

Losing the API: Developing Novel Methods for Scraping Black Twitter

Andrew Lowe Mohammed <alowemoh_at_umd_dot_edu>, The University of Maryland, College Park 
<https://orcid.org/0009-0008-1358-2587>

Alisa Hardy <ahardy_at_umd_dot_edu>, The University of Maryland, College Park  <https://orcid.org/0000-0002-2853-1941>

Rianna Walcott <rwalcott_at_umd_dot_edu>, The University of Maryland, College Park  <https://orcid.org/0000-0002-2261-8616>

Tynesha McCullers <tmccull2_at_ncsu_dot_edu>, North Carolina State University  <https://orcid.org/0009-0007-4243-8974>

Abigail Vazquez Rosario <rosario_at_umd_dot_edu>, The University of Maryland, College Park  <https://orcid.org/0000-0002-3507-5753>

DOI: <https://doi.org/10.63744/gxtb2db7xazp>

Abstract

As digital platforms continually evolve, rapid changes to platform affordances quickly render digital tools and data collection methods obsolete. Researchers of digital culture therefore must proactively adapt to the ephemerality of data. This paper examines these challenges within the context of Twitter (X) following its 2022 acquisition by Elon Musk, and the subsequent limiting of access to the API for data collection. Using a combination of manual data-collection practices and Zeeschuimer [Peeters 2025], a browser extension that collects social media data while browsing, researchers developed a novel data collection method, and model methodological adaptability within shifting digital terrains.

Introduction

In the constantly changing landscape of digital platform affordances, there is a need to develop research methods that withstand such changes. The rapid pace of changes to platforms can quickly render digital methods and tools unusable for researchers in the digital humanities. It is therefore imperative that scholars of digital culture react proactively to the ephemerality of data when undertaking data collection to support analysis of digital spaces. This article explores how these dynamics play out in the changing landscape of Twitter (X) after Elon Musk's acquisition of the platform in 2022. Through development and expansion of a novel data collection method, we demonstrate the need for methodological adaptability within shifting digital terrains.

1

On October 31, 2022, Twitter users were met with a peculiar “trick or treat” as Musk, owner of the well-known electric-car brand Tesla, entered discussions about acquiring the platform. Rumors surfaced of Musk becoming Twitter's new owner, leaving many users concerned about the platform's future. Social media users could not have foreseen the arduous seven-month dramedy that would follow. Months of conversations about Twitter's ownership culminated in Musk's purchase of the social network platform, which he rebranded as X on July 23, 2023 [Musk 2023]. Our study concerns the migration of users away from Twitter in the wake of Musk's acquisition of the platform and particularly focuses on *Black* Twitter's activity between March 2022 to July 2023. Our attention on Black Twitter as a case study is due to its position as a discourse community made up of multiple heterogeneous subcommunities that nevertheless behave as a collective. Black Twitter's reputation for innovation and virality positions its members as tastemakers, with many ‘high-net worth individual’ (HNWI) members whose behaviors initiate ‘herd effect’ migration patterns [Kumar, Zafarani, and Liu 2011];[Walcott 2024]. Evaluating the behaviors of such users provides insights into broader migration

2

narratives occurring at a platform level. In addition, Twitter's crisis of governance and subsequent rapid platform changes instigated meta-narrative discussions about platform migration within the Black Twitter community in particular that we were able to pinpoint due to our long-standing participation in the discourse collective. By focusing on community affective responses to platform-level disruption, we offer a replicable test case for researchers seeking data samples from a digital counterpublic. Over the course of our study the platform changes impacting the community also restricted our data collection methods, and our response to those restrictions forms the basis of this paper.

To contextualize the meta-narrative platform migration discourse we observe on Black Twitter, we acknowledge that Black collective community practices on digital platforms predate 'Black' Twitter. Prior to Twitter's inception, Black people navigated to blogs and forum websites to connect and build community with one another. As [Banks 2018] discusses, during the mid-to-late 1990s, United States' based websites such as NetNoir and BlackVoices were catalysts for fostering online communities for Black digital users. Unfortunately, these and similar websites only lasted a few short years due to challenges with site maintenance and server hosting costs, leading to these platforms being bought out [Brock 2020]. Black digital users regularly reimagine social networks in the early 2000s, including MySpace and Facebook, which emerged as platforms for community building. While both sites offered Black users' digital space to connect with one another, the sites' features imposed age limits and educational status requirements for users interested in utilizing them. Blogging websites such as Afrobella, Jack and Jill Politics, Racialicious, Prometheus6, and WhatAboutOurDaughters became homestays for Black online users in the United States. Eventually, traffic to these blog sites were overshadowed by social media platforms.

However, Black digital engagement on blogs and forums cultivated practices we see on Black Twitter, now X, today. Twitter's specific technological affordances are particularly compatible with existing Black rhetorical practices, as "a source of user-generated culturally relevant content, combining social media features and reach to share information" [Brock 2012, 530]. Black Twitter is "an online gathering of Twitter users," most of whom identify as Black and use the platform to "perform Black discourses, share Black cultural commons places, and build social affinities" [Brock 2020, 81]. This online community was founded in the platform's affinity for immediacy, lending itself to several discursive strategies — signifyin', call-and-response, and storytelling — rooted in offline Black communities [Brock 2012], [Brock 2020]; [Florini 2014], [Florini 2019]. Black Twitter exemplifies the reconstruction of physical spaces of Black reality-building, such as the "barbershop," through digital networks by employing the legacies of oral communication that comprise Black rhetorical tradition [Steele 2016, 2]. Black Twitter users expertly reimagine signifyin', when social media posts embody multiple messages that can be understood through shared cultural knowledge emerge on digital spaces by using platform features such as hashtagging, image sharing, and strategic use of metric tools like retweeting and subtweeting [Florini 2014, 226–227]. This online community is a space for Black users to engage in dialogues about social issues relevant within Black communities and allows for Black creative expression and innovation within a white supremacist structure. Scholars have studied Black Twitter as an area for social change, counterpublics, and pleasurable feelings [Squires 2002]; [Florini 2014];[Jackson et al. 2020];[Brock 2020];[Steele 2018].

We use Black Twitter as a case study to examine how digital humanities scholars can continue researching digital counterpublics amidst technological changes that further obscure and marginalize those voices. As part of our practice of excavating and centering counterpublic perspectives, despite the official rebrand we acknowledge the lasting impact of the name "Twitter" amongst Black users, and use it here in deference to its continued relevance to the 'Black Twitter' community we are concerned with. We understand Black users' continued use of 'Twitter' as an act of resistance to undesired platform changes, a way of maintaining a countercultural positioning in refusal of Musk's rebranding and encroaching changes to the dominant reputation, functioning, and demographics on the platform.

Investigating these responses to change necessitated a conversation about method, as the very nature of the changes we sought to investigate impacted our ability to gather data. This article discusses how the Black Communication and Technology Lab (BCaT) collaborative team ^[1] sought to counteract increasingly limited access to community discourse by developing novel research methods. As we discuss our experiences analyzing user movement amid platform changes, it is important to establish what we mean by "Black digital migration." Black digital migration is when users within the African diaspora navigate the different push and pull factors that alter the habitability of a digital space, expressing their desires to remain on the platform or migrate to a different one [Walcott 2024]. Black digital migration is

rooted in Black users' refusal to be excluded from digital practice, specifically the use of social media platforms to organize and resist hegemonic digital landscapes [Everett 2002];[Lu and Steele 2019]. This refusal is also evident in our research praxis to continue studying Black Twitter as an act of social change even if back-end data availability is inaccessible.

Between March to July 2023, we gathered Tweets by Black users that reflected their sentiments toward the platform changes, and corresponding migratory intentions, from remaining on the site “till the wheels fall off” or migrating to competing platforms. While we had planned to use automated scraping tools such as Twitter Archiving Google Sheets (TAGs), we were limited in access to the back-end application program interface (API) data on Twitter following the monetization of its access in February 2023. API is a software intermediacy from two disconnected applications that allows developers to programmatically access and interact with a platform's data and functionalities. Platform API's connect backend digital code with frontend user interfaces. APIs are useful for extracting tweets based on customized parameters such as hashtags, user IDs, or timestamps [Perera et al. 2010]. Companies and developers created tailored tools like TAGS that interacted with Twitter's API to study the spread of messages over time and across networks.

This article evaluates the efficacy of incorporating manual scraping practices *alongside* automated data collection and data visualization to demonstrate how scholars can employ adaptive methods when met with barriers that impact their ability to gather data. We employed mixed-method approaches, including (1) manual collection of relevant tweets and associated metadata collaboratively compiled using spreadsheets; (2) Zeeschuimer, the automated over-the-shoulder data collection tool for analyzing social media content, and (3) Tableau, a data visualization tool that allows for research storytelling using interactive graphics.

First, we discuss how digital scholars have employed both automated data collection and manual scraping as methods to study Black Twitter phenomena. Second, we outline the possibilities and implications for using both manual scraping and the automated digital tools. Third, we describe our use of Tableau to incorporate graphs analyzing and dissecting patterns from the data we collected. We conclude with a discussion regarding mixed-methods approaches to study online platforms amid changing data access.

Automated Data Collection vs. Manual Scraping

Scholars have employed both automated data collection and manual scraping techniques to analyze Twitter and noted the possibilities and limitations of using both independently. Automated data collection consists of harvesting data automatically via computational tools and storing the information within a system or database. Scholars have used automated data techniques to gather information from Twitter using the app's publicly available API. For instance, [Levenberg et al. 2018] catalog a variety of tools available for scholars to collect, code, and analyze tweets using automated scraping methods. The edited collection *Research Methods for the Digital Humanities* demonstrates how software such as TAGs, Apify, and Google Chrome's Twitter Scraper extension computationally collect data based on the API to gather information on public conversations about ongoing topics and cultural data. In addition to collecting the content of the Tweets themselves, these digital tools also harvest associated metadata, including hashtags, 'threaded' tweets, replies from other users, and any included images and videos, and metric information about the user such as their numbers of followers and users they follow.

Another strategy that scholars have used to gather big data is to buy millions of tweets directly from Twitter. In *#Hashtag Activism*, [Jackson et al. 2020] worked with Twitter to gather data from counterpublics via hashtags like #YesAllWomen, #WhyIStayed, and #MeToo. Their mixed-methods approach combines computational tools that can extract incredibly large amounts of digital data with qualitative analysis, examining nearly four million tweets and organizing them into networks to conduct a discursive analysis highlighting the “social and political labor undertaken by hashtag activists as they work to tell stories and make meaning” [Jackson et al. 2020]. Similarly, the authors of the “Beyond the Hashtags” report, which examines big data from tweets following the 2016 Black Lives Matter movement, purchased more than 40 million tweets from Twitter [Freelon et al. 2016]. They analyzed the data using the Python module Twitter Subgraph Manipulator and separated the tweets into nine time periods. Other scholars have used the cloud-based text and social network analyzer Netlytic to collect and analyze big data from Twitter. This software uses APIs to collect publicly

accessible posts from Twitter, YouTube, and RSS feeds and uploads these datasets into comma-separated values (CSV) or Google Sheets.

Following automated data collection, researchers must subsequently “clean” their collected datasets to remove irrelevant information such as advertisements, spam, or inclusions otherwise unrelated to the research parameters. This also includes procedurally fixing or removing incorrect, corrupted, incorrectly formatted, duplicate, or incomplete data within a dataset. This process is arduous and time consuming, particularly when working with large datasets requires a lot of time from the researchers, especially those doing big data analysis who need to limit the amount of incorrect information in their dataset.

12

While web-scraping and analysis tools allow quantitative access to user data, they are made defunct by restricted access to APIs. In February 2023 free access ended to Twitter’s API, and a paid tier program launched in March 2023 [X Developers 2023]. Citing “self-serve” access, Twitter developers began limiting free API usage to 1,500 tweets a month, with steep pricing increases for paid tiers of access from \$3 to \$16 a month (About X Premium, 2023), which can be prohibitively expensive for small groups and individual researchers. In the case of this study, cleaning computationally collected data was incompatible with our needs and available resources – our search parameters and keywords used in the search returned a high volume of irrelevant content; for example tweets that contained the word ‘black’ but related to homonyms such as the colour black rather than the racial category. Similarly, searching for different social media platform names such as ‘Mastodon’ or ‘Bluesky’ returned many tweets that discussed those respective platforms, but may not have been related to our desired context of Black users’ migration intentions or affective relationships towards those platforms. Finally, computational processes also do not allow for restricting data collected to tweets generated by a specific community ability (ethnic or otherwise). Some limitations of using Twitter’s search API in research are outlined by [González-Bailón et al. 2014], who note that the search API does not provide an exhaustive search of Tweets, but rather an index of recent tweets (posted within 6-9 days prior to date of retrieval) and therefore needed to be set up to collect relevant data at the beginning of the instance of its occurrence. The ‘search API’ is typically used by researchers for targeted retrieval, as opposed to the ‘firehose API’ which delivers a live feed of all new data in near real-time. They also find that the search API over-represents central users, as ‘smaller samples do not offer an accurate picture of peripheral activity’ due to the higher influence of snowballing in identifying relevant nodes for inclusion. This may obscure relevant content from computation data collection methods, particularly when peripheral communities of users. Our intervention addresses this problem of over-representation, as our use of multiple researcher accounts and iteratively defined parameters for inclusion allowed for collection of tweets that may have been missed in automated scraping processes.

13

To this end, we employed an analog or ‘hand coded’ data collection process. This involved manually entering our desired data into a collaborative Google spreadsheet — in this case, the data we collected included user metadata such as usernames, verification status, bio, geo-location if available, follower counts as a metric to establish their popularity as basis for inclusion, the race and gender identity as assumed by the research team, the tweet content and ID, and time/date posted. We also included engagement metrics at the time of collection, such as any hashtags used, ‘like’, ‘comment’, ‘retweet’, ‘bookmark’ and ‘quote tweet’ counts. Lastly, we included any significant researcher observations and the initials of the researcher who ‘scraped’ the tweet, alongside our date of collection. Manual data collection is more time-consuming than computational scraping, but reduced the time spent on data cleaning by generating only data relevant to the study.

14

In manually scraping Black Twitter, many researchers employ screenshots as a method of data collection. This is a fundamentally Black archival practice of gathering, describing, and digitizing Black collections or materials that speak to the complex ways Black life is lived, documented, and remembered [Hartman 2008];[Helton et al. 2015];[Sutherland and Collier 2022]. As [Clark 2020] explains, screenshotting is a culturally Black approach “for collecting, storing, filing, and retrieving proof of online utterances” (p. 203). These online methods mirror offline ones such as scrapbooking family milestones, writing about narratives of enslavement, and photographing material realities of the Black lived experience, and are grounded in ‘receipts culture’ (Clark, 2020). Receipts culture allows scholars to document the realities of Black Twitter and can be used when studying Black online spaces beyond Black Twitter to read multiple meanings inscribed in these visual artifacts <https://www.zotero.org/google-docs/?aQgEwV>.

15

Our methods did not involve screenshotting images of user tweets; we privileged user privacy throughout the course of our data collection and analysis. Instead, we maintained hyperlinks to the original tweets in our spreadsheet, meaning that tweets deleted by the original poster were no longer usable within our dataset at the time of writing; allowing deletion to represent removal of consent to publish. We were influenced by how Black user identities are constituted, captured, recorded, and processed amid modern landscapes of data surveillance [Leaver 2017]. Separating users into categories based on their Twitter popularity also helped protect user privacy as we could obfuscate individual identities, even anonymizing users with celebrity status outside of the platform. This practice makes clear when it would be inappropriate to quote user tweets directly and highlights when it is important to paraphrase user sentiments. Since we were not actively seeking user permission to collect these tweets, identifying users — even celebrities — by their individual usernames would be breaching our ethics with regard to digital privacy. When records of past surveillance and data repositories like our dataset are made public, they take on new meaning [Wisser et al. 2016]. Privacy is an important concern since identifying individuals — especially those from marginalized groups — positions them in the crosshairs of criticism from the hegemonically powerful.

16

What we offer to digital humanities is a data collection technique that affords critical and nuanced ways to observe digital culture amid constantly shifting platform ecologies. Our collaboration as members who are all participants of Black Twitter but encompass a range of social identities was a component of our research design. Our collaborative practice mitigated the impact of black box algorithms that curate individual user experiences, and encouraged reflexive dialogue about our orientation to the site under study. We were also intentional in not reducing Black users or digital performances of Black life to datasets and code ([Johnson 2020]), and this collaborative data collection that unfolded over multiple sessions of parallel work allowed for the research team to assess together that users met the criteria for inclusion in the dataset (for example, in the cases of users that a team member believed was Black based on various context clues and signifiers, but required dialogic confirmation from the rest of the team due in the absence of embodied confirmation of Black identity). We were also able to share contextualizing observations in real time; accounting for representation of Blackness beyond data points in narrativizing Black users' experiences. Midway through our data collection process, in July 2023, Twitter implemented rate limits of 600 tweets. Rate limiting meant that viewers could not see new tweets on their feeds once they reached the limit of 600 per day. The refresh rate limit halted our progress significantly, as we would regularly be stopped during the course of our data collection due to reaching the 600-tweet limit. Additionally, the personalized Twitter algorithms associated with each researcher's account meant that different sections of the discourse were highlighted over others on our individual feeds. These constraints and our response to them are also indicative of our collective praxis — our collaboration allowed for the flexible and adaptive response to a (thankfully short-lived) structural impediment to our data collection that may have stymied a single researcher.

17

Data Collection Using Manual Scraping and Zeeschuimer

The care we center in our methodology draws from Black feminist (and digital) methods, which examine digital phenomena in the context of race, gender, power, and other systemic factors interacting in a matrix of relations [Collins 2000]; [Noble 2018]. [Steele et al. 2023] developed a Black feminist intervention of “radical intentionality” based on ethics of radical love, justice, care, and liberation. These ethics seek “to remedy the tensions at the intersection of community and the academy” while breaking through “the social change industrial complex of the neoliberal academy” [Steele et al. 2023, 3]. Similarly, we strive to be accountable to those we studied on Twitter, and those with whom we established a community. Radical intentionality centers Black traditions of speaking truth to power, acknowledging the status quo and power dynamics present on Twitter and offline. To this end, we center critical perspectives in our analysis that engage the structural inequalities that impact how Black users navigate and leverage platforms to build community online, express cultural identity, and resist structural inequalities that both predate and are related to new platform governance, such as racism, sexism, surveillance and censorship practices.

18

We embraced grounded theory as a methodology to empower the wide range of expertise and familiarity with Black Twitter between research team members embodying different lived experiences. Grounded theory encourages developing theory from systematic collection of data to explain observed cultural norms [Glaser and Strauss 1967]; [Glaser 1992]. Grounded theory is a set of “related but distinguished methods, often sharing names, spaces, and points of view but also often disagreeing and even defining themselves in opposition to one another” [Martin et al. 2018, 11–

19

22]. In our analysis, this defined our approach as we developed our theorization of Black digital migration practices and affective relationships to platforms as geographical sites based on insights gathered from the tweets. We mitigated the influences of researcher preconceptions by refining our dataset and search terms based on observations over different stages of data collection. Grounded theory enabled us to consider the implications of our findings as we gathered our data, supporting adaptivity and reflexivity in our research methods as a multi-staged, developing process.

The BCaT team took detailed notes corresponding with specific dates and terms relevant to how counterpublics were discussing the state of the platform. We compiled the dates based on our “Musk Acquisition Timeline,” a chronological timeline of the changes to Twitter’s governance and platform affordances we developed using datapoints drawn from ‘Twitter Support’ updates and news media articles from sources such as the *Washington Examiner*, *The Verge*, and *TechCrunch* between March 2022 through April 2023. As participant observers, we were able to extract recurring terms and phrases that were most relevant to the target demographic of Black Twitter users. Using Twitter’s advanced search features, we tailored our search to align with specific date ranges of high-activity and major updates as dictated by our timeline.

Twitter’s advanced search, added to the platform in 2011, allows users to input specific keywords, phrases, dates, and time frames to search for posts centering specific discourses. Our use of Twitter’s advanced search feature exemplifies what Caliandro states is “following the thing, the medium, and the natives,” where we use the interface’s ordering principles to guide research [Caliandro 2018, 560]. This meant inputting events from our timeline and collaboratively agreed search terms through the advanced search feature to extrapolate targeted discourses. Some of the terms we searched were specific names of other platforms such as “Spill,” “Threads,” and “BeReal,” along with Black racial signifiers such as “ApartheidClyde,” “BlackTwitter,” and “#TwitterHomegoingService.” We also searched terms associated with the shift in Twitter’s ownership such as “Twitter Blue,” “Elon Musk,” and “verified.” Over the data collection period, we adapted our search terms iteratively to exclude low-engagement terms. The term bank consisted of keywords and phrases surrounding Musk’s new ownership of the social media platform, focusing on older social media platforms, contemporary social media platforms, self-definitions of Black Twitter users, Black migration narratives, platform-ownership narratives, and responses to feature changes on Twitter. Creating our term bank was an iterative process: we regularly added terms that we saw appearing frequently. We also coded the contents of the Tweets collected based on mutually agreed codes: sentiment codes for the specific emotions that appeared in the tweets (Positive/Negative/Neutral/Ambiguous) and descriptive codes to identify and summarize overarching recurring themes(s) (Political/Humor/Satire/Informative/Interface Change/Migration).

The parameters for including tweets in our dataset required that they were primarily by Black users with at least 100 likes, reposts, and/or quote tweets. We searched terms both with and without a hashtag in Twitter’s search engine, accounting for the specific context in which these tweets emerged [Florini 2019].

As our manual scraping methods faced challenges in the wake of affordance changes on Twitter, we sought out other methods to continue engaging with Black Twitter. Along with manual data collection, the BCaT team also employed automated data collection through the use of Zeeschuimer. This digital tool is a browser extension that monitors the internet traffic of the researcher’s interface while they are browsing social media sites like Twitter, TikTok, and Instagram <https://www.zotero.org/google-docs/?vwBIYE>. Zeeschuimer gathers data about the items in a platform’s web interface and allows a researcher to upload this information to 4CaT, a tool that can be used to analyze and process data from social media <https://www.zotero.org/google-docs/?KaRNM1>. Zeeschuimer is an alternative to conventional scraping or API-based data collection as it metaphorically ‘looks over the researcher’s shoulder’ to gather data from the tweets that appear on their timeline while scrolling the site on a desktop browser.

We used computational methods comparatively alongside our manual data collection to inform our refinement of the data collection process. We used Zeeschuimer to scrape tweets using a ‘clean’ researchers’ account and the same search parameters in advanced search as with our manual collection process. The resultant dataset was transferred to 4CaT for analysis, after collecting approximately 4,000 tweets. This exploratory exercise was helpful for several reasons. First, we identified large clusters of content centered around the dates highlighted in our interactive timeline, which confirmed the credibility to our chosen time frames. Second, we found that the aggregate dataset required a

prohibitively time-consuming amount of ‘cleaning’; results were less targeted and related to our research questions than within our manually collected dataset. Scraping just the names of websites, for example, would most often pull results of users simply discussing their usage of those sites.

Data Storytelling of Black Digital Migration

Data visualization tools are digital applications that convert numeric and/or textual data into variable tables, charts, word clouds, and other images. Although there are many data visualization tools available in digital humanities research, our team chose to use Tableau due to its analytic platform that makes it accessible to researchers to explore, manage, organize, and interact with a variety of data. Tableau is apt for data storytelling, creating interactive workbooks that allow researchers and audiences to see and understand the data associated with a study easily. Using Tableau, our team was able to recreate the story of how Black users responded to Musk’s acquisition of Twitter, focusing on the timeline, characters, and affects surrounding the ownership changes.

Tableau also offers a variety of tools for data exploration, including different filters and editing tools to identify trends within our dataset. However, Tableau lacks its own data-cleaning systems, which increased our workload as we needed to organize, transform, and aggregate the data manually before inputting it into Tableau. The free version of Tableau has limited collaboration capabilities: the workbook can be accessed via the cloud, but researchers cannot work on it simultaneously. With Tableau’s paid account – free to users with an educational license, following verification of an institutional email address – working collaboratively is easier, but on the desktop version each researcher was required to have their own workbooks.

25

26



Figure 1. This visualization showcases a word cloud developed based on what social media platform Black Twitter users were discussing the most.

Our analysis is informed by data-driven storytelling, or the use of data to create visual narratives, offering a structured approach for communication insights within a dataset. Data visualization is employed in a vast array of projects in the digital humanities since visual aids enhance the thought processes of both researchers and audiences as they digest aggregate data []. By using the five-step ‘FLOAT’ method developed by [Ossom-Williamson 2021], which provides a framework for creating data-driven narratives. This was executed, for example, in the data visualizations we created with Tableau, using interactive bar charts, line graphs, and other charts to narrate the complex ways Black users responded to platform changes, which may serve as a guide for other digital humanities researchers. Having a structured and organized dataset was integral to our goal of data storytelling through Tableau.

27

The FLOAT method defines the processes required to create a data story: formulating a good research question, locating a sustainable data source, organizing data, analyzing data, and telling a data-driven story [Ossom-Williamson 2021]. We completed the first three steps iteratively, revisiting our research questions and figuring out what was feasible to answer based on the data we had collected. A data sheet helped us organize and analyze our data, making clear which parts of the data story were most relevant and important to communicate.

"Who all gon' be there?" - Black Twitter and Digital Migration

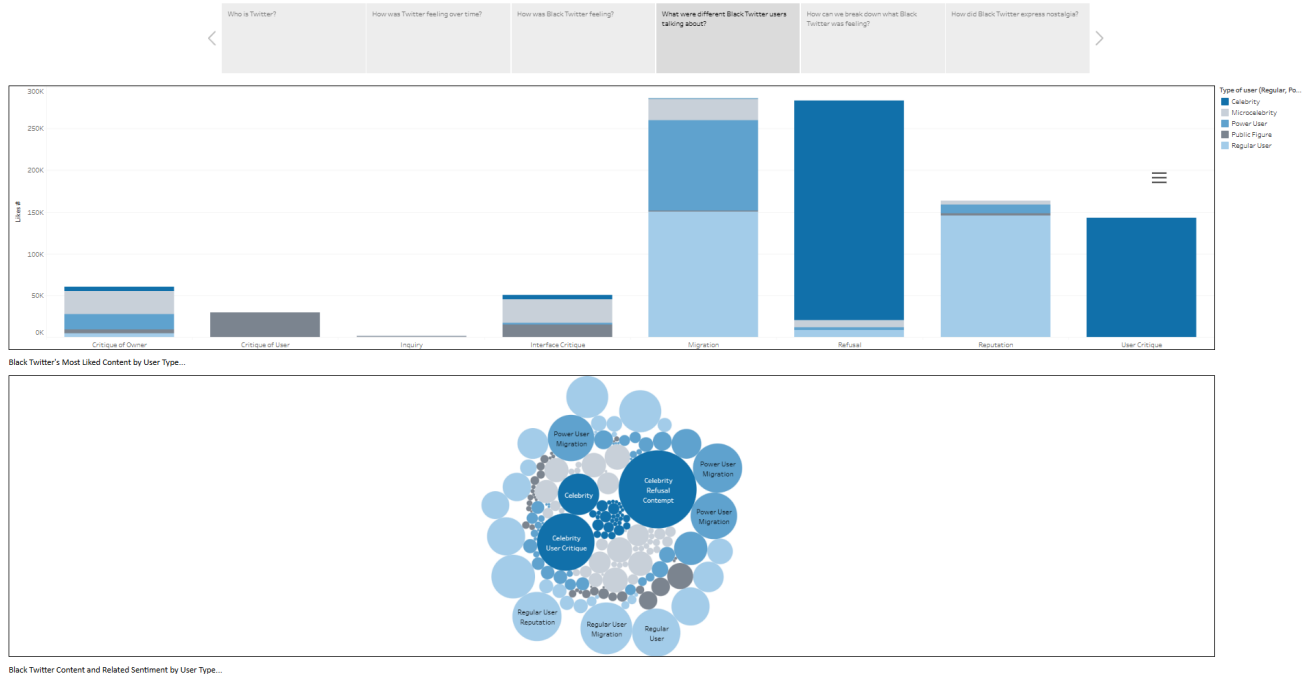


Figure 2. This visualization showcases Black users' primary descriptive codes and primary sentiment codes by type of user based on number of likes.

"Who all gon' be there?" - Black Twitter and Digital Migration



Figure 3. Timeline of Black Twitter's primary sentiment trends (top) and primary descriptive codes (bottom) alongside spikes in discursive activity.

Figure 3 showcases two of the visualizations that were made possible due to our early attention to the dates of Twitter's new ownership. Using our data appendices, descriptive content, and sentiment codes, we were able to tell the story of primary sentiments and content throughout the timeline alongside spikes in discourse about Twitter's new ownership. The upper visualization highlights the primary sentiments expressed by Twitter users based on the amount of likes individual tweets received. This information shows the most popular sentiments circulating through Black Twitter during specific time periods and events associated with ownership changes.

The creation of these visualizations highlights the importance of the data organization step in the FLOAT method. Several visualizations were created only after researchers added new categories of data during the development of a data sheet and data appendix. While automated methods are apt for collecting data revealing larger-scale trends, manual scraping allowed researchers to make judgments about which users were Black. This was done through analysis of users' self-identification, use of Black rhetorical practices, and researcher analysis of user profiles [Saldana 2021]. Our positionality as researchers from within the community under study was critical to our ability to evaluate our participants' belonging to identity categories that are signalled by symbolic and discursive performance. Naturally, researcher judgment of such fluctuating and contentious categories as ethnicity is not infallible, even when conducted collectively. Therefore, the team was careful not to overgeneralize by labeling accounts without any of these identity signifiers as "N/A" in our dataset.

Tweets Manually Scraped: 216

Tweets Computationally Scraped with Zeeschuimer: 3767

Scope of the Dataset: March 25th, 2022 - July 23rd, 2023

Key Variables: Verified, Date/Time, Location, Type of User (Race, Identity), Likes, Retweets, Quotes, Bookmarks, Followers, Comments, Visuals, Bio, Attachment, Observations

Twitter Observations

27 appearances of the term 'Black Twitter'	23 appearances of the term 'Spill'
19 appearances of the term 'Elon Musk'	17 appearances of the term 'Platform'
15 appearances of the terms 'Verified/verifications'	15 appearances of 'Apartheid Clyde'
3 appearances of the term '#BlackTwitter'	4 appearances of the term '#OnHere'
1 appearance of the term '#Twitterhomegoingservice'	1 appearance of the term '#TwitterTakeover'
1 appearance of the term 'BeReal'	4 appearances of the term 'Bird App'
1 appearance of the term 'Black Planet'	3 appearances of term 'Blue Tick'
1 appearance of the term 'Clubhouse'	1 appearance of the term 'Discord'
3 appearances of the term 'Facebook'	4 appearances of the term 'Free speech'
7 appearances of the term 'Instagram (IG)'	5 appearances of the terms 'Leave/Leaving AND Twitter'
3 appearances of the term 'Mastodon'	2 appearances of the term 'MySpace'
2 appearances of the term 'Reddit'	2 appearances of the terms 'RIP Twitter/#RipTwitter'
7 appearances of the term 'Social media'	4 appearances of the term 'TikTok'
2 appearances of the term 'Truth Social'	1 appearance of the term 'Tumblr'
2 appearances of the term 'Youtube'	10 appearances of the term 'X'
6 appearances of the term 'Rebrand'	

Figure 4. This image is of the "Data Sheet" which we developed as a guide for our study.

Discussion

Analyzing Musk's takeover of Twitter creates space for a critical comparison between computation and manual data collection methods. In this section, we compare computational and manual data collection to highlight the complex interplay of advantages and disadvantages each method offers in contemporary research. Using this comparison, we situated a method between computational and manual data collection that is adaptive to constantly shifting platform affordances.

Computational data collection automated through software like Zeeschuimer offers incomparable scale and speed of

data gathering. The massive amounts of data available through Zeeschuimer allow researchers to conduct large-scale analyses and mirror much of the analytical work done by businesses. One major advantage of using Zeeschuimer is that it does not require nearly as much labor from researchers as manual data collection does. A major limitation of this software is that it requires extensive cleaning of the data to correct or delete inaccurate information from the dataset such as advertisements or irrelevant tweets.

Our manual data collection was a far more laborious process, as it required each researcher to read and define metadata from individual tweets over an extended period of time. This made the manual data collection process feel more intimate for researchers as many were using their own Twitter accounts and are members of the Black Twitter community. In alignment with Black feminist digital methods, we cultivated an environment of joy, care, and reflexivity for both Black Twitter users and each other by wading through the “toxic technoculture” of Twitter that reinforces harmful narratives about Black communities and upholds white supremacy [Massanari 2017, 329–342]. We created our own liberated space in the physical environment of the BCaT lab, and the virtual space of Zoom allowed us to share our individual experiences using Twitter and offer insights on how the platform changes impacted user experience, as well as work with collaborators who were geographically distant. Having an ethics of care allows digital humanities scholars to prioritize the emotional capacities of each team member to mitigate stress and tensions, find pockets of joy, and practice compassion with each other.

Manual data collection along with the creation of our Musk Acquisition Timeline allowed us to situate individual tweets within a specific context and become more immersed in public conversation. Oftentimes, automated methods, if not paired with qualitative approaches, do not allow researchers to extract information regarding the broader political and social contexts of the tweets or the variety of conversations that a particular post is a part of. Therefore, we provided thick descriptions of how users engaged with Twitter by observing and analyzing detailed accounts of Black adaptive rhetorical practice [Geertz 2008]. This involved cataloging the specific ways Black users employed strategic language, images, GIFs, hashtags, and other digital strategies to interact and traverse the platform while illuminating the historical, cultural, and social context that inform these practices. This deeper understanding was made possible due to our weekly engagement on the app and with the broader digital media detailing platform changes and their implications for active users.

Our team found that a balanced approach combining the efficiency of automated scraping tools with nuances of manual scraping was effective for our chosen research subject. Manual scraping allows us to iteratively refine and target specific discourses, and in tandem with computational scraping tools, we found that we could combine efficiency and small-data collection sensibilities.

By exploring both manual and computational data collection methods, we refined a data collection technique that included the best affordances from both methods. We used Zeeschuimer to automatically gather metadata from tweets, including the content, user location, like count, retweet count, attachments, time stamp, and other information that the tool deems important to researchers studying social media. Since Zeeschuimer reads over the researcher’s shoulder, we had to act as a search engine and therefore navigate to the appropriate material for inclusion. Our manual data collection stage allowed us to assess other pertinent categories of analysis, such as the users’ assumed racial and gender identity, follower count, and descriptions of tweet content based on our thematic coding. However, we had to manually examine each tweet to differentiate between authentic users and the bots/fake accounts that can be included in the data collected by Zeeschuimer. Relying only on the automated dataset would have included some irrelevant tweets, thus skewing the data integral to our data story. To combine both methods, we ‘liked’ the relevant tweets we assessed for inclusion during the manual scraping process, *then* turned to Zeeschuimer to scrape *only* the tweets present in the research account’s ‘likes’ collection, thus quickly generating a dataset that reflected researcher’s pre-selected corpus of user profiles and tweet content, to collect metadata and data samples pertinent to our study.

The integration of both Zeeschuimer and manual scraping alongside our timeline affirms the need for researchers to familiarize themselves with the social media landscape prior to data collection. Simply scraping the terms between the Musk Acquisition Timeline ([BCaT Lab 2023]) start and end dates would not have been sufficient, because our understanding of what was relevant to inclusion was refined through iterative, grounded practice, and our familiarity and

immersion within the field site to identify relevant search terms and spikes in activity.

Our manual scraping alongside the timeline made us more sensitive to the discourse happening on Twitter. We strategically used Zeeschuimer in the end-stages to gather and aggregate metadata. This mixed-method approach offers researchers a model to adapt their own inquiry into counterpublic discursive communities.

38

Following data collection, Tableau allowed us to create a variety of graphs to narrate our research findings. Following the creation of both descriptive and sentiment codes from the dataset, Tableau allowed us to demonstrate the relative frequency with which specific codes appeared across our dataset. It helped us visualize the time periods when Black Twitter users were most active and consider the type of sentiment they expressed online during these heightened times of popularity. One of the advantages of Tableau is its user-friendly interface, which makes the platform easy to use without requiring high-end technical skills.

39

Conclusion

Musk's acquisition of Twitter and subsequent changes to platform governance highlighted challenges digital humanities researchers face amidst changing affordances of social media networks. Ultimately, Musk's Twitter takeover raises questions about the future of data collection on social media networks. Embracing nuanced approaches that leverage the access of automated scraping tools with the sensibilities of manual data collection creates new possibilities for data collection. Specifically, these innovative data collection methods offer solutions to withstand shifting platform affordances that impact the ability to engage in social network research.

40

Methodological adaptability is critical when engaging with platform affordances that can change rapidly. This article sought to dissect the differences between manual scraping and automated data collection and illustrate how they might work in tandem. These challenges ultimately allowed for the cultivation of iterative scraping solutions and the prioritization of group learning as method. The BCaT collaborative team hopes this study empowers researchers to take a flexible and reflexive mixed method approach to studying the discourse of social media subcultures.

41

Notes

[1] The BCaT collaborative team is based in the Communication Department at the University of Maryland and is composed of a collective group of faculty, graduate and undergraduate fellows researching Black communication practices in/and technoculture.

Works Cited

Jessop, M., (2008) "Digital visualization as a scholarly activity", *Literary and linguistic computing* 23, pp. 281–293. <https://doi.org/10.1093/lc/fqn016>

BCaT Lab 2023 BCaT Lab (2023) Musk Acquisition Timeline [Interactive Timeline]. Available at: <https://www.bcatlab.org/black-digital-migration-project/interactive-timeline>.

Banks 2018 Banks, C. (2018) "Disciplining Black activism: post-racial rhetoric, public memory and decorum in news media framing of the Black Lives Matter movement", *Continuum*, 32(6), pp. 709–720. Available at: <https://doi.org/10.1080/10304312.2018.1525920>.

Brock 2012 Brock, A. (2012) "From the blackhand side: Twitter as a cultural conversation", *Journal of Broadcasting & Electronic Media*, 56, pp. 529–549. Available at: <https://doi.org/10.1080/08838151.2012.732147>.

Brock 2020 Brock, A. (2020) *Distributed blackness: African American cybercultures*. New York: NYU Press.

Caliandro 2018 Caliandro, A. (2018) "Digital methods for ethnography: Analytical concepts for ethnographers exploring social media environments", *Journal of Contemporary Ethnography*, 47, pp. 551–578. Available at: <https://doi.org/10.1177/0891241617702960>.

Clark 2020 D. Clark, M. (2020) DRAG THEM: A brief etymology of so-called "cancel culture". *Communication and the Public*, 5(3–4), pp. 88–92. Available at: <https://doi-org.proxy-um.researchport.umd.edu/10.1177/2057047320961562>

Collins 2000 Collins, P.H. (2000) *Black feminist thought: Knowledge, consciousness, and the politics of empowerment*. (2nd edn.) New York, NY: Routledge. Available at: <https://doi.org/10.4324/9780203900055>.

- Everett 2002** Everett, A. (2002) "The revolution will be digitized: Afrocentricity and the digital public sphere", *Social Text*, 20(2), pp. 125–146. Available at: https://doi.org/10.1215/01642472-20-2_71-125
- Florini 2014** Florini, S. (2014) "Tweets, tweeps, and signifyin': Communication and cultural performance on "Black Twitter"", *Television & New Media*, 15, pp. 223–237. Available at: <https://doi.org/10.1177/1527476413480247>
- Florini 2019** Florini, S. (2019) *Beyond hashtags: Racial politics and Black digital networks*. New York, NY: NYU Press Scholarship Online. Available at: <https://doi.org/10.18574/nyu/9781479892464.001.0001>
- Freelon et al. 2016** Freelon, D. and McIlwain, C. D. and Clark, M. (2016) *Beyond the hashtags: #Ferguson, #Blacklivesmatter, and the online struggle for offline justice*. Center for Media & Social Impact, American University. Available at: <http://cmsimpact.org/blmreport>
- Geertz 2008** Geertz, C. (2008) "Thick Description: Toward an Interpretive Theory of Culture 1973", *The Cultural Geography Reader*. New York, NY: Routledge.
- Glaser 1992** Glaser, B. (1992) *Basics of Grounded Theory Analysis*. Mill Valley, CA: Sociology Press.
- Glaser and Strauss 1967** Glaser, B., and Strauss, A. (1967) *The Discovery of Grounded Theory: Strategies for Qualitative Research*. Mill Valley, CA: Sociology Press.
- González-Bailón et al. 2014** González-Bailón, S. et al. (2014) "Assessing the bias in samples of large online networks", *Social Networks*, 38, pp. 16–27. Available at: <https://doi.org/10.1016/j.socnet.2014.01.004>
- Hartman 2008** Hartman, S. (2008) Venus in Two Acts. *Small Axe Number* 26, 12(2), pp. 1-14. Available at: <https://muse.jhu.edu/article/241115>.
- Hartman 2021** Hartman, S. (2021) "Intimate history, radical narrative", *The Journal of African American History*, 106, pp. 127–135. Available at: <https://doi.org/10.1086/712019>
- Helton 2021** Helton, L.E. (2021) "Schomburg's library and the price of Black history", *African American Review*, 54, pp. 109–127. Available at: <https://www.jstor.org/stable/27111721>
- Helton et al. 2015** Helton, L et al. (2015) "The question of recovery: An introduction", *Social Text* 4(125): pp. 1–18. Available at: <https://doi.org/10.1215/01642472-3315766>.
- Jackson et al. 2020** Jackson, S., Bailey, M., Welles, B. (2020) *#HashtagActivism: Networks of race and gender justice*. Massachusetts: Cambridge MA: The MIT Press. Available at: <https://doi.org/10.7551/mitpress/10858.001.0001>
- Johnson 2020** Johnson, J., (2020) "Wicked flesh: Black women, intimacy, and freedom in the Atlantic world", *Early American Studies*. Philadelphia, PA: University of Pennsylvania Press.
- Kumar, Zafarani, and Liu 2011** Kumar, S., Zafarani, R., Liu, H. (2011) "Understanding user migration patterns in social media", *Proceedings of the AAAI Conference on Artificial Intelligence*, 25(1), pp. 1204-1209. Available at: <https://doi.org/10.1609/aaai.v25i1.8089>.
- Leaver 2017** Leaver, T., (2017) "Born digital? Presence, privacy, and intimate surveillance", *Translingual Transcultural Transmedia. Studies in narrative, identity, and knowledge*, pp. 146-160. Shanghai: Fudan University Press. Available at: <https://doi.org/10.31235/osf.io/ay43e>
- Levenberg et al. 2018** Levenberg, L., Neilson, T., Rheams, D. (2018) *Research methods for the digital humanities*. Switzerland: Palgrave Macmillan Cham. Available at: <https://doi.org/10.1007/978-3-319-96713-4>
- Lu and Steele 2019** Lu, J. H. and Steele, C. K. (2019) "'Joy is resistance': cross-platform resilience and (re)invention of Black oral culture online", *Information, Communication & Society*, 22(6), pp. 823–837. Available at: <https://doi-org.proxy-um.researchport.umd.edu/10.1080/1369118X.2019.1575449>
- Martin et al. 2018** Martin, Vi B.; Scott, Clifton; Brennen, Bonnie; and Durham, Meenakshi G. (2018) "What is grounded theory good for?", *College of Communication Faculty Research and Publications*, p. 502.
- Massanari 2017** Massanari, A. (2017) "#Gamergate and the fapping: How Reddit's algorithm, governance, and culture support toxic technocultures", *New Media & Society*, 19(3), pp. 329-346. Available at: <https://doi-org.proxy-um.researchport.umd.edu/10.1177/1461444815608807>
- Musk 2023** Musk, E [@elonmusk]. (2023) <https://x.com/elonmusk/status/1683171310388535296?lang=en> (Accessed 25 November 2025).
- Noble 2018** Noble, S., (2018) *Algorithms of oppression*. New York, NY: NYU Press.

- Ossom-Williamson 2021** Ossom-Williamson, P., Ramsby, K. (2021) "Part 2. The FLOAT Method". Available at: https://uta.pressbooks.pub/datanotebook/part/2_float/
- Peeters 2023** Peeters, S. (2023) "Analyzing online expression affordances on IRC and Twitter", in M. Saunders, & L. Gee (Eds.), *Ego-Media : Life Writing and Online Affordances*. Redwood City, CA: Stanford University Press. Available at: <https://doi.org/10.21627/2023em>
- Peeters 2025** Peeters, S. (2025) "Zeeschuurmer." Zenodo. doi:10.5281/zenodo.15281614.
- Peeters and Hagen 2022** Peeters, S. and Hagen, S. (2022) "The 4CAT capture and analysis toolkit: A modular tool for transparent and traceable social media research", *Computational Communication Research* 4(2): pp. 571-589. Available at: <http://dx.doi.org/10.2139/ssrn.3914892>
- Perera et al. 2010** Perera, R.D, et al. (2010) "Twitter analytics: Architecture, tools and analysis", in *MILCOM 2010 MILITARY COMMUNICATIONS CONFERENCE*. San Jose, CA: IEEE. 2186-2191. Available at: <https://ieeexplore.ieee.org/document/5680493>
- Saldana 2021** Saldana, J. (2021) *The coding manual for qualitative researchers*. Thousand Oaks, CA: SAGE Publications Limited.
- Squires 2002** Squires, C.R. (2002) "Rethinking the Black public sphere: An alternative vocabulary for multiple public spheres", *Communication Theory* 12, pp. 446–468. Available at: <https://doi.org/10.1111/j.1468-2885.2002.tb00278.x>
- Steele 2016** Steele, C.K. (2016) "The digital barbershop: Blogs and online oral culture within the African American community", *Social Media + Society* 2, 2056305116683205. Available at: <https://doi.org/10.1177/2056305116683205>
- Steele 2018** Steele, C.K. (2018) Black Bloggers and Their Varied Publics: The Everyday Politics of Black Discourse Online. *Television & New Media* 19, 112–127. Available at: <https://doi.org/10.1177/1527476417709535>
- Steele et al. 2023** Steele, C.K., Lu, J., Winstead, K. (2023) *Doing Black digital humanities with radical intentionality: A practical guide*. New York: Routledge.
- Sutherland and Collier 2022** Sutherland, T. and Collier, Z. (2022) "Introduction: The promise and possibility of Black archival practice", *The Black Scholar*, 52, pp. 1–6. Available at: <https://doi.org/10.1080/00064246.2022.2043722>
- Walcott 2024** Walcott, R., (2024) "#RIP Twitter: The Conditions of Black Social Media Platform Migration", Just Tech. Social Science Research Council. Available at: <https://doi.org/10.35650/JT.3068.d.2024>
- Wisser et al. 2016** Wisser, K.M. and Blanco-Rivera, J.A. (2016) "Surveillance, documentation and privacy: an international comparative analysis of state intelligence records", *Archival Science* 16, pp. 125–147. Available at: <https://doi.org/10.1007/s10502-015-9240-x>
- X Developers 2023** X Developers (2023) *Introducing paid access to the Twitter API*. Available at: <https://developer.x.com/en/docs/x-api>.

Recommendations

DHQ is testing out three new article recommendation methods! Please explore the links below to find articles that are related in different ways to the one you just read. We are interested in how these methods work for readers—if you would like to share feedback with us, please complete our short evaluation survey. You can also visit our documentation for these recommendation methods to learn more.

SPECTER Recommendations

Below are article recommendations generated by the SPECTER model:

1. My DH Present, Past, & Future, 2022, Moya Bailey, Northwestern University
2. Welcome to Digital Humanities Quarterly, 2007, Julia Flanders, Brown University; Wendell Piez, Mulberry Technologies, Inc.; Melissa Terras, University College London
3. The Fulll Monty, 2022, Elizabeth Losh, William & Mary
4. A Review of Twitter and Tear Gas, 2021, Nanditha Narayanamoorthy, York University
5. Now is the Future Now? The Urgency of Digital Curation in the Digital Humanities, 2013, Alex H. Poole, University of North Carolina at Chapel Hill

DHQ Keyword Recommendations

Below are article recommendations generated by DHQ Keywords:

1. My DH Present, Past, & Future, 2022, Moya Bailey, Northwestern University
2. #transform(ing)DH Writing and Research: An Autoethnography of Digital Humanities and Feminist Ethics, 2015, Moya Bailey, Northeastern University
3. Lab and Slack. Situated Research Practices in Digital Humanities - Introduction to the DHQ Special Issue., 2020, Mila Oiva, University of Turku (Finland); Urszula Pawlicka-Deger, Aalto University
4. #BlackScholarJoy: The Labor, Resistance and Joy Practices of Black Women Graduate Students, 2022, Ravynn K. Stringfield, University of Richmond
5. Studying Up: A Review of Alice Marwick's Status Update, 2015, Luke Fernandez, Weber State University

TF-IDF Recommendations

Below are article recommendations generated by the TF-IDF Model:

1. "Whole Self to the World": Creating Affective Worlds and Black Digital Intimacy in the Fandom of The Misadventures of Awkward Black Girl and Insecure, 2022, Brienne A. Adams, Georgetown University
2. #BlackScholarJoy: The Labor, Resistance and Joy Practices of Black Women Graduate Students, 2022, Ravynn K. Stringfield, University of Richmond
3. Attention Ecology: Trend Circulation and the Virality Threshold, 2016, Nicholas M Van Horn, Capital University; Aaron Beveridge, University of Florida; Sean Morey, University of Tennessee, Knoxville
4. Pertinent Discussions Toward Modeling the Social Edition: Annotated Bibliographies, 2012, Ray Siemens, University of Victoria; Meagan Timney, University of Victoria; Cara Leitch, University of Victoria; Corina Koolen, University of Victoria; Alex Garnett, University of Victoria
5. Debates in #BlackDH: Key Moments and Queer Directions in Black Studies Scholarship, 2024, Faithe J. Day



This work is licensed under a Creative Commons Attribution-NoDerivatives 4.0 International License.