# COVID-19 Community Archives and the Platformization of Digital Cultural Memory

Amelia Acker
University of Texas at Austin
School of Information
aacker@ischool.utexas.edu

Lucy Flamm
University of Texas at Austin
School of Information
lucy.flamm@gmail.com

## Abstract

*In this study we aim to understand how GitHub is used by COVID-19 interest groups for organizing community archives to protect their knowledge from the Chinese government's censorship efforts. We introduce two case studies of such COVID-19 community archives published with GitHub that appeared online in early 2020. Using public GitHub repository documentation and web archive web crawls from the Internet Archive's Wayback Machine, we describe how these digital community archives emerge and exist on the platform, how knowledge of them circulated on other US based social media sites and show strategies and tactics these volunteers used to keep these community archives alive, resist censorship, and guard the safety of these collections. We argue that these COVID-19 community archives are at risk because of their platform accessibility as much as the content they document, and that understanding how organizers use GitHub's platform affordances is essential to theorizing how platforms are impacting approaches to preserving cultural memory.*

## 1. Introduction

Since the beginning of the COVID-19 outbreak in December of 2019, people all over the world have been using online platforms, mobile phone apps, and social media to share documents, health information, and personal experiences about the pandemic's trajectory. Shortly after the impact of coronavirus became known, researchers began to observe the spread of rumors, health misinformation, conflicting reports, hoaxes and inaccurate data across the web [1]. Public health researchers and epidemiologists began to call for increased access to public health data, vetted sources, and more peer reviewed research. Legislators and policy makers began to turn towards the scourge that health misinformation on platforms can cause in fomenting a public health crisis and demanded that platforms make more efforts to moderate. Information researchers,

historians, journalists and archivists began to call for digital archiving efforts to document this moment and ask, in an age of platforms and the aftermath of the crisis how will we remember this infodemic [2], [3]? The research reported in this paper responds to these calls, by examining how communities of volunteers have used GitHub to create community archives to, in their own words, remember COVID-19.

In February 2020, activists and journalists began sharing links on social media platforms like Twitter and WeChat to community archive projects coming from China [4], [5]. A handful of community archive projects organized by Chinese citizens had quickly emerged, attempting to preserve news stories, personal narratives, and first-hand accounts of the novel coronavirus outbreak, anticipating that they would likely be removed from the Internet by Chinese government censors. Some of these documentation projects use GitHub, a US-based platform that allows teams to collaborate, share software projects and documentation, collaborate with code, and create repositories of information for people to access multiple versions of repositories as projects are developed [6].

Groups of Chinese volunteers started repositories (known as "repos") to create collections of documents, solicit contributions, and coordinate the translation of first-hand accounts of the pandemic. Initially, GitHub appeared to be used by volunteers in order to evade state censorship efforts at blocking citizens' accounts of the pandemic from spreading online but many of these repos have since been taken down, gone private, have been "404'd" and are no longer accessible on the live web. While many of the surviving community archive projects risk being scrubbed from the Internet by Chinese government censors, it is instructive for researchers of critical and ethical perspectives of digital social media to consider how organizers use, subvert, and protect the knowledge that they seek to preserve using platforms. GitHub provides observation entry points for researchers because the platform strives to provide cloud storage, open access and public

HICSS

repositories as primary platform features (as well as powerful network effects). And like many other platforms striving to enter into global markets with information infrastructure, GitHub has a history of responding to state censorship efforts and takedown requests from governments.

Platforms present themselves as huge barriers to digital archiving efforts and remembering the pandemic because of the way that data flows are commodified, circulate, and remain accessible to users on the Internet as public information infrastructure [7]. For example, open source investigators find that bearing witness and documenting human rights violations are increasingly challenged in the face of content moderation policies that delete and ban evidence on YouTube [8]. Many racial justice activists and scholars have observed that social media platforms have become sites where the ritualized public mourning and remembering of Black lives who have been brutalized and murdered assumes content where "death and trauma are continuously re-inscribed" [9]. Such content that becomes viral and may risk desensitizing viewers to the police violence and brutality that users who post these images seek to critique and draw awareness to.

These access challenges, among others, such as datafication and commodification shape and are shaped by contemporary memory practices that spill across platforms. We call this the *platformization of digital cultural memory*. Here we draw on scholars who theorize the impact of platforms at the juncture of digital, economic, and governmental infrastructures [10], as a process involving human and non-human actors [11], and as a means for critiquing the programmability of data from platforms [12]. We employ the approach of Langlois et. al  and use "disaggregation" to evaluate the functionality and implications of platform features as they pertain to Github's publics and archiving efforts [13]. GitHub features simultaneously foster and inhibit practices of documenting journalistic and personal narratives. Cultural memory becomes institutionalized on the platform, and platform policies consequently mediate long-term accessibility and preservation.

In this paper we discuss two case studies of COVID-19 community archives published using GitHub that appear to have encountered Chinese censorship efforts in the first few months of 2020. We argue that these community archives are at risk because of their accessibility as much as their content, but that understanding how organizers use GitHub's platform affordances is essential to theorizing how platforms are shifting the concerns of digital preservation from storage to ensuring long-term access and authenticity. Using public GitHub repo documentation and web archive web crawls from the Internet Archive's Wayback Machine (IAWM), we describe how these digital community archives emerged on GitHub, how knowledge of them circulated on other social media platforms and detail some strategies these volunteers have used to keep these community archives alive, resist censorship, and guard their safety through translation, making copies, and other reproduction strategies.

Following on previous ethical and critical digital social media research published in HICSS, we seek to show these memory practices beginning on GitHub that spill over to other platforms are confrontations between communities and the tools they use to communicate information [14]. These confrontations are productive for digital social media scholars because they surface conflicting values of platforms, government interests, and communities seeking self-determination over repositories of information, accessing information publicly and building community archives that commemorate victims of the pandemic. What's at stake when grassroots organizers and community archivists use platforms as memory infrastructures? What can we learn from how community archives have enrolled platform features into their memory practices, self-determination efforts and community-building? After presenting some background literature, we briefly describe our methods and limits of the study, then we provide a statement on our research team's positionality. In the fourth and fifth sections we introduce two case studies of Chinese community archives documenting COVID-19 using GitHub. The paper finishes by discussing the platformization of digital cultural memory and the need for research that examines how community archives are created, accessed, and impacted by platform features.

## 2. Background literature

### 2.1 Community Archives

The GitHub repos that we researched are both identified as archives and community efforts at documenting and preserving memories of the coronavirus by Chinese users. In documentation and in news coverage of these efforts, organizers of the repos identify themselves as Chinese citizens who are concerned "volunteers", "organizers", and "netizens" committed to collecting personal accounts of their experiences, which in addition to the pain and trauma of pandemic also include censorship and threats from the government. As such, we take these creators' identification and efforts on their own terms and consider these repos to be digital community archives that are organized and enacted on GitHub where the platform operates as an intermediary between users and third parties such as advertisers or government

regulators. We argue that these COVID-19 community archives materialized in response to misrepresentations and absences of personal narratives, but also efforts to address health misinformation, suppression of on-the-ground first-person accounts, and government censorship of online personal experiences of the pandemic in China.

As people have adopted new digital technologies—from digital cameras to email listservs and Facebook groups—community archives have increasingly leveraged digital tools on the Internet. Some initiatives are entirely organized and accessed via platforms like Instagram. Community archives are defined as non-traditional collections and archival communities made up of people who are not well-documented, or not documented at all, by traditional cultural heritage institutions like archives, libraries, or museums [10]. The critical archival theorist Michelle Caswell has argued for the more precise description of "marginalised identity-based community archives" in order to distinguish from other kinds of community archives which may come from dominant or oppressive power structures that ultimately maintain their dominance through preservation [11, p. 23]. Thus, marginalized identity-based community archives are formed by people not only to create documentation around their experiences but also to counteract misrepresentations and dominant documentation practices.[1] Crucially, for the COVID-19 community archives and other activists' documentation efforts, GitHub is a free, open, and publicly available platform in China. While other US-based platforms such as Google and Facebook are presently banned, many Western platforms with enterprise and e-commerce purposes have sought to grow their markets in China while maneuvering the Great Firewall [15].

Researchers have observed that as community archive organizations have transitioned to born-digital records and online communication for access they increasingly use platforms to capture documentation from social media activity, circulate information resources, and even collaborate on writing projects [16]. However, many archival scholars and media studies researchers have examined the challenges that groups face when establishing digital community archives that reflect their values, support their needs, and protect their knowledge as minoritized groups [17]–[19]. Platforms frequently update features, policies, and affordances that can impact users in a variety of ways that put their identities, community practices, and commemoration with platformed content at risk [20], [21]. As such, digital community archives using platforms can at once enact self-determination efforts while risking oppressive scrutiny, or even ambivalence and obfuscation in relation to the platform's governance.

## 2.2 GitHub

Since 2005, GitHub has operated as cloud-based hosting services for the Git version control system ensuring documentation of non-linear collaborative development of software [22]. "Git" is a version control system that offers a method of open collaboration for distribution enabling documentation of workflow through producing working copies of a repository for contributors and smooth merging processes. The open source software project has grown immensely popular, with a recent survey by Stack Overflow revealing that nearly 90% of respondents utilize Git in their coding process [23]. GitHub self-articulates as a space for developers to come together and work on code as reflected in the more than 100 million repositories of code hosted on the site [22], [24]. In 2018 Microsoft acquired GitHub and since then has expanded its scope, describing itself as an online platform for connection, storage, and retrieval for teams and enterprise users. GitHub as a platform has increasingly capitalized its services to private companies around the world as enterprise software for documentation management for 3 million organizations. Approximately 60% of Fortune 500 companies use GitHub for their internal development and documentation processes [22].

While the platform makes money from tailored enterprise plans with varying capabilities for administrative actions, security, and support [25], there are many features available to users for free, allowing them to publish and make actions and documentation public on the Internet. Many thousands of individual users use the free version of GitHub for community-oriented endeavors [26], and groups in China have used GitHub repos to organize actions and volunteer efforts for many years. For example, 996.ICU is a group of information technology workers fighting to disrupt the 9am-9pm, 6 days per week schedule, and they use GitHub as a bulletin board for visibility and labor advocacy [27]. Because GitHub is not blocked by Chinese Internet service providers (ISPs), it is also used to document and publish content that would otherwise be censored and inaccessible. For more than four years, prominent anti-censorship repos have been utilizing GitHub to combat Chinese government censorship and advocate for free speech [28].

It is important to interrogate the value of such a platform for preservation of information as much as it

---

[1] Caswell has theorized this elsewhere as the "symbolic annihilation" of minoritized groups. For more see: Caswell, M. (2014). Seeing yourself in history: community archives and the fight against symbolic annihilation. *The Public Historian*, 36(4), 26-37.

affords the ability to access materials that are designed to be openly available online. For our purposes here, its software documentation and versioning control features are of great interest to the study of community archive practices because they capture chains of evidence in metadata information about repository updates. Both analog and digital archiving communities believe in the importance of tracing and confirming provenance. In this regard, GitHub provides a constant chain of documentation on additions, changes, merges, and contributors. This metadata information can be understood as valuable contextual information and valuable provenance. The chain of provenance is continuously preserved within the structure of the platform and continuously produces metadata on changes to each line of code or other form of file. However, this reliable and openly accessible provenance information can also put users at risk of censorship crackdowns as part of China's Internet censorship and surveillance policies.

### 2.3 Platforms and infrastructuralization in China

Scholars who research China's Internet governance typically focus on techno-nationalist influences in the development of digital platforms, ranging from data flow policy and its localization regimes [29], to the programmability and "infrastructuralization" of platforms such as WeChat [30]. As a result of China's Great Firewall and longstanding censorship policies since the establishment of ISPs [15], [31], China's citizenry has practiced self-censorship in online and digital communications [32]. In their discussion of the Chinese model of platform infrastructuralization through Tencent's WeChat, Plantin and de Seta emphasize the protectionist environment where platforms like Tencent and Weibo allow governmental control and user surveillance in an effort to become, essentially, public information infrastructures in China. US-based and European platform companies aim to become public information infrastructures in many contexts as well, but eschewing state oversight for more private, corporate regulatory frameworks. As Western platforms such as GitHub attempt assert their "infrastructural ambitions" in Chinese markets, researchers can witness the infrastructuralization of digital communication services such as software development and knowledge repositories unfold, while observing how foreign privatized firms like Microsoft must negotiate China's techno-nationalist platform policies [30, p. 259]. In building on previous research examining Internet censorship in China and digital communication technologies [15], [29], [32], this study highlights how governments, platforms, and users can be at odds with the values that appear to be 'baked in' to

platform features, such as public versioning, open access to documents, and cloning repositories. Platform confrontations like the ones we present here instantiate values while engendering power, revealing how a functional sovereignty is enforced by platforms as intermediaries of states (or alongside them) [7]. These confrontations can also illustrate how alternative uses of platform mechanisms are possible, revealing the stakes of platforms in and across nation states, and society.

## 3. Methodology

In this study we aim to understand how GitHub repositories are being used by COVID-19 interest groups for organizing community archives and memory infrastructures, while protecting their knowledge from government censorship. How do groups build, organize and provide access to their GitHub community archives? How do they use platform features to protect, preserve, and represent community documentation? In addressing these research questions, we show how public GitHub repos can be cloned, forked, and taken private in order to preserve personal narratives and counteract government censorship and health misinformation. Due to the sensitivity of these materials, we decided to only use data and repositories which have been already published and cited by Chinese and international news outlets. We primarily use contextual metadata from documentation describing each archive's GitHub repo, and not the records that they have been published as some contain identifying information, have been since made "private" or have been unpublished or are likely to be taken down. While all these sources were publicly available online at the beginning of our analysis (in Spring 2020), some have disappeared and many more are likely to be deleted. One strategy that we find and report on when users copy, fork or clone these sources. Even as we investigate and report on these archival practices, we seek to protect these copies and clones from censorship efforts too. Unless Chinese and international news outlets have previously published these repos and usernames, we do not identify users in these research findings, in order to protect the privacy of these GitHub users and the confidentiality of these important repos.

Documentation about public GitHub repos and user accounts are publicly available online, including metadata on account updates and actions. Each user account page contains information about the repositories that user has created. GitHub automatically generates a data visualization concerning activity in the form of issues and pull requests called. Contribution activity is also documented in the form of a timeline showing which commits are linked to which repository (Figure 1.).
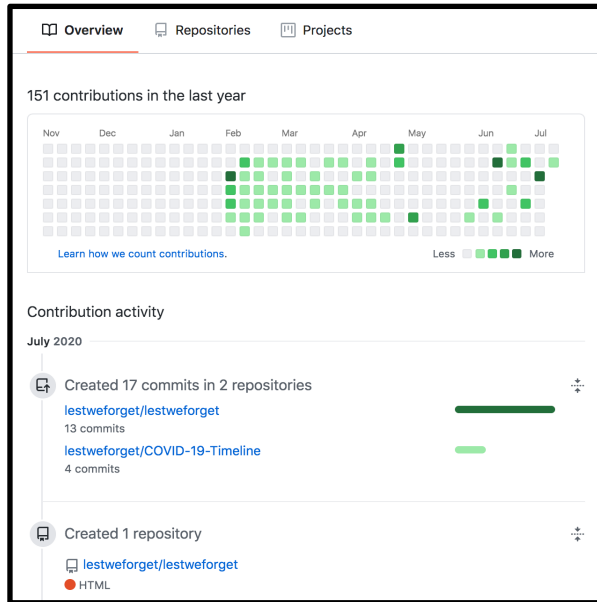
**Figure 1. Contribution activity timeline and visualization of a GitHub user account.**

The timeline feature is mirrored within each GitHub repo document activity. To trace the arc of these community archive initiatives, we analyzed existing repo timelines and web archives of the platform. This enabled consideration of the changing scope and added content in the context of the COVID-19 pandemic and domestic politics. For the repos which have since been made private, we utilized the Internet Archive's Wayback Machine web archives. Public repos on GitHub are frequently crawled for web archives like the Internet Archive, so we were able to utilize provenance information from the web archives of the GitHub pages previously operational through comparing snapshots of each repo's original URL. For more on this method of using provenance information from IAWM web craws, see [33].

International media coverage of the public health crisis, COVID-19 documentation initiatives, and Chinese web content provided further source context for the climate in which these community archives projects operated. Further, Chinese and English social media posts provided reactionary commentary and personal insight into the status of these community archive projects as they emerged and were subject to censorship. In observing these repo actions from web archived URLs and comparing them to news stories and other social media posts published at the same time, we were able to construct a timeline of events, actions, and in some cases verified motivations from organizers describing these changes in repo documentation and

news reporting. Reading this digital evidence in concert enabled us to visualize and trace how two teams of Chinese community archivists have worked to document narratives concerning COVID-19 with GitHub creating community archives.

### 3.1 Limitations

It is imperative to state the limits of this research when observing archival practices in 'real-time' as organizers respond to government censorship and negotiate layers of accessibility on platforms. To avoid further highlighting and naming users operating in this vulnerable and quickly-changing context, we have chosen not to publish usernames or cite content we know remains at risk. As a result of masking sources (and users' identities) that are still available online, we limit the primary sources and data drawn upon for evidence and presented to readers for verification to instances previously reported in media coverage.

Translation of these sources are another limitation we acknowledge in this study. Both repositories featured some English prose in their self-written descriptions of identities, motivations, and archive content but most documents were in Mandarin Chinese. With the understanding that word selection in bilingual texts carries weight and power, a native speaker contributed Chinese-to-English translations of the data.[2] This research focuses on the occurrence of these community digital archives, and our translator collaborator was able to assist in translating materials that were key to understanding the trajectory of these repos and confirming semantic meanings.

Critics of our choice to interpret archival efforts and memory practices so shortly after they unfolded online during a time of great upheaval during the pandemic would be right to caution their generalizability, especially since the end of the pandemic is yet to arrive. However, we believe that these findings from the 'beginnings' of community documentation responses to the pandemic will be meaningful in future accounts and theories of information practices documenting the COVID-19 aftermath, as well as understanding the impact of platform affordances on minoritized communities, efforts at self-determination and their archives.

### 3.2 Statement on our positionality

We acknowledge our team positionality and our subjectivities as researchers as we approach the topic and interpret these findings. Our positionality informs

---

[2] Due to the nature of the documents and the context of the research, our research team translator wishes to remain anonymous.

not only our individual perspectives, but perspectives on methodology and construction of the research process. As white, cisgender researchers, both co-authors have inherently benefited from the structural racism of the academy and the higher education landscape in the US. Having both been trained and working at American institutions, we are cognizant of the centrism of Western scholarship and the Euro-centric methodologies which inform our understanding of information studies, archival practices, and cultural stewardship of digital memory and platforms. As a result of our locations and institutional affiliation, we have had open Internet access and continue to be able to access the web with limited concerns for web tracing, tracking, and censorship. We recognize that each of these privileges impact our research positionality when investigating digital archives, community preservation, and information networks in China and beyond.

## 4. Ensuring Access with Translations

One COVID-19 community archive projects, "/nCovMemory," in particular illustrates coordination and tactics that community archivists use to confront Chinese Internet censorship of speech in China while using platform features for collaboration and managing documentation. According to /nCovMemory repo's description, the archive began collecting accounts in January 2020. Shortly after the first /nCovMemory repo was created the account's owner, Memoryhonest created another repo entitled, /nCovMemory-en featuring English translations of the same articles published in the Chinese /nCovMemory repo [34].

The repo has a document that introduces the purpose and motivations of the team of volunteers. These community archivists call themselves the Chengji Translation team, a small group of volunteers committed to remembrance through authentic documentation. From January to February 2020, the Chengji Translation Team posted English translations of first-hand reports from the /nCovMemory repo and published them to the /nCovMemory-en repo.

US journalists interviewed the creators of the /nCovMemory GitHub repository and reported that a team of 7 volunteers had been archiving media reports, as well as non-fiction works, and personal narratives that conflicted with Chinese state media reports of the pandemic on the ground [35]. After the US media coverage of the archive project, the first-hand accounts and articles in Chinese as well as the English

translations stopped being published to both GitHub repos.

In April, the Chinese repo /nCovMemory began to display a 404 page according to archived webcrawls of the website from the Internet Archive's Wayback Machine [36]. Concurrently, it was reported that organizers of similar COVID-19 community archives using GitHub were becoming targets of a Chinese government crackdown censoring accounts of the pandemic [35]. Some Chinese GitHub repo creators were arrested and put under house arrest, while others are reported to still be missing [37]. While the /nCovMemory-en English translation GitHub repo is still publicly available online, the Chinese language repo /nCovMemory has been taken down or gone private [38].[3] However, before the Memoryhonest account went quiet and took down public access to the /nCovMemory, both the Chinese and English repos were downloaded and republished as cloned repos by other GitHub users. Such clones of the original /nCovMemory community archive allow those interested to continue to "browse [their] archives," of the English translations but also access the materials through different copied repos, including clones that have now been published outside of GitHub as well.

For the Chengji translation team, the act of translation ensures a broader audience, providing a means for more readers to witness first-hand accounts that are not represented currently in Chinese media coverage. In describing their motivations in the repo ReadMe and "about" section, the volunteers regard these acts of translation as both resistance and self-determination in giving authentic accounts. They acknowledge the likelihood of the repo receiving more traction, longevity, audience reach and endurance of the materials if other versions of the Chinese collections are translated into English and posted in the English language repo.

## 5. Lots of Clones Keeps Stuff Safe

The self-articulated memory practices and preservation efforts guiding /nCovMemory can be observed across GitHub in other repos emerging in the early months of 2020. On February 4th, just over a week after Wuhan was placed under government-mandated quarantine [39], another repo, /wuhan2019, emerged on GitHub [40]. The archive's mission was clearly articulated in the repository's description, proclaiming: "Lest we forget this pandemic, or at least I won't forget" acting as an archive for mainstream media articles that

---

[3] Repos that have been taken down, unpublished, or have gone from public to private each display the same '404 not found' page dialog, so it is difficult to infer motivations without verification from the

account holder. In this case, published interviews with the Memoryhonest account organizers report that they went 'private' in response to pressure from government censorship efforts.

organizers anticipate will be censored or removed [41]. As Chinese citizens began to voice concern about accuracy and access to information, both /nCovMemory and /wuhan2019 appear as community efforts to document mainstream narratives, facts, and figures relating to coronavirus and referencing each other's efforts, while bolstering documentation with personal narratives and first-hand accounts. Both scholarly media analysis and journalism highlighted the beginning of 2020 as a period of increasing worry, agony, and despair over not only the virus but the government's response to limit and censor oppositional content originating from China [28].

In the context of limited access to reliable public health information concerning coronavirus, /wuhan2019 functions to document disappearing coronavirus information as well as misinformation preserving both reliable and unreliable accounts. Amongst the state sponsored media emerging that COVID-10 community archives sought to document using GitHub repos, the /wuhan2019 repository dedicates itself to preserving mainstream media articles. The repo contains a ReadMe file written by the creator, outlining the mission and introducing the trajectory of the initiative. It has been updated as the archive grows, illustrating the narrative changes and enduring motivations of those creating and maintaining the repos. The language describes a reality in which the actors expect all the documents hosted on the repo to "be erased by CCP [sic: Chinese Community Party] someday" [42]. In recirculating previously published online documents as PDF files the /wuhan2019 acts as a curator anticipating both a present and a future in which these reports will be destroyed.

The documents are uploaded, added, and preserved within the /wuhan2019 repository and are grouped by media outlet into a folder structure of chronologically organized publications. The earliest article dates to late December 2019, and the latest commit (or version) to the repository on April 24th, 2020 [41]. 43 news outlets are represented with varying amounts of documentation per outlet. The outlets range from *Xiao Xiang Morning News* published in Changsha capital of China's Hunan province to *GQ China*, a popular men's magazine. Within each folder, articles are preserved as individual digital objects with naming conventions including date and article title. For example, the /wuhan2019/New York Times (中文) repository hosts 319 pdfs of Chinese *New York Times* articles covering the outbreak in China from US perspectives that were published in the Chinese version of the newspaper.

Among the documents stored in /wuhan2019 are internationally-recognized features on the evolution, ramifications, and human-toll of the virus in China and abroad [35]. On February 3rd, the magazine branch of

Caixin Media published a 40,000 word cover story chronicling the impact of how stalled dissemination of information pertaining to the virus led to the scale and scope of crisis within Wuhan [35]. With such a critical stance of local government, publishers who ran articles such as this assumed swift censorship, accepting the risk of deletion and forced edit by government censors. Lestweforget was then busy, locating published articles that would soon be destroyed or censored.

Another genre preserved within the archive are diary-like first person accounts reflecting intimate reflections. Such was the case with the diaries of the writer Fang Fang—a well-known literary figure whose works have historically concerned the lower strata of society—published on social media. As noted by the Chinese public and international outlets, each of her accounts were consistently and immediately deleted by censors within approximately an hour [43]. /wuhan2019 hosts 57 of Fang's diary entries. Such personal testimonies like Fang Fang's are pivotal in shaping understanding of the concerns and constraints which shaped Hubei in the context of COVID-19. Reading commentary allows one to understand the intrinsic details of events both personal and universal. These types of historical accounts and individual narratives of crises are pertinent when access to accurate information is inhibited or controlled by oppressive governments. As posed by the Chinese author and academic Yan Lianke, without documentation like Fang's which digitally pens the experience, emotions, and socio-cultural landscape, "What would we have heard? What would we have seen?" [44].

The maintenance of /wuhan2019 with new, reputable articles before censorship edits, alterations, and deletion from the government continued throughout February, March, and the beginning of April. As the toll of the virus exponentially grew, so did Chinese censorship mechanisms. In response, social media users began to use fictional languages like Klingon and Elvish, or emojis in their posts to try and evade censors using automated language identification trackers [4].

With the disappearance of three individuals affiliated with another Chinese GitHub project, Terminus2049, and the lack of additions to 2019nCovMemory before disappearance, the landscape of community archiving using Git was rapidly changing. On April 26th, /wuhan2019 suspended contributions while maintaining status as a public repository. In a recent update to the ReadMe, the decision to suspend updates to the repository came in part because the pandemic was showing signs of decline in China, and in part due to the fear of having identity disclosed and personal security compromised [38]. As concisely articulated by a GitHub user who messaged the repository and subsequently made public in a recent

statement, "We all want to remember history, but there are always people who do not want us to remember" [38]. On June 5th, however, an update was published reasserting the aim of the project and calling for the release of the Terminus2049 contributors, and freedom of speech in China [45].

## 6. Discussion

Platforms like GitHub promise an easier way to collaborate, host information online (like code and documents) and share access to team projects. We see many of the GitHub features that support collaboration used in innovative ways by community archives, from taking a repo private to protect volunteers, to cloning or forking a parent repository to make public copies for ensuring further access. Each of these tactics can outmaneuver China's Great Firewall of censoring speech, at least for a while. These Chinese community archive repos (and their clones) documenting personal COVID-19 experiences surface ideological differences between US-based platforms, the Chinese government's Internet censorship efforts, and the desire of communities to document and share their personal experiences of the pandemic. For scholars concerned with power, access to knowledge, self-determination efforts through commemoration, what can we learn from how these community archives of personal COVID-19 experiences use platforms like GitHub?

In the aftermath of earthquakes and pandemics, platforms like WeChat and Twitter become public information infrastructures that shape and are shaped by new information practices. Platforms have also become venues where communities coordinate and commemorate trauma, violence, and loss with documentation. Information historian Megan Finn, has shown in her work on information practices before and after disasters, private platforms like Facebook convene groups of people in novel ways with algorithms, location data, and affinity groups [46]. However, the ways in which platforms convene such publics are generally unknown, their mechanisms are 'black boxed', providing outsiders with low visibility into their construction, development, and evaluation. We can add another process to the black box with the COVID-19 community archives 'going dark' because researchers have little ways of knowing when government censorship via takedown requests has occurred, or whether ISPs are blocking access to URLs if platforms do not publish evidence of complying with such requests. It is too soon to confirm whether the repos that went dark and are now 404s are the result of take down requests from government, censorship intimidation or police custody, or simply the chilling of publishing public repos.

Whether it has been removed or gone private, the 404 of /nCovMemory repo shows there are conflicting values of access, commemoration, and control over first-hand accounts of the coronavirus pandemic. Information science and STS scholars have examined the epistemic tensions that platforms represent when they are used as infrastructures to preserve heritage, document evidence, or to commemorate people we've lost [9], [21]. By specifically examining platform developments where data stewardship and digital preservation strategies are enacted by users of platforms, we can more fully understand the impact of platformization on our digital cultural memory and the future of archives [12].

These platform accounts also reveal the power of governance, sovereignty, and state censorship campaigns in and through features like public repos over digital cultural memory. The /wuhan2019 repo provides a lens for considering how, in an effort to combat censored news narratives, repo creators become historical curators of both COVID-19 accounts and subject to censorship themselves as platform users. In making commits and changes to the archive to protect these collected contributions, the contributor to /wuhan2019 authoritatively appraises the changing informational landscape and government suppression of health information in the media. On March 17, 2020, Lestweforget issued an update to /wuhan2019 noting the suspended additions of two media outlets coverage of the novel coronavirus. The commit described that due to the "quality" of the articles being published, items from *China Business News* and *Xinmin Weekly* would be stopped being included [45]. The /wuhan2019 repo updates consequently serve as records of reputability or lack thereof, and the performative (and powerful) actions of contributing to community archives. With a public repository, user activity of the creator and those participating in community archiving efforts are represented in the form of public notification and digital records.

Like the /nCovMemory repo, /wuhan2019 attempts to depict the realities of lived experience of coronavirus in Wuhan, Hubei province, and beyond. In contrast to a cultural landscape where narratives of experience were being erased by censorship and distorting societal representation and collective memory, Lestweforget contends the "INTERNET never forgets" [40]. The feelings and ideals motivating both COVID-19 community archives' repos are facing the confrontation between memory, state oppression, and access to information within platforms and across them. With community archives like /wuhan 2019 and /nCovMemory, we conclude by asking ourselves if platforms are being used as community archives, what can platforms learn from users that leverage GitHub's

features for connecting communities to information resources and knowledge by usurping Chinese Internet policy with a US-based platform?

## 7. Conclusion

We began this research to investigate platform dynamics in relation to memory practices and community archives.[4] By examining how GitHub is being used by volunteers for organizing COVID-19 community archives, we are able to show how platform affordances can be used to evade suppression of Chinese content published online with team translation organizing, cloning, and migrating content. Each of these emerging and innovative community archive practices uses code, documentation, and software repository features appear to be new strategies of memory and resistance using the GitHub platform. But, as the disappeared repos illustrate, China continues to suppress personal narratives and media coverage of the coronavirus pandemic. Platforms, like GitHub, can leveraged to extend suppression by putting users and their publicly available content at risk because governments can be users of platforms too.

Whether enforcing the power of states, disciplining bodies through institutions, or building counter-narratives that resist hegemony, archives have always been implicated in projects of domination. Researchers of digital cultural memory should pay attention to the platformization of memory practices and community archives. These confrontations over accessing personal accounts of the pandemic in China by organizing community archives on GitHub remind us again that it remains to be seen where platforms see themselves in this struggle over what will be remembered and what we'll sooner forget.

## 8. References

[1] A. Acker and M. Chaiet, "The weaponization of web archives: Data craft and COVID-19 publics," *Harvard Kennedy School Misinformation Review*, vol. 1, no. 3, Sep. 2020, doi: 10.37016/mr-2020-41.

[2] J. Hua and R. Shaw, "Corona Virus (COVID-19) 'Infodemic' and Emerging Issues through a Data Lens: The Case of China," *IJERPH*, vol. 17, no. 7, p. 2309, Mar. 2020, doi: 10.3390/ijerph17072309.

[3] B. Xie *et al.*, "Global health crises are also information crises: A call to action," *Journal of the Association for Information Science and Technology*, vol. n/a, no. n/a, Mar. 2020, doi: 10.1002/asi.24357.

[4] R. Broderick, "Chinese WeChat Users Are Sharing A Censored Post About COVID-19 By Filling It With Emojis And Writing It In Other Languages," *BuzzFeed*, Mar. 11, 2020. https://www.buzzfeednews.com/article/ryanhatesthis/coronavirus-covid-chinese-wechat-censored-post-emojis (accessed Jun. 23, 2020).

[5] R. Broderick, "As Chinese Internet Users Try To Track The Coronavirus, Their Government Is Tracking Them," *BuzzFeed News*, Feb. 07, 2020. https://www.buzzfeednews.com/article/ryanhatesthis/as-chinese-internet-users-try-to-track-the-coronavirus (accessed Jul. 13, 2020).

[6] GitHub, "GitHub features: the right tools for the job," *GitHub*, 2020. https://github.com/features (accessed Jul. 12, 2020).

[7] A. Acker, "A Death in the Timeline: Memory and Metadata in Social Platforms," *Journal of Critical Library and Information Studies*, vol. 2, no. 1, p. 27, 2018.

[8] A. V. Banchik, "Disappearing acts: Content moderation and emergent practices to preserve at-risk human rights–related content," *New Media & Society*, p. 1461444820912724, Mar. 2020, doi: 10.1177/1461444820912724.

[9] T. Sutherland, "Making a Killing: On Race, Ritual, and (Re)Membering in Digital Culture," *Preservation, Digital Technology & Culture*, vol. 46, no. 1, pp. 32–40, 2017, doi: 10.1515/pdtc-2017-0025.

[10] T. Poell, D. B. Nieborg, and J. van Dijck, "Platformisation," *Internet Policy Review*, vol. 8, no. 4, pp. 1–13, 2019, doi: 10.14763/2019.4.1425.

[11] J. van Dijck, "Seeing the forest for the trees: Visualizing platformization and its governance:," *New Media & Society*, Jul. 2020, doi: 10.1177/1461444820940293.

[12] A. Helmond, "The Platformization of the Web: Making Web Data Platform Ready," *Social Media + Society*, vol. 1, no. 2, Sep. 2015, doi: https://doi.org/10.1177/2056305115603080.

[13] G. Langlois, G. Elmer, F. McKelvey, and Z. Devereaux, "Networked Publics: The Double Articulation of Code and Politics on Facebook," *Canadian Journal of Communication*, vol. 34, no. 3, Art. no. 3, Aug. 2009, doi: 10.22230/cjc.2009v34n3a2114.

[14] A. Acker and B. Beaton, "How Do You Turn a Mobile Device into a Political Tool?," Jan. 2017, doi: 10.24251/HICSS.2017.281.

[15] G. Austin, *Cyber Policy in China*. John Wiley & Sons, 2014.

[16] A. Cocciolo, "Community Archives in the Digital Era: A Case from the LGBT Community," *Preservation, Digital Technology & Culture (PDT&C)*, vol. 45, no. 4, pp. 157–165, Feb. 2017, doi: 10.1515/pdtc-2016-0018.

[17] M. Caswell, M. Cifor, and M. H. Ramirez, "'To Suddenly Discover Yourself Existing': Uncovering the Impact of Community Archives," *The American Archivist*, vol. 79, no. 1, pp. 56–81, Jun. 2016, doi: 10.17723/0360-9081.79.1.56.

[18] J. Lingel and D. Boyd, "'Keep it secret, keep it safe': Information poverty, information norms, and stigma,"

*Journal of the American Society for Information Science and Technology*, vol. 64, no. 5, pp. 981–991, 2013, doi: 10.1002/asi.22800.

[19] D. K. Wakimoto, C. Bruce, and H. Partridge, "Archivist as activist: lessons from three queer community archives in California," *Arch Sci*, vol. 13, no. 4, pp. 293–316, Dec. 2013, doi: 10.1007/s10502-013-9201-1.

[20] A. Acker, "Radical Appraisal Practices and the Mobile Forensic Imaginary," *Archive Journal*, Nov. 2015, Accessed: Jan. 08, 2018. [Online]. Available: http://www.archivejournal.net/?p=6204.

[21] A. Acker and J. R. Brubaker, "Death, Memorialization, and Social Media: A Platform Perspective for Personal Archives," *Archivaria*, vol. 77, no. 0, pp. 1–23, May 2014.

[22] "GitHub Enterprise: a smarter way to work together," *GitHub*, 2020. https://github.com/enterprise (accessed Jun. 23, 2020).

[23] "Stack Overflow Developer Survey 2018," *Stack Overflow*, 2018. https://insights.stackoverflow.com/survey/2018/?utm_source=so-owned&utm_medium=social&utm_campaign=dev-survey-2018&utm_content=social-share (accessed Jun. 23, 2020).

[24] "Microsoft + GitHub = Empowering Developers," *The Official Microsoft Blog*, Jun. 04, 2018. https://blogs.microsoft.com/blog/2018/06/04/microsoft-github-empowering-developers/ (accessed Jun. 23, 2020).

[25] GitHub, "Plans for every developer," *GitHub*, 2020. https://github.com/pricing (accessed Jun. 23, 2020).

[26] GitHub, "The State of the Octoverse," *The State of the Octoverse*, 2020. https://octoverse.github.com/ (accessed Jul. 13, 2020).

[27] T. Qu, "How GitHub became a bulletin board for Chinese tech worker complaints," *Quartz*, Apr. 09, 2019. https://qz.com/1589309/996-icu-github-hosts-chinese-tech-worker-complaints/ (accessed Jul. 13, 2020).

[28] J. Horwitz, "China's fierce censors try a new tactic with GitHub—asking nicely," *Quartz*, Jun. 28, 2016. https://qz.com/718465/chinas-fierce-censors-try-a-new-tactic-with-github-asking-nicely/ (accessed Jul. 14, 2020).

[29] J. Liu, "China's data localization," *Chinese Journal of Communication*, vol. 13, no. 1, pp. 84–103, Jan. 2020, doi: 10.1080/17544750.2019.1649289.

[30] J.-C. Plantin and G. de Seta, "WeChat as infrastructure: the techno-nationalist shaping of Chinese digital platforms," *Chinese Journal of Communication*, vol. 12, no. 3, pp. 257–273, Jul. 2019, doi: 10.1080/17544750.2019.1572633.

[31] D. Bamman, B. O'Connor, and N. Smith, "Censorship and deletion practices in Chinese social media," *First Monday*, vol. 17, no. 3, Mar. 2012, doi: 10.5210/fm.v17i3.3943.

[32] R. MacKinnon, "Flatter world and thicker walls? Blogs, censorship and civic discourse in China," *Public Choice*, vol. 134, no. 1, pp. 31–46, Jan. 2008, doi: 10.1007/s11127-007-9199-0.

[33] A. Ben-David and A. Amram, "The Internet Archive and the socio-technical construction of historical facts," *Internet Histories*, vol. 2, no. 1–2, pp. 179–201, Apr. 2018, doi: 10.1080/24701475.2018.1455412.

[34] Memoryhonest, "memoryhonest/nCovMemory-en, 'Our Team,'" *GitHub*, Feb. 24, 2020. https://github.com/memoryhonest/nCovMemory-en (accessed Jun. 08, 2020).

[35] J. Li, "Chinese citizens are racing against censors to preserve coronavirus memories on GitHub," *Quartz*, Mar. 02, 2020. https://qz.com/1811018/chinese-citizens-use-github-to-save-coronavirus-memories/ (accessed Jun. 08, 2020).

[36] Internet Archive, "Wayback Machine, Web crawls of https://github.com/2019ncovmemory/nCovMemory," *Wayback Machine*, Jun. 08, 2020. https://web.archive.org/web/*/https://github.com/2019ncovmemory/nCovMemory (accessed Jun. 08, 2020).

[37] P. Zhang, "Chinese activists detained after sharing censored coronavirus material on crowdsourcing site GitHub | South China Morning Post," *South China Morning Post*, Apr. 25, 2020. https://www.scmp.com/news/china/politics/article/3081569/chinese-activists-detained-after-sharing-censored-coronavirus (accessed Jun. 08, 2020).

[38] J. Li, "Chinese internet users who uploaded coronavirus memories to GitHub have been arrested," *Quartz*, Apr. 27, 2020. https://qz.com/1846277/china-arrests-users-behind-github-coronavirus-memories-page/ (accessed Jun. 08, 2020).

[39] "Hubei Timeline," *Johns Hopkins Coronavirus Resource Center*, 2020. https://coronavirus.jhu.edu/data/hubei-timeline (accessed Jun. 23, 2020).

[40] "lestweforget - Overview," *GitHub*, 2020. https://github.com/lestweforget (accessed Jun. 23, 2020).

[41] "lestweforget/wuhan2019," *GitHub*, 2020. https://github.com/lestweforget/wuhan2019 (accessed Jun. 23, 2020).

[42] "Update README.md · lestweforget/wuhan2019@64d6653," *GitHub*, 2020. https://github.com/lestweforget/wuhan2019/commit/64d665306ce906cdf09506538059059e48065709 (accessed Jun. 23, 2020).

[43] H. Adlakha, "Fang Fang: The 'Conscience of Wuhan' Amid Coronavirus Quarantine," *The Diplomat*, Mar. 23, 2020. https://thediplomat.com/2020/03/fang-fang-the-conscience-of-wuhan-amid-coronavirus-quarantine/ (accessed Jun. 23, 2020).

[44] Y. Lianke, "Yan Lianke: What Happens After Coronavirus? | Literary Hub," *LitHub*, Mar. 11, 2020. https://lithub.com/yan-lianke-what-happens-after-coronavirus/ (accessed Jun. 23, 2020).

[45] Lestweforget, "wuhan2019/README.md," *GitHub*, Jun. 14, 2020. https://github.com/lestweforget/wuhan2019/blob/master/README.md (accessed Jul. 14, 2020).

[46] M. Finn, *Documenting Aftermath: Information Infrastructures in the Wake of Disasters*. Cambridge, Massachusetts: The MIT Press, 2018.