens thus emphasizes an interrogation of both the affordances and constraints of a site of activity (or the networks made up by such sites) as well as the critical role of the infrastructure's human elements, such as how human actors navigate sociotechnical possibilities and limitations and to what ends (Plantin et al., 2018).

Moran and Nechushtai (2023) add that trust is a resource that can be developed and exploited within the sociotechnical and physical infrastructure that supports journalism at multiple stages of newsmaking. It is therefore important to examine not only how journalistic outlets (re)shape their products to increase trust but also how they utilize existing infrastructure to engage in activities and foster relationships that can be reasonably expected to increase trust. One way to do this is by examining how journalistic outlets utilize platforms generally seen to be situated *outside* of traditional journalism to advance transparency, innovation, and collaboration. These three concepts are not only viewed by practitioners and scholars alike as being increasingly important for supporting journalism's core mission but have also been rhetorically constructed—with some empirical support—as key avenues for increasing trust in journalism (Boyles, 2020b; Curry & Stroud, 2021; Vos & Craft, 2017).

Transparency and Trust

While scholars have defined *transparency* in different ways, those definitions can usually be boiled down to the general idea of promoting 'openness' (Heim & Craft, 2020; Karlsson, 2010). Karlsson (2010, 2020) distinguishes between three forms of transparency. The first form, *disclosure transparency*, refers to the extent to which journalistic actors are open about how news is produced and their organizational standards, with the objective of helping make journalistic processes discernible to outsiders. However, as Karlsson (2010) notes, disclosure transparency concerns itself primarily with communicating *to* an audience and not *with* them. It includes explaining news selection and analytic processes. The second form, *participatory transparency*, aims to involve audiences in news production processes in different ways. This includes inviting audience input into what an ongoing news investigation should focus on and asking them for help on interpreting information. The third form, *ambient transparency*, involves journalists "add[ing] information around the edges of news stories" (Karlsson, 2020, pp. 1808–1809) but not necessarily incorporating audiences within the frame of the content. This includes connecting audiences to source documents through affordances like hyperlinks or adding signals that help individuals make sense of the content (e.g., adding a clear label to distinguish native advertising from a news story).

Karlsson (2021) proposes the Performative Transparency Model to help scholars examine how transparency is enacted. The model includes four building blocks: the stage, actors, script, and aesthetics and delivery. The stage is the site where transparency occurs, and it must be durable in time, spreadable in space, and enable reciprocity. In the context of journalism, the actors occupying the stage include at minimum journalists and their publics, but may also include peers, critics, and regulators (among others). The script refers to the guidelines and shared expectations—rarely codified but negotiated over time through professional claims, popular depictions, and social interactions—of what constitutes transparent and non-transparent performances. The aesthetics and delivery cover how transparency measures are translated into language, signs, and elements that are accessible to and easily interpreted by other actors to reduce information asymmetry. Ultimately, as Karlsson (2021) contends, a transparency performance must be convincing to be useful. This does not mean it needs to *improve* an outcome (e.g., increase trust); sometimes, it need only be convincing enough to *maintain* a status quo in the face of instability, such as a crisis of trust.

Much of the empirical work examining journalistic transparency has focused on two streams of inquiry. The first stream examines the ways through which journalistic outlets explain their processes and include audience input within journalistic products. The second stream concerns itself with the ways in which journalists aim to engage directly with audiences beyond journalistic products. The empirical evidence regarding the oft-hypothesized positive relationship between increased transparency and greater trust (or, at least, increased perceived credibility) is decidedly mixed, though. While some scholars have found support for that relationship (e.g., Curry & Stroud, 2021; Masullo, Curry, Whipple, & Murray, 2021), others have found those effects to be practically insignificant or heavily qualified (e.g., Henke, Leissner, & Möhring, 2020; Karlsson, 2020; Karlsson & Clerwall, 2018).

While those findings are valuable, they focus on transparency within journalism's traditional products (e.g., news stories) or within a singular (though key) piece of journalistic infrastructure: social media. There is considerably less scholarship examining the application of transparency in other parts of the infrastructure supporting journalism. This is important given the number of practitioners who are either tasked with engaging in the "labor of trust" or see that endeavor as part of their role orientation (Zahay, Jensen, Xia, & Robinson, 2021, p. 1055). Within this less-explored stream of work, Moran (2021) and Bunce and colleagues (2018) have examined how some newsrooms used the business communication platform Slack to create 'virtual newsrooms' that were open to audiences. Their findings were generally consistent in that while the studied journalists believed the platform's affordances enabled multiple forms of transparency to be enacted, transparency was ultimately "limited both by its questionable public desirability and by the nature of Slack," with "the actual enactment of transparency requir[ing] readers undertake the bulk of the labor" (Moran, 2021, p. 14).

Innovation and Collaboration

The concept of *innovation*—new ideas, products, or ways of doing things within or outside organizations—has received considerable attention in journalism studies in recent years, especially as new technologies lowered the barriers for software development, as consumer hardware became sufficient to perform advanced journalistic tasks, and as technologists and developers became more common (and integrated) within newsrooms (Belair-Gagnon & Steinke, 2020; Zamith & Braun, 2019). A journalistic innovation can focus on improving a particular aspect of journalism or engage with many of its stages; it can have a broad, feature-rich scope or a narrow, highly specialized scope; it can focus on improving operations within a specific context (e.g., a single organization) or be abstracted to be useful across multiple contexts; it can target specialized (e.g., technically adept) or general audiences; and it can originate both within traditional journalistic institutions or from actors that have traditionally operated outside it (Belair-Gagnon, Holton, & Westlund, 2019; Holton & Belair-Gagnon, 2018; Lowrey et al., 2019).

While some journalistic innovations are the product of singular actors, many are the outgrowth of *collaborations* among actors operating within the same field, among actors originating from distinct fields, and among actors that actively seek to cross fields (Belair-Gagnon & Steinke, 2020). Open-source ethos has been of particular interest to scholars who study collaboration around technological innovation in journalism, especially within boundary-crossing contexts (Boyles, 2020b; Lewis & Usher, 2013; Usher, 2016). Open-source ethos is rooted in the ideals of transparency, collaboration, and the wisdom of the crowds, and is frequently enacted through sharing computer code, inviting participation from different actors, and making it possible for users to identify and even correct errors in some work (Usher, 2016). This ethos is also frequently connected to data journalism because of its congruence with data journalists' role conceptions, narrations, and performances (Camaj et al., 2022). Martin and colleagues (2022, p. 13) observe that "many data journalists see openness to collaboration as a shift in culture in their field" and that they use platforms like GitHub to publish code and datasets as a way of expressing "open-data activism"—especially in countries with poor public data transparency infrastructure. However, Zamith's (2019) analysis of the 'typical' data journalism produced by *The New York Times* and *The Washington Post* found that collaboration was the exception rather than the norm, and that relatively few stories provided direct access to the supporting data or code.

The most visible manifestation of open-source ethos is the adoption of an open-source license that explicitly allows a project to be reused and modified under certain conditions while relinquishing some copyright privileges. These licenses are typically categorized as permissive or copyleft. Permissive licenses, which include the MIT and Apache 2.0 licenses, provide the most freedom for personal and commercial reuse and modification. Copyleft licenses, which include the GNU General Public License, are relatively more restrictive because they often require at least a portion of a derivative to be released under the same licensing terms. There is some evidence that the type of license associated with a project—and sometimes, even just the inclusion of an open-source license—impacts the project's active lifespan and the amount of collaboration it receives (Almeida, Murphy, Wilson, & Hoye, 2019).

Collaborative development of an innovation within the context of journalism often involves frictions and tensions. For example, it is often challenging for collaborators with distinct backgrounds to communicate with one another, and they may not share important values or ways of working (Lewis & Usher, 2014). Journalistic actors in particular sometimes get caught between their own desire to be open and institutional traditions that promote opaqueness (Usher, 2016) or struggle internally with the tension between professional control and participation (Lewis, 2012). And, in many cases, an innovation simply isn't adopted by others or the desire for collaboration never materializes, resulting in a feeling of isolation among innovators (Boyles, 2016; Haim & Zamith, 2019). Many journalistic innovations and efforts to collaborate therefore ultimately fail. Weber and Kosterich (2018) have highlighted the need for more permanent 'trading zones' where heterogeneous actors can regularly collaborate in order to more meaningfully impact journalistic spaces.

Innovation and collaboration around journalism do not have to be limited to technology. Organizations like Airwars, Bellingcat, and Forensic Architecture have used existing tools and infrastructure to enable novel approaches to investigative journalism and to establish themselves as trustworthy actors through their enactment of open-source ethos (Müller & Wiik, 2023). This includes leveraging existing infrastructure on the margins of journalism to not only make their journalistic methods more transparent but to help recruit individuals with specialist competencies and gain legitimacy within the field of journalism (Müller & Wiik, 2023). Innovation can also spur from challenges to collaboration, as in the cases of large-scale investigations such as the Panama Papers and Football Leaks, where collaborators developed and open-sourced tools like secure search engines to enable journalists within the network to locate and share information (Larrondo-Ureta & Ferreras-Rodríguez, 2021).

GitHub as a Case Study

The intersection of transparency, innovation, and collaboration, both within and outside of journalism, manifests itself prominently on open-source development platforms (Haim & Zamith, 2019; Usher, 2016). While there are multiple such platforms, GitHub has arguably been the center of networked, open-source development for at least the past decade. The platform builds on the distributed version control affordances of the widely used open-source software Git and adds social networking-like functionality. GitHub offers free and paid tiers that allow individuals and organizations alike to host multiple projects, invite internal and external collaborators, and document information. GitHub reported having over 100 million users on its platform—ranging from individual hobbyists to institutional developers working at Fortune 500 companies—and more than 375 million public repositories as of February 2023.

Activity on GitHub is generally oriented around ‘repositories’ (see Figure 1). These are distinct spaces that are managed by a single account or a group of collaborators and usually contain a specific project. Each repository contains its own set of files, a complete chronicle of changes to each file, and multiple affordances designed to facilitate collaboration. Every repository is expected—but not required—to contain a text file named “README” that describes the repository contents and provides relevant information. When a repository is created, the accountholder is also encouraged (but not required) to provide a short description of the repository’s contents and either select a preformatted “LICENSE” file that describes the copyright restrictions (if any) or provide a custom LICENSE file of their own.

GitHub developers publish a variety of materials on the platform, such as source code for internal production tools, digital notebooks that detail data analyses, and code used to render multimedia projects. There are multiple ways in which individuals can participate in the projects. Users with limited technical aptitude can ‘star’ a project to signal that they consider it to be of value or use GitHub’s issue-reporting affordance to indirectly contribute by describing a bug they encountered, detailing a feature request, or suggesting corrections. Technically adept users can contribute directly to existing projects by submitting a public ‘pull request’ that contains changes to files in the repository, which the repository maintainer(s) can accept or reject. Such users can also ‘fork’ a project (create an associated copy that they can independently modify or extend).

Through these affordances, open-source repositories on collaborative development platforms can potentially serve as trust-building journalism infrastructure by increasing ambient, disclosure, and participatory transparency; by offering accessible, free, and highly visible spaces for developing and distributing journalism-enhancing innovations; and by making it relatively easy for a variety of actors with different levels of technical aptitude to collaborate and contribute in a semi-structured way.

Despite GitHub’s general popularity, there has been relatively little systematic examination of the extent of its actual use within the context of journalism. Boyles (2020a) analyzed the GitHub repositories of seven high-profile North American news outlets and found projects that aimed to improve user experience and accessibility, heightened existing product functionality, crafted new interfaces and platforms, and promoted newsroom productivity tools. However, Boyles also found that although newsroom developers invited collaboration and discussed their work through the language of collaboration, there was limited evidence of actual collaboration occurring around those projects. Nevertheless, the newsrooms saw an active presence on GitHub as a status marker (Boyles, 2020a). Haim and Zamith (2019) examined a random sample of repositories that included the terms ‘news’ or ‘journalism’ and found that most projects were initiated by non-traditional actors (e.g., independent coders and educational institutions). They found that projects focused on providing technological solutions to challenges associated with news distribution and sought to make journalism more transparent by sharing source documents. They also observed that the median developmental lifespan of a project was 17 weeks, offering ample opportunity for collaboration. However, like Boyles (2020a), Haim and Zamith (2019) found little evidence of collaboration. They also observed that less than one-quarter of the examined repositories included a copyright license, which they critiqued as a potential deterrent to collaboration.

Research Questions and Hypotheses

While there has been extensive writing around the concepts of transparency, innovation, and collaboration within the context of journalism, there is little empirical evidence to support extensive hypothesizing within the context of GitHub. For example, while it would stand to reason, based on theory, that news organizations would use GitHub to advance different forms of transparency, it is unclear which forms of transparency are most (or best) enacted in practice. As such, this study aims to contribute empirical evidence by positing the following research questions and hypotheses:

RQ1: How widespread is the use of GitHub among prominent news outlets that appeal to U.S. audiences?

RQ2: What kinds of original projects do those organizations typically publish on GitHub and what is the main project scope, use-case scenario, and target audience of the technological innovations?

RQ3: What forms of transparency manifest within the repositories for those projects?

H1: Most original repositories will involve little collaboration, namely by (a) being forked fewer than 10 times; (b) receiving fewer than 25 stars; (c) having no pull requests from external accounts; and (d) having no issues reported by external accounts.

H2: The average original repository will have an active lifespan of roughly 17 weeks.

H3: Most original repositories will not include a copyright license.

RQ4: Has the use of GitHub by prominent news outlets become more or less prevalent over time?

Method

Sampling

A purposive sampling strategy was used to identify 124 news outlets that appeal to U.S. audiences. These outlets were previously classified by the Pew Research Center as having sizable audiences based on Pew's analysis of digital audience data from Comscore Media Metrix, which included news-related websites appearing in 11 content categories, as well as newspaper circulation data (for a detailed methodology, see Pew Research Center, 2021). The author deviated from Pew in a handful of instances by grouping closely related outlets (e.g., CBS and CBS News). While this strategy excluded some news outlets that make extensive use of GitHub (e.g., *ProPublica*) and some relevant international outlets (e.g., *Financial Times*), it does offer a representative overview of major news organizations that U.S. audiences routinely access.

The author then performed a mixture of URL-based and name-based searches on both GitHub and Google to identify which of those organizations had a GitHub account. Those searches only sought out institutional accounts (i.e., official organizational accounts). Several organizations had multiple institutional accounts, and all eligible accounts were included. Personal accounts (e.g., employee accounts) were excluded because the lack of a directory of such accounts was expected to produce systematic sampling biases.

API Information Retrieval

The first stage of data collection was computational. The author developed a computer script to access the GitHub application programming interface (API) and gather information about all identified accounts and all their public repositories on August 1-4, 2022. The API was used to collect information about the following key variables, among others: the *number of repositories* each organization had; the *date of creation* and *date of last push* for each repository; the *number of stars* and *forks* associated with each repository, as well as the listed *copyright license*; and extended information about each *contributor*, *pull request*, *issue*, and *fork* associated with each repository. To ensure the repositories would have sufficient time to grow and benefit from collaboration, only the repositories created by the end of 2021 were retained for analysis.

Because many organizations did not associate their institutional accounts with individual user accounts (i.e., they did not list their “members”), this study utilized a conservative classification strategy to consistently label contributors as being either internal (members of the organization) or external (non-members) in relation to individual repositories. Specifically, a contributor was classified as being internal if they either committed a change to at least three different repositories associated with the organization or if they accounted for at least 5% of the total commits to all the organization’s repositories. When an organization listed its members, they were all classified as internal contributors. Accounts identified as bots were excluded from analyses.

Content Analysis

The second stage of data collection involved a manual content analysis of a subsample of repositories. The codebook was generated in a semi-inductive manner, with the author drawing on prior work and refining the variable operationalizations by reviewing two randomly sampled original repositories (i.e., not 'forked' from another account) from each organization.

The first set of variables pertained to the project characteristics. The categories for the *project type* variable were drawn from Haim & Zamith (2019) but modified to fit the present context. They included: 'News Production Materials' (e.g., R notebooks detailing data analyses and source code for interactive web content), 'News Production Technology' (e.g., government data retrieval programs and data visualization tools), 'News Distribution Technology' (e.g., load balancers for web servers and web styling templates), 'News Interaction Technology' (e.g., mobile news apps and chatbots), 'General-Purpose Technology' (e.g., brainstorming tools and database management helpers), 'Education and Events' (e.g., resources for hackathons and organizational standards documents), 'Other,' and 'Unclear.'

An additional three innovation-related variables were evaluated for projects coded as technologies: *project scope*, *project use-case*, and *target audience*. Project scope referred to whether this was a 'Minor' project (added a small amount of functionality to an existing technology or served as an original technology that performed simple tasks or offered modest affordances) or a 'Major' project (added a substantial amount of functionality to an existing technology or served as an original technology that tackled complex tasks or offered significant new affordances). Project use-case referred to whether the project primarily targeted, benefited, or was communicated in terms of its ‘Internal’ usefulness (i.e., to the account-holder’s organization) or its ‘External’ usefulness (i.e., the general public or to other organizations). Target audience evaluated whether the main beneficiary or apparent audience for the project were ‘Non-Technical’ individuals (e.g., journalists, editors, designers, or general news audiences) or ‘Technical’ individuals (e.g., software developers or system administrators) based on the project’s stated purpose and accessibility. In instances where the project required technical know-how but targeted non-technical groups (e.g., data journalists who use R), the variable was coded as ‘Non-Technical.’

The second set of variables pertained to the forms of transparency. While Karlsson’s (2010, 2020) work focused on online news articles, their conceptualizations can be applied to how projects are presented on GitHub and introduced through the README file (see Figure 1). Nine elements associated with Karlsson’s three forms of transparency were coded independently on a yes or no basis. For *ambient transparency*, this included the presence of a short, informative description of the project in the ‘About’ section of the page; the association of the project with a URL or the inclusion of relevant external hyperlinks within the README file; and the inclusion of visual badges that denote the project status or aspects of its contents, tags that identify releases or milestones, or package information for the project. For *disclosure transparency*, this included a description of the project within the README file that offered a clear sense of the contents or purpose of the project; a clear description of how the project works or came together, or a listing or application of its major features; and a clear description of how to make use of the project contents. For *participatory transparency*, this included the presence of a general statement or clear heading that invited audiences to participate in some fashion; information about how to contact an individual or group to learn more about the project, ask questions, report problems, or become otherwise involved; and specific instructions for how others may/should contribute, or encouragement to contribute in a specific way.

Intercoder reliability was assessed by having two coders double-code a random sample of 100 eligible repositories. Such testing yielded Krippendorff's alpha coefficients ranging from 0.84 (*project type*) to 1 (multiple variables). Upon establishing acceptable reliability, the full content analysis was carried out on a stratified random sample of 15 original repositories from each eligible organization to ensure broad representation. If an organization had fewer than 15 original repositories, all their repositories were coded. A total of 737 repositories were manually analyzed.

The list of organizations, data collection scripts, codebook, final datasets, and the data analysis report are available at https://github.com/rodzam/open_source_repos_as_trust_building_journalism_infrastructure.

Findings

Use of GitHub

The first research question asked about the extent of the use of GitHub among prominent news outlets that appeal to U.S. audiences.

Out of the 124 organizations examined, 77 (62.1%) had at least one institutional account on GitHub. However, nine of those organizations did not have a single public repository. In other words, the GitHub account associated with their organization was either a placeholder or all development occurred within private repositories. Another three organizations did not have a single original (i.e., not forked from another account) repository.

There were several organizations that made extensive use of GitHub, though. Collectively, there were 6,827 repositories, 5,342 (78.2%) of which were original. Forty-three of the organizations (34.7%) had more than 10 original repositories in the dataset, with organizations like *The Guardian*, BBC, and *The Seattle Times* having the greatest number of such repositories (see Figure 2). Among those 43 outlets, the median number of original repositories was 52 and the median number of all repositories (forked and original) was 73. Thus, the organizations that were active on GitHub used it both to develop original projects and to connect with projects initiated by other actors.

Innovation and Forms of Transparency

The second research question asked about the kinds of original projects those organizations published on GitHub and the main project scope, use-case scenario, and target audience of their technological innovations.

Of the 737 repositories that were sampled and manually coded, the main project types were News Production Materials (*n* = 260, 35.3%), General-Purpose Technology (*n* = 149, 20.2%), News Production Technology (*n* = 130, 17.6%), News Distribution Technology (*n* = 83, 11.3%), Education and Events (*n* = 24, 3.3%), and News Interaction Technology (*n* = 20, 2.7%). A total of 71 (9.6%) repositories were coded as either Other or Unclear. In other words, 51.8% of the repositories pertained to some kind of technology and 38.5% pertained to materials intended to shed light on news products or otherwise educate different audiences.

The vast majority of the 382 technology-oriented repositories had a minor project scope (*n* = 345, 90.3%). These included many "starter kits" designed to preload existing technologies and assets to get a project off the ground faster, as well as productivity tools that made it easier to automate simple tasks, access external services in a simplified way, and add specialized functionality to existing technologies. Just 37 (9.7%) repositories had a more ambitious, major project scope. These included a new markup language (with multiple associated tools) designed to yield more portable journalistic content, a utility that made it possible to simultaneously transcode videos into multiple formats, and a static site generator for multimedia news projects. There was a nearly equal split in the use-case scenarios, with 204 (53.4%) being primarily oriented toward internal use (i.e., within the organization) and 178 (46.6%) having clear external applications (i.e., beyond the organization). Most of these technology-oriented repositories targeted technical audiences (*n* = 235, 61.5%), such as system administrators and database specialists. More than one-third (*n* = 147, 38.5%) targeted non-technical audiences, such as journalists, editors, and news users.

The third research question asked about the forms of transparency that manifested within the repositories for those projects.

As shown in Figure 3, two forms of transparency manifested themselves a substantial portion of the time. There was at least one element of ambient transparency in 87.8% of the 737 repositories analyzed, with all possible elements appearing roughly one-third of the time. Similarly, there was at least one element of disclosure transparency in more than three-fourths of the repositories and all three elements appeared more than one-third of the time. However, participatory transparency rarely manifested itself in the repositories, with just 13.2% of repositories having a single element of participatory transparency.

Collaboration and Licensing

The first hypothesis posited that most original repositories would involve little technical and non-technical collaboration, namely by (a) being forked fewer than 10 times; (b) receiving fewer than 25 stars; (c) having no pull requests from external accounts; and (d) having no issues reported by external accounts.

Of the 5,342 original repositories in the data, 93.8% (*n* = 5,009) were forked fewer than 10 times, 92.6% (*n* = 4,946) received fewer than 25 stars, 80.5% (*n* = 4,298) had no pull requests from an external account, and 85.5% (*n* = 4,569) had no issues reported by an external account. The first hypothesis was therefore supported.

However, some repositories did involve a substantial amount of collaboration. A total of 87 (1.6%) original repositories were forked at least 50 times, with *FiveThirtyEight*'s 'data' repository being forked a remarkable 10,556 times. Additionally, 160 original repositories (3.0%) were starred at least 100 times, with *FiveThirtyEight*'s 'data' repository again leading the way with 15,666 of them. In terms of direct technical collaboration, 127 (2.4%) original repositories received at least 25 pull requests from external accounts, with *The Guardian*'s 'frontend' repository receiving an impressive 4,683 such requests. In terms of direct non-technical collaboration, 82 (1.5%) original repositories had at least 25 issues reported by external accounts, with the BBC's 'simorgh' repository receiving a notable 2,288 such reports.

The second hypothesis posited that the average original repository would have an active lifespan of roughly 17 weeks. The median lifespan was 18.1 weeks and the longest lifespan, *The Guardian*'s 'content-api-scala-client' repository, was 615.4 weeks (11.8 years). The second hypothesis was therefore supported. Notably, if only the repositories that received two or more commits and had a lifespan of two or more days are included—that is, the repositories that received at least *some* development—then the median lifespan increases threefold to 55.2 weeks, or just over one year.

The third hypothesis posited that most original repositories would not include a copyright license. Of the 5,342 original repositories, 2,992 of them (56.0%) did not list a license. The third hypothesis was therefore supported. As shown in Figure 4, when a license was specified, it tended to be a permissive license open-source license.

Use of Platform Over Time

The fourth research question asked about the prevalence of the use of GitHub by those organizations over time.

The oldest repository in the dataset (the *Los Angeles Times*' 'latimes-mappingla-geopy') was created on March 18, 2009. As shown in Figure 5, 2012 was a major year in news organizations' use of GitHub, with 16 organizations publishing their first original repository that year. Indeed, of the 65 organizations that published at least one original repository, 36.9% (*n* = 24) had done so by the end of 2012 and 81.5% (*n* = 53) had done so by the end of 2016. Thus, many prominent news organizations have been using GitHub for nearly a decade now.

The creation of new original repositories picked up notably in 2015 and peaked in 2017, when 883 new repositories were published. Although fewer new repositories were published in subsequent years, GitHub remains actively used. In 2021, for example, 333 new original repositories were published by 31 different outlets.

Regarding collaborative activity, issue reporting activity increased quickly at first, with an initial peak in 2014, when 1,754 issues were reported by external actors. However, such activity slowed through 2017 (1,106 issues) before rebounding with a second peak in 2019 (2,023 issues). In 2021, however, that number slid back down to 931. The number of pull requests by external actors continually rose through 2017 and peaked in 2019, when 5,346 such pull requests were registered. However, the 2,100 requests registered in 2021 was almost equal to the amount registered in 2014. The forking patterns reveal a similar trajectory. There was a continued increase in the number of times an original repository was forked by others through 2018 (5,098 forks). However, there was an unexpected dip in 2019, followed by a peak in 2020 (7,036 forks) and another dip in 2021 (4,777 forks).

Discussion

This study elicited five key findings. First, the use of GitHub by prominent news outlets that appeal to U.S. audiences is far from widespread. However, GitHub has been an active site of activity for many such outlets for well over a decade and remains actively used by several. Second, the open-source repositories on the platform were used to share a mixture of technologies and journalistic materials, with the former being a bit more prevalent. Third, the introductory documentation for the repositories routinely included at least partial amounts of both ambient and disclosure transparency, but rarely exhibited participatory transparency. Fourth, collaboration, measured across different markers, was almost non-existent in the vast majority of the repositories, though a small subset did involve significant amounts of collaboration. Finally, there has been an apparent decline in news organizations' use of GitHub—at least through their institutional accounts—and in the use of multiple collaboration affordances within their repositories in recent years.

A New Space for Performing Transparency

While prior work has focused on news websites and social media as the main stages for enacting transparency (Karlsson, 2021), this study illustrates that the infrastructure afforded by open-source repositories and platforms like GitHub have the requisite elements to offer another stage for performing transparency: they are durable in time, spreadable in space, and enable reciprocity. Moreover, this is an active stage within the context of journalism, as evidenced by the number of news organizations and repositories present on GitHub alone. While this study did not systematically examine the qualities of the collaborators, the evidence of at least *some* external collaboration paired with the knowledge that diverse actors occupy GitHub and open-source environments more broadly (Boyles, 2020a; Haim & Zamith, 2019) allows for a plausible argument that the infrastructure afforded by open-source repositories permits news organizations to perform transparency in new ways and reach new actors—or reach them differently—through those performances.

The sociotechnical features of open-source repositories and platforms like GitHub allow existing scripts to be refined by making it easier for journalists to at least share the source code behind their work and innovation (Usher, 2016; Weber & Kosterich, 2018). Prior work has identified the introduction of open-source ethos within newsrooms and argued that it could alter existing scripts by introducing new values and ways of working (Lewis & Usher, 2013, 2014). This study offers further evidence that open-source ethos is not just a rhetorical novelty in journalism. A significant number of prominent news organizations have been sufficiently influenced by it to create, in some cases, hundreds of repositories covering a range of materials and technologies. Additionally, while prior research has suggested that GitHub can be an avenue for sharing data journalism materials (Martin et al., 2022), this study adds empirical evidence of such use—especially by organizations like *FiveThirtyEight*, which highlight their analytic rigor as a differentiator. It also underscores how open-source repositories can facilitate the performance of journalistic roles more closely associated with (open-data) advocacy, a phenomenon observed in prior work (Camaj et al., 2022), by providing infrastructure to support the enactment of radical transparency. Indeed, while prior work has critiqued 'typical' data journalism for not living up to the rhetoric about its transparency (Zamith, 2019), platforms like GitHub offer an accessible technical solution for moving closer to those objectives.

However, the new script—or, at minimum, its delivery—still borrows heavily from earlier performances, which have been critiqued as being more akin to monologues (see Lewis, 2012). While the use of ambient and disclosure forms of transparency suggests a desire to make technologies, materials, and processes more scrutable, the non-use of participatory elements reflects a performance that still struggles to welcome outsiders and relinquish control. This is particularly notable given the centrality of collaboration within open-source ethos (Usher, 2016). While this is by no means indicative of a professional culture that *rejects* participation, it does reflect the perception that participation remains a side act when journalists have the spotlight (see Lewis, 2012).

The performance of transparency was far from uniform, though. Some organizations frequently offered detailed README files while others routinely provided curt descriptions and others rarely offered any documentation at all. In short, it appears that the script for performing transparency on open-source repositories remains in draft form. However, a qualitative impression indicated that this draft has evolved over the years. In particular, the abundance of "starter kits" developed by organizations in recent years to help their journalists and developers create open-source repositories in standardized ways (including with templated documentation) suggests that such scripts are maturing—at least within a subset of organizations.

Innovations that Aren't Innovative

While there has been some bullish rhetoric around news innovation, multiple scholars have critiqued news organizations for being reluctant to innovate or resorting to mimicry to remain competitive (see Belair-Gagnon & Steinke, 2020; Lowrey et al., 2019). This study's findings present similar concerns about the state of technological innovation originating from news organizations. While the sampled technologies had a broad range of applications, they tended to address minor productivity challenges and most technologies were unlikely to stand out amid the millions of other projects on GitHub.

This is to be expected in part: most open-source projects are not overly complex. However, the apparent rarity of major projects by prominent news organizations calls into question whether newsroom developers have the necessary resources to engage in bold endeavors (Belair-Gagnon & Steinke, 2020; Boyles, 2016). Alternatively, it is plausible that their practical enactment of open-source ethos is limited to certain kinds of projects, namely those that are simpler or have lower monetization potential. In other words, *real* or *disruptive* innovation may be seen as a resource that must be protected to provide a competitive advantage or revenue stream. This would present a challenge to the institutionalization of open-source ethos that merits further examination (see also Lewis, 2012). It also raises questions about whether news organizations are able to keep up with the innovations coming from journalistic outsiders that end up structuring journalistic spaces and practices (Belair-Gagnon & Steinke, 2020; Lowrey et al., 2019). While there are certainly examples of significant open-source innovations originating from news organizations, those appear to be the outliers.

It is important to note that this study did not examine the innovativeness of the news production materials (e.g., interactive stories) that news organizations shared through open-source repositories. Indeed, only a small fraction of the stories produced by the organizations were open-sourced on GitHub, and that fraction seemingly involved longer-term projects. It is therefore plausible—though there was qualitatively little evidence of it (see also Zamith, 2019)—that news organizations were innovating through their core product (news content) and sharing those innovations by making the underlying code (and processes) accessible to others.

The Isolation of Open-Source Collaboration

GitHub has the requisite elements to offer something akin to a "permanent 'trading zone'" (Weber & Kosterich, 2018, p. 324). The evidence provided here and elsewhere (e.g., Boyles, 2020a; Haim & Zamith, 2019) shows that the platform is used by a wide range of actors to engage in myriad activities around a broad spectrum of projects over sustained periods of time. This study adds that prominent institutional actors in journalism are among the active participants in that zone. Moreover, the evidence that the technological innovations targeted both technical and non-technical actors and were frequently designed or documented in a way that makes them useful to other actors points to a commitment to increase journalism's footprint within open-source environments. Building on prior literature, this may instill more of the field's values and ideals in other spaces (Lewis & Usher, 2013, 2014; Usher, 2016). In other words, journalistic actors are not just recipients of values and practices within the trading zone of GitHub. They are active senders as well.

However, the low amount of collaboration in the vast majority of repositories is concerning. The patterns observed here reflect and presumably add to the isolation newsroom developers already report (Boyles, 2016). These results are also consistent with prior evidence about the lack of collaboration around news-related open-source projects (Haim & Zamith, 2019). While those findings could be explained by the wide range of actors examined—many of which lacked prominence—the present findings show that even prominent organizations are frequently unable to mobilize collaboration (see also Boyles, 2020a).

This study's design precludes it from explaining *why* this is the case, but theory and empirical evidence offer some plausible reasons. First, collaboration around news innovation has been shown to be difficult due to differences in values and practices among would-be collaborators, both within and across newsrooms (Lewis & Usher, 2014, 2016). Second, the absence of clear structures and processes for making participation feel meaningful or worthwhile may inhibit collaboration (Belair-Gagnon & Steinke, 2020; Usher, 2016). Third, the frequent absence of explicitly identified copyright licenses may curtail at least some collaboration (Almeida et al., 2019; Haim & Zamith, 2019).

Notably, this study only examined public repositories. It is plausible that there is substantially more collaboration occurring within private repositories or on private branches of public repositories, with a project manager (single account) overseeing reviewing and pushing out public-facing changes. However, such a phenomenon would presumably mostly affect internal collaboration, as it is unlikely that external actors are being routinely given access to private institutional repositories. Additionally, post-hoc analyses indicated that there was very little collaboration occurring *between organizations* in the dataset (i.e., accounts associated with different professional news organizations working together). While prior work has observed an increase in inter-organizational collaboration—especially in the cases of data and investigative journalism (Larrondo-Ureta & Ferreras-Rodríguez, 2021; Martin et al., 2022)—it may be the case that such efforts only materialize on GitHub for larger-scale projects, occur largely out of public view due privacy concerns and embargo constraints, or manifest mostly through other channels altogether.

However, some projects did involve high levels of collaboration. Future work may thus seek to examine what kinds of external actors are choosing to collaborate with news organizations in open-source environments and why. It may also examine which project attributes tend to increase collaborative success. While scholars have examined such questions in general open-source contexts and hypothesized about such participation within journalism, empirical evidence is lacking.

Open-Source Repositories as Trust-Building Infrastructure

Convincing performances of transparency, useful innovations, and rewarding collaboration provide conditions for producing trust-building outcomes (Karlsson, 2021), especially when integrated within a single site of activity. To that end, open-source repositories—and GitHub in particular—do indeed offer the potential to be useful trust-building infrastructure within journalism. However, the present findings align with prior work in underscoring the challenges of institutionalizing sustained use of *journalism-adjacent* infrastructure (e.g., Bunce et al., 2018; Holton & Belair-Gagnon, 2018; Moran, 2021).

The clear upward trajectory in the use of GitHub through 2017 by some measures and 2019 by others shows that the rhetorical valuation of transparency, innovation, and collaboration (Lewis & Usher, 2013; Vos & Craft, 2017; Weber & Kosterich, 2018) has manifested itself in tangible form through the infrastructure afforded by open-source repositories. However, the subsequent decline in the use of GitHub, both in the number of projects published and in the amount of collaborative activity around those projects, raises the possibility that the enactment of open-source ethos on the world's most popular development platform has already peaked within the field of journalism—at least in the near term. To be clear, this is not to say that such ethos has been rejected or is destined for irrelevance. Rather, it is plausible that following an initial period of excitement, the prioritization of values and resources have been renegotiated. Further work is needed to empirically assess this possibility, and some elaboration is therefore warranted to provide direction for such work.

The high proportion of repositories that received little apparent engagement and collaboration offers a simple and intuitive metric to support the claim that the return on investment in the "labor of trust" (Zahay et al., 2021, p. 1055) through this kind of infrastructure is poor. In a resource-limited environment, this alone may be sufficient to justify reduced organizational involvement in open-source environments. Moreover, the frequent lack of tangible rewards for such labor may also prove discouraging for newsworkers who see community organizing ideals to be important to their role orientation (see also Moran, 2021; Zahay et al., 2021), which in turn encourages reprioritizing values in a constrained environment. This reprioritization may be accentuated by the growing body of recent evidence that at least partly decouples concepts like transparency from credibility and credibility from trust (Henke et al., 2020; Karlsson, 2020; Karlsson & Clerwall, 2018; Peifer & Meisinger, 2021), again raising questions about whether this form of trust-building labor is worth the costs. Scholarship examining the motivations and perceptions for engaging in this labor through the infrastructure afforded by open-source repositories would be a great addition to the literature (e.g., Boyles, 2020a).

However, the discussion so far here, and elsewhere, presumes two notions that may be flawed: (1) that engagement and collaboration are necessary for such infrastructure to be useful in building trust and (2) that a general audience is the target of such labor. First, the infrastructure provided by open-source repositories may be useful for trust-building simply for its *signaling* ability (see also Karlsson, 2021). For example, open-source software is often perceived to be more secure than its closed-source counterparts simply because of the impression that transparency allows for close scrutiny by others—despite the ample evidence that many projects never receive such scrutiny and that open-source software is also prone to serious bugs. A variant of such an effect might be applicable within the present context. Audiences may not need to evaluate open-source repositories or become active participants themselves for the infrastructure to produce trust-building (or at least trust-affirming) outcomes. This presumes, of course, that audiences are aware of news organizations' use of such infrastructure. That remains an open question that can be explored in future work.

Second, this infrastructure may be used less to appeal to general audiences—indeed, news organizations don't often promote their open-source repositories—and more to communicate trustworthiness and legitimacy to a narrower set of actors (e.g., their peers or critics) while performing a ritual to gain or maintain membership within desired groups (see also Müller & Wiik, 2023). Indeed, Boyles (2020a) argued that participation in open-source environments served newsroom developers by increasing their status among professional peers; served news organizations by expanding their prestige; and served the institution of journalism by bonding like-minded practitioners within an interpretive community. To that we may add that using the infrastructure of open-source repositories may be useful simply for its ability to convey strategic performances around transparency, innovation, and collaboration.

Viewed in this light, the sociotechnical possibilities arising from open-source repositories and collaborative development platforms can be theoretically and practically useful for advancing trust-related objectives in both material and nonmaterial ways.

References

Almeida, D. A., Murphy, G. C., Wilson, G., & Hoye, M. (2019). Investigating whether and how software developers understand open source software licensing. *Empirical Software Engineer*, *24*(1), 211–239. https://doi.org/10.1007/s10664-018-9614-9

Belair-Gagnon, V., Holton, A. E., & Westlund, O. (2019). Space for the liminal. *Media and Communication*, *7*(4), 1–7. https://doi.org/10.17645/mac.v7i4.2666

Belair-Gagnon, V., & Steinke, A. J. (2020). Capturing digital news innovation research in organizations, 1990–2018. *Journalism Studies*, *21*(12), 1724–1743. https://doi.org/10.1080/1461670X.2020.1789496

Boyles, J. L. (2016). The isolation of innovation. *Digital Journalism*, *4*(2), 229–246. https://doi.org/10.1080/21670811.2015.1022193

Boyles, J. L. (2020a). Deciphering code: How newsroom developers communicate journalistic labor. *Journalism Studies*, *21*(3), 336–351. https://doi.org/10.1080/1461670X.2019.1653218

Boyles, J. L. (2020b). Laboratories for news? Experimenting with journalism hackathons. *Journalism*, *21*(9), 1338–1354. https://doi.org/10.1177/1464884917737213

Bunce, M., Wright, K., & Scott, M. (2018). “Our newsroom in the cloud”: Slack, virtual newsrooms and journalistic practice. *New Media & Society*, *20*(9), 3381–3399.

https://doi.org/10.1177/1461444817748955

Camaj, L., Martin, J., & Lanosga, G. (2022). Professional ideals of data journalists around the globe: Congruencies and divergences between role conceptions and narrated role performances. *Journalism Studies*, *23*(12), 1450–1471. https://doi.org/10.1080/1461670X.2022.2094822

Curry, A. L., & Stroud, N. J. (2021). The effects of journalistic transparency on credibility assessments and engagement intentions. *Journalism*, *22*(4), 901–918. https://doi.org/10.1177/1464884919850387

Haim, M., & Zamith, R. (2019). Open-source trading zones and boundary objects: Examining GitHub as a space for collaborating on "news." *Media and Communication*, *7*(4), 80–91. https://doi.org/10.17645/mac.v7i4.2249

Heim, K., & Craft, S. (2020). Transparency in journalism: Meanings, merits, and risks. In L. Wilkins & C. G. Christians (Eds.), *The Routledge Handbook of Mass Media Ethics* (pp. 308–320). New York: Routledge. https://doi.org/10.4324/9781315545929-21

Henke, J., Leissner, L., & Möhring, W. (2020). How can journalists promote news credibility? Effects of evidences on trust and credibility. *Journalism Practice*, *14*(3), 299–318. https://doi.org/10.1080/17512786.2019.1605839

Holton, A. E., & Belair-Gagnon, V. (2018). Strangers to the game? Interlopers, intralopers, and shifting news production. *Media and Communication*, *6*(4), 70–78. https://doi.org/10.17645/mac.v6i4.1490

Karlsson, M. (2010). Rituals of transparency. *Journalism Studies*, *11*(4), 535–545. https://doi.org/10.1080/14616701003638400

Karlsson, M. (2020). Dispersing the opacity of transparency in journalism on the appeal of different forms of transparency to the general public. *Journalism Studies*, *21*(13), 1795–1814. https://doi.org/10.1080/1461670X.2020.1790028

Karlsson, M. (2021). *Transparency and journalism: A critical appraisal of a disruptive norm*. London: Routledge.

Karlsson, M., & Clerwall, C. (2018). Transparency to the rescue? *Journalism Studies*, *19*(13), 1923–1933. https://doi.org/10.1080/1461670x.2018.1492882

Larrondo-Ureta, A., & Ferreras-Rodríguez, E.-M. (2021). The potential of investigative data journalism to reshape professional culture and values. *Communication & Society*, *34*(1), 41–56. https://doi.org/10.15581/003.34.1.41-56

Lewis, S. C. (2012). The tension between professional control and open participation. *Information, Communication and Society*, *15*(6), 836–866. https://doi.org/10.1080/1369118X.2012.674150

Lewis, S. C., & Usher, N. (2013). Open source and journalism: Toward new frameworks for imagining news innovation. *Media Culture & Society*, *35*(5), 602–619. https://doi.org/10.1177/0163443713485494

Lewis, S. C., & Usher, N. (2014). Code, collaboration, and the future of journalism. *Digital Journalism*, *2*(3), 383–393. https://doi.org/10.1080/21670811.2014.895504

Lewis, S. C., & Usher, N. (2016). Trading zones, boundary objects, and the pursuit of news innovation: A case study of journalists and programmers. *Convergence*, *22*(5), 543–560. https://doi.org/10.1177/1354856515623865

Lowrey, W., Sherrill, L., & Broussard, R. (2019). Field and ecology approaches to journalism innovation: The role of ancillary organizations. *Journalism Studies*, *20*(15), 2131–2149. https://doi.org/10.1080/1461670X.2019.1568904

Martin, J. A., Camaj, L., & Lanosga, G. (2022). Scrape, request, collect, repeat: how data journalists around the world transcend obstacles to public data. *Journalism Practice*, Advance Online Publication, 1–18. https://doi.org/10.1080/17512786.2022.2142837

Masullo, G. M., Curry, A. L., Whipple, K. N., & Murray, C. (2022). The story behind the story: Examining transparency about the journalistic process and news outlet credibility. *Journalism Practice*, 16(7), 1287–1305. https://doi.org/10.1080/17512786.2020.1870529

Moran, R. E. (2021). Subscribing to transparency: Trust-building within virtual newsrooms on Slack. *Journalism Practice*, *15*(10), 1580–1596. https://doi.org/10.1080/17512786.2020.1778507

Moran, R. E. (2021). Subscribing to transparency: trust-building within virtual newsrooms on slack. *Journalism Practice*, *15*(10), 1580–1596. https://doi.org/10.1177/14648849211048961

Müller, N. C., & Wiik, J. (2023). From gatekeeper to gate-opener: open-source spaces in investigative journalism. *Journalism Practice*, *17*(2), 189–208. https://doi.org/10.1080/17512786.2021.1919543

Peifer, J. T., & Meisinger, J. (2021). The value of explaining the process: How journalistic transparency and perceptions of news media importance can (sometimes) foster message credibility and engagement intentions. *Journalism & Mass Communication Quarterly*, *98*(3), 828–853. https://doi.org/10.1177/10776990211012953

Pew Research Center. (2021, July 27). Methodology: State of the news media. Retrieved August 15, 2021, from https://www.pewresearch.org/journalism/2021/07/27/state-of-the-news-media-methodology/

Plantin, J.-C., Lagoze, C., Edwards, P. N., & Sandvig, C. (2018). Infrastructure studies meet platform studies in the age of Google and Facebook. *New Media & Society*, *20*(1), 293–310. https://doi.org/10.1177/1461444816661553

Riffe, D., Lacy, S., Watson, B. R., & Fico, F. (2019). *Analyzing media messages: Using quantitative content analysis in research* (4th ed.). New York: Routledge.

Usher, N. (2016). *Interactive journalism: Hackers, data, and code*. Champaign, IL: University of Illinois Press.

Usher, N. (2018). Re-thinking trust in the news. *Journalism Studies*, *19*(4), 564–578. https://doi.org/10.1080/1461670X.2017.1375391

Vos, T. P., & Craft, S. (2017). The discursive construction of journalistic transparency. *Journalism Studies*, *18*(12), 1505–1522. https://doi.org/10.1080/1461670X.2015.1135754

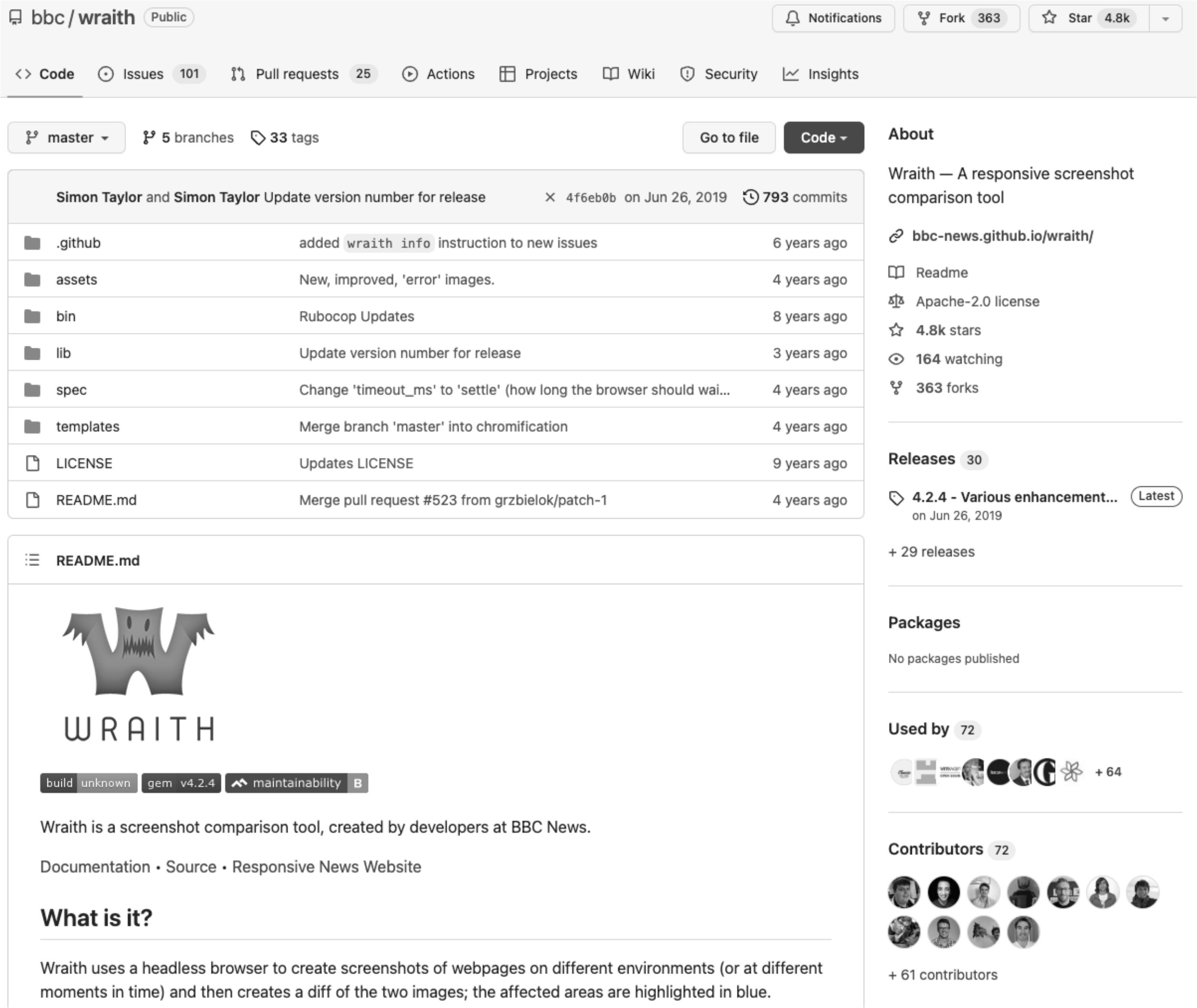
Weber, M. S., & Kosterich, A. (2018). Coding the news. *Digital Journalism*, *6*(3), 310–329. https://doi.org/10.1080/21670811.2017.1366865

Zahay, M. L., Jensen, K., Xia, Y., & Robinson, S. (2021). The labor of building trust: Traditional and engagement discourses for practicing journalism in a digital age. *Journalism & Mass Communication Quarterly*, *98*(4), 1041–1058. https://doi.org/10.1177/1077699020954854

Zamith, R. (2019). Transparency, interactivity, diversity, and information provenance in everyday data journalism. *Digital Journalism*, *7*(4), 470–489. https://doi.org/10.1080/21670811.2018.1554409

Zamith, R., & Braun, J. A. (2019). Technology and journalism. In T. P. Vos, F. Hanusch, D. Dimitrakopoulou, M. Geertsema-Sligh, & A. Sehl (Eds.), *The International Encyclopedia of Journalism Studies* (pp. 1–7). Hoboken, NJ: Wiley. https://doi.org/10.1002/9781118841570.iejs0040

Figures and Tables



Simon Taylor and Simon Taylor Update version number for release × 4f6eb0b on Jun 26, 2019 793 commits

.github	added `wraith info` instruction to new issues	6 years ago
assets	New, improved, 'error' images.	4 years ago
bin	Rubocop Updates	8 years ago
lib	Update version number for release	3 years ago
spec	Change 'timeout_ms' to 'settle' (how long the browser should wai...	4 years ago
templates	Merge branch 'master' into chromification	4 years ago
LICENSE	Updates LICENSE	9 years ago
README.md	Merge pull request #523 from grzbielok/patch-1	4 years ago

README.md

Wraith is a screenshot comparison tool, created by developers at BBC News.

Documentation • Source • Responsive News Website

What is it?

Wraith uses a headless browser to create screenshots of webpages on different environments (or at different moments in time) and then creates a diff of the two images; the affected areas are highlighted in blue.

About

Wraith — A responsive screenshot comparison tool

bbc-news.github.io/wraith/

Readme

Apache-2.0 license

4.8k stars

164 watching

363 forks

Releases 30

4.2.4 - Various enhancement... Latest
on Jun 26, 2019

+ 29 releases

Packages

No packages published

Used by 72

+ 64

Contributors 72

+ 61 contributors

Figure 1. A screenshot of the GitHub page for the BBC’s ‘wraith’ repository.

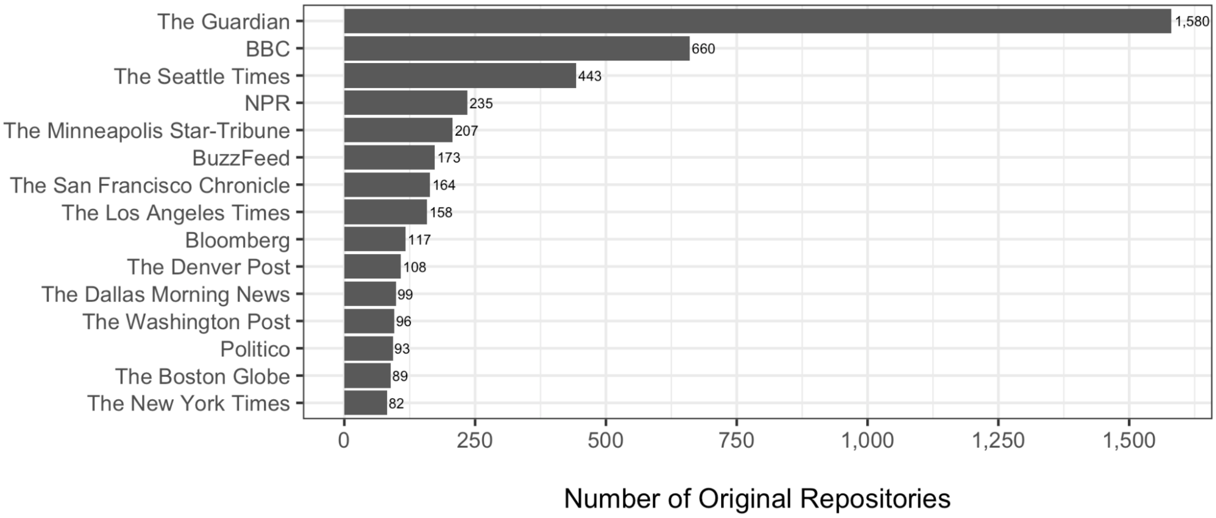


Figure 2. The number of original (non-forked) repositories for the 15 organizations with the most repositories.

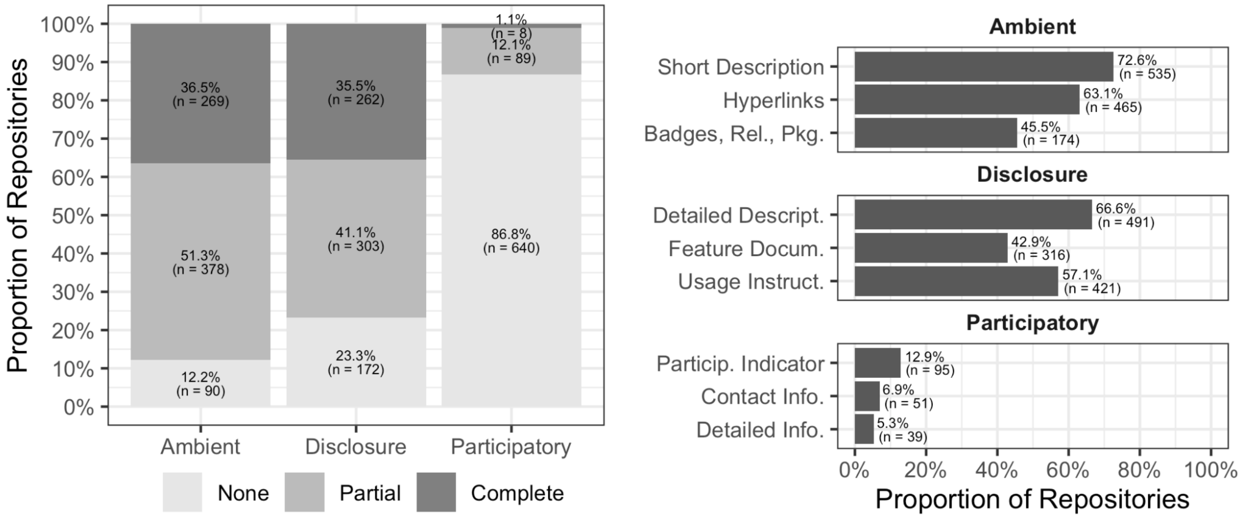


Figure 3. The presence of three forms (left) and nine elements (right) of transparency in the 737 repositories that were manually coded. The element of Badges, Releases, and Packages was only coded for technology-related repositories (*n* = 382).

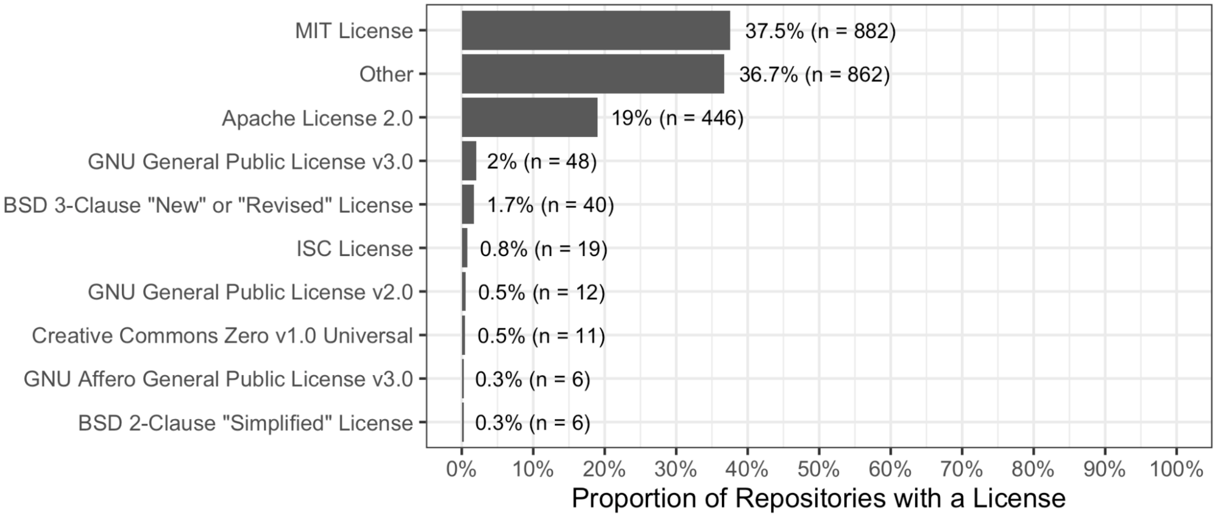


Figure 4. The ten licenses most often used in repositories identified by GitHub as having a license (*n* = 2,350). An informal review of repositories categorized as Other showed that they frequently referred to one of the top three licenses in a manner that simply did not meet GitHub's formatting expectations.

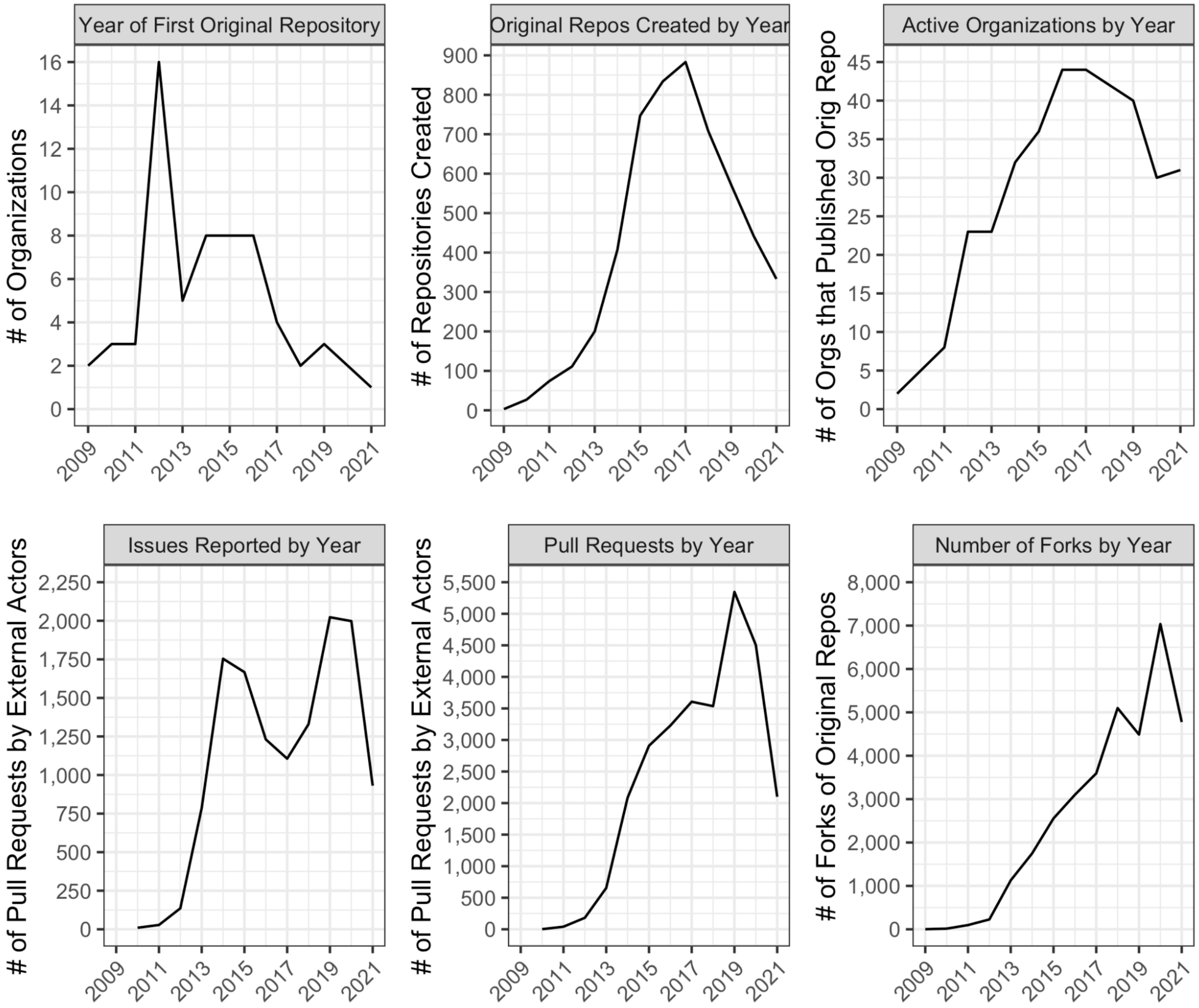


Figure 5. The extent of the use of GitHub by prominent news organizations between 2009 and 2021, as measured across three dimensions.