



Selected Papers of #AoIR2023:
The 24th Annual Conference of the
Association of Internet Researchers
Philadelphia, PA, USA / 18-21 Oct 2023

WHO WATCHES THE BIRDWATCHERS? CREATING A ROGUE ARCHIVE OF TWITTER'S ONGOING COLLAPSE

Ben T. Pettis
University of Wisconsin-Madison

Crowdsourced Fact-Checking as Platform Metacommentary

Community Notes (formerly named “Birdwatch”) is Twitter’s crowdsourced fact-checking program to combat mis- and dis-information.¹ It “aims to create a better-informed world, by empowering people on Twitter to collaboratively add helpful notes to Tweets that might be misleading ” (Twitter n.d.). By signing up to be a Birdwatch contributor, a user can add contextual notes and commentary to other tweets as well as rate the contributions of others. Since 2021, community notes that are rated as “helpful” are displayed underneath Tweets that are identified by Birdwatch users as needing additional context. User submissions to the Community Notes program also serve as metacommentary on the platform more generally. Beyond their fact-checking role, Birdwatch notes also illuminate how some users perceived Elon Musk’s recent purchase of the platform and how the subsequent changes aligned with their own understandings of what the platform ought to be.

This paper describes Birdwatch Archive, a project to archive Twitter’s Community Notes program, which has been an under-acknowledged component of the platform’s history. The website processes the data that Twitter publicly releases from the Birdwatch program and displays it in a searchable and organized fashion that is accessible and useful to researchers. As Muira McCammon and Jessa Lingel (2022) explain, studying dying platforms is a way “for designers and practitioners to think about the lifecycles of the devices and services they create, and for users seeking to understand the relationships between people, platforms and data” (2). I argue that studying a dying platform relies upon the creation of what Abigail De Kosnik (2016) describes as “rogue archives,” collections which exist beyond traditional memory institutions. This includes

¹ The online platform now named “X” was previously named “Twitter.” I continue to use this original name in following the practices of the platform’s users and communities. Similarly, although Twitter renamed the “Birdwatch” program to “Community Notes” shortly after the Musk purchase, I continue to use both interchangeably as both names continue to be used within Twitter’s publicly posted documentation as well as within the public URLs for downloading data.

Suggested Citation (APA): Pettis, B. T. (2023, October). *Who Watches the Birdwatchers? Creating a Rogue Archive of Twitter’s Ongoing Collapse*. Paper presented at AoIR2023: The 24th Annual Conference of the Association of Internet Researchers. Philadelphia, PA, USA: AoIR. Retrieved from <http://spir.aoir.org>.

turning to less publicized materials from social media platforms, and to preserve and make these available to other researchers.

Even though major platforms such as Twitter are typically well-documented and may offer some data access, researchers cannot rely solely on the platform itself to document its histories. I use notes from the Birdwatch Archive to analyze how some Twitter users experienced the weeks and months surrounding Musk's purchase of the platform. How did some users understand the platform and its online space? How did some users understand their own role in shaping the platform? What can we take from that to our analyses of other online spaces?

Elon Musk Purchases Twitter and Transforms the Platform

In late 2022, Elon Musk purchased Twitter and enacted many significant changes across the platform, such as firing thousands of employees, reinstating previously banned users, and selling the blue checkmark “verified status” for a monthly fee (Scribner 2022). In the months following Musk's purchase, many users abandoned the platform and there have been numerous failures of technology and infrastructure. In their “platform biography,” Jean Burgess and Nancy Baym (2020) argue that much of Twitter's success can be traced to novel practices developed by users and later co-opted by the platform, such as hashtags and retweeting. The sociocultural meaning of any online platform emerges, in part, from the balance between user practices and the business interests of its corporate owners (Gillespie 2010). As with the decline of earlier platforms like Friendster, MySpace, and Tumblr, the sudden changes enacted on Twitter felt like a seismic shift in the makeup of the contemporary internet (McCammon and Lingel 2022; Corry 2022; Ogden 2021). But, as those previous cases demonstrate, new platforms will emerge may soon take Twitter's place. Speculating in 2020 on the future of Twitter and similar platforms, Burgess and Baym (2020) wrote that “[t]he challenge for us all is how to maintain the creative power users have to reshape the cultures of digital media platforms, especially when they seem to be veering precipitously toward interests that may work against their users' security, privacy, well-being, and civic agency” (115). Even as Twitter continues to decline and the balance tips toward those interests, internet scholars can assess what's left behind on the dying platform and imagine how future platforms might better balance the interests of their users with their own business priorities.

The Birdwatch Archive

There has been significant development of methods to study Twitter by retrieving and analyzing Tweets, and these will continue to be critical for studying the platform's ongoing decline. However, the use of meta-commentary such as data from the Community Notes program offers additional opportunities to understand how users perceived their role in shaping the platform. In the interest of transparency, Twitter publicly distributes Birdwatch data, but this is limited in its scope and utility (Twitter n.d.). Data is batched and updated approximately once per day as a set of TSV (tab-separated values) files. Typically, only the past week's worth of data continues to be accessible at their public URLs, and only one day's files listed on the data download page. Furthermore, although the data is distributed in a well-documented and structured

format, it is difficult for a researcher to parse and search through the data in an efficient manner. To address these limitations, I have developed an archiving project to preserve copies of the public Birdwatch data and make it accessible in a more convenient format for researchers. The project uses a series of Python scripts to copy TSV files into a public Google Cloud Storage Bucket. The files are then imported into a relational database which is used as the backend for a public-facing website: www.birdwatcharchive.org. Following Shawn Graham, Ian Milligan, and Scott Weingart's (2016) "macroscopic" methods for large data sets, the website facilitates sorting, querying, identifying relations across the data at scale. Using the anonymous user identification strings from each TSV file, the website enables researchers to assess how frequently users contribute to the Community Notes program by grouping notes and ratings they have provided.

Even as Twitter continues to devolve and collapse, we can try to learn from how users described and understood the platform. When studying major platforms, we cannot rely solely upon the data made accessible by the platform itself. Instead, we must look for opportunities to create "rogue archives" of online settings, which includes turning sources that are not as frequently viewed by most users. As Jessica Ogden (2021) puts it, "web archiving practices constrain how we will come to know and understand dead and dying platforms in future" (3). Therefore, we should try to cast as wide a net as possible when archiving how various users understood an online space. Preserving significant tweets is important, but other aspects of the platform—including Birdwatch—are important for future analyses of Twitter as well. Part of imagining and building the future of the internet should include sifting through the rubble of what came before. Researchers can lean upon rogue archives such as the Birdwatch Archive to begin to do this work.

References

- Brock, André. 2012. "From the Blackhand Side: Twitter as a Cultural Conversation." *Journal of Broadcasting & Electronic Media* 56 (4): 529–49. <https://doi.org/10.1080/08838151.2012.732147>.
- Bruns, Axel, Tim Highfield, and Jean Burgess. 2014. "The Arab Spring and Its Social Media Audiences: English and Arabic Twitter Users and Their Networks." In *Cyberactivism on the Participatory Web*. Routledge.
- Burgess, Jean, and Nancy K. Baym. 2020. *Twitter: A Biography*. New York: New York University Press.
- Corry, Frances. 2022. "Why Does a Platform Die? Diagnosing Platform Death at Friendster's End." *Internet Histories* 6 (1–2): 31–47. <https://doi.org/10.1080/24701475.2021.1985360>.
- Crosset, Valentine, Samuel Tanner, and Aurélie Campana. 2018. "Researching Far Right Groups on Twitter: Methodological Challenges 2.0." *New Media & Society* 21 (4): 939–61. <https://doi.org/10.1177/1461444818817306>.

- De Kosnik, Abigail. 2016. *Rogue Archives: Digital Cultural Memory and Media Fandom*. Cambridge, Massachusetts: The MIT Press.
- Fuchs, Christian. 2018. *Digital Demagogue: Authoritarian Capitalism in the Age of Trump and Twitter*. London: PlutoPress.
- Gillespie, Tarleton. 2010. "The Politics of 'Platforms.'" *New Media & Society* 12 (3): 347–64. <https://doi.org/10.1177/1461444809342738>.
- Graham, Shawn, Ian Milligan, and Scott Weingart. 2016. "The Joys of Big Data for Historians." In *Exploring Big Historical Data: The Historian's Macroscope*. London: Imperial College Press.
- McCammon, Muira, and Jessa Lingel. 2022. "Situating Dead-and-Dying Platforms: Technological Failure, Infrastructural Precarity, and Digital Decline." *Internet Histories* 6 (1–2): 1–13. <https://doi.org/10.1080/24701475.2022.2071395>.
- Ogden, Jessica. 2021. "'Everything on the Internet Can Be Saved': Archive Team, Tumblr and the Cultural Significance of Web Archiving." *Internet Histories* 0 (0): 1–20. <https://doi.org/10.1080/24701475.2021.1985835>.
- Scribner, Herb. 2022. "A Timeline of the Musk-Twitter Deal so Far." *Axios*. October 26, 2022. <https://www.axios.com/2022/05/17/elon-musk-twitter-timeline>.
- Twitter. n.d. "Community Notes Guide." Accessed November 15, 2022a. <https://twitter.github.io/communitynotes/overview/>.
- . n.d. "Downloading Data | Community Notes Guide." Accessed November 15, 2022b. <https://twitter.github.io/communitynotes/download-data/>.