



**Selected Papers of #AoIR2022:  
The 23rd Annual Conference of the  
Association of Internet Researchers**  
Dublin, Ireland / 2-5 Nov 2022

## **ARCHIVING THE INSURRECTION: THE CASE OF R/DATAHOARDER**

Edward Chapman  
Marquette University

### **Introduction**

On January 6th, 2021, right-wing demonstrators forcibly entered the United States Capitol building in an effort to disrupt the certification of the 2020 presidential election. Rioters documented their efforts on social media and live-streaming platforms, creating a digital footprint of their actions. In response, members of the r/DataHoarder subreddit initiated an ad-hoc archival project to collect, deduplicate, and preserve incriminating evidence from the Capitol insurrection [5]. These Redditor's efforts were not alone: other groups and individuals sought to collect and analyze footage of the insurrection, some going as far as to apply facial recognition technology or crowdsourcing in order to identify suspected participants [2, 6]. The confluence of self-incriminating social media content, collaborative online platforms, and open-source investigatory methods created an environment where participatory activism mixed freely with punitive vigilantism and historical preservation. This research examines the nature of the r/DataHoarder archival work as an example of a participatory, crowdsourced online investigation with far-ranging practical and ethical implications.

### **Background**

The 2013 Boston Marathon bombing, and its associated subreddit, r/FindBostonBombers, provides a historical comparison to the r/DataHoarder insurrection archive. The event set precedent as the first time a crowdsourced investigation ran parallel to a police investigation for a US terror attack [4]. Reddit's unsuccessful search for the Boston Marathon bombers was notable for the analytic capabilities they mustered from crowdsourced volunteers. The investigation also served as a warning due to the trail of misinformation and false accusations left in its wake. The memory of r/FindBostonBombers was often invoked throughout the r/DataHoarder insurrection archive project as a reminder of past ethical oversteps.

Unbeknownst to many of the Capitol rioters, the incriminating and personally identifiable nature of their social media postings left them vulnerable to personal and professional

Suggested Citation (APA): Chapman, E. (2022, November). *Archiving the Insurrection: The Case of r/DataHoarder*. Paper presented at AoIR 2022: The 23rd Annual Conference of the Association of Internet Researchers. Dublin, Ireland: AoIR. Retrieved from <http://spir.aoir.org>.

humiliation. Doxing describes the public dissemination of an individual's personal information as a means to deanonymize, intimidate, or humiliate them [1]. The term originated in 1990's hacker culture, where "dropping dox" on a rival could reveal their offline identity and exile them from the community [3]. The risk of doxing is ever present in crowdsourced investigations such as the r/DataHoarder insurrection archive, where information collection can be used to enact a witch hunt. I am interested in the discourse surrounding doxing within the r/DataHoarder community, knowing that the insurrection archive could provide a rich source of material for such endeavors.

## Methods

A Python script was used to download comment text and metadata for all comments in the Capitol insurrection archive thread within r/DataHoarder. This formed a dataset of 2,788 comments posted by 1,311 unique users. I imported comment text into a qualitative coding software and coded each comment for users' stated motivations as well as any archival activities that I could observe. My code list developed throughout the coding process and focused on different forms of participation in the archive. See Appendix A for the full list of codes. As I worked through the dataset, I added codes for salient topics in my research such as *doxing* for encouraging the exposure of rioters' personally identifiable information and *fbi* for calls to submit evidence to law enforcement agencies. I also removed or merged codes that were needlessly specific, such as distinctions between requests for technical help and responses to these requests, both of which were merged into *strategy*. I did not code every comment. I skipped comments which were deleted, comments directed at or posted by bots, very brief comments, and comments whose contents were contextually irrelevant to the aims of the thread.

## Findings

Throughout the r/DataHoarder insurrection archive project, Redditors expressed a range of motivations for their efforts. I found that stated motivations generally fell into one of the following categories: *contribute to law enforcement efforts*; *enact punitive justice upon rioters*; *engage in public discourse*; and *preserve information for posterity*. Redditors expressed multiple and, at times, contradictory motives, yet *preserving information for posterity* was the most frequent reason observed. One of the main aims in coding the dataset was to identify which activities the Redditors engaged in during the insurrection archive project. Such activities included emotional arguments, political analysis, calls for doxing and punishment, calls for evidence submission to the FBI, as well as activities traditionally associated with digital archiving such as locating sources, deduplicating material, and preserving material. The most common activity was submitting evidence for archiving in the form of social media URLs. Other top activities included coordinating efforts (deduplicating and vetting submissions, providing technical advice) and scraping or downloading evidence from URL submissions. Non-archival activities such as doxing or arguing were the 5th and 6th most frequent activities, respectively. Grouping participants according to the number of times they posted in the insurrection archive revealed that the most heavily invested participants focused their efforts on technical archival tasks. Calls for doxing were most commonly posted by first-time participants who did not return to the project.

## Significance

Findings indicate that the r/DataHoarder insurrection archive project was ultimately successful in their attempts to collect and preserve evidence of the Capitol insurrection. Many of the archive's images and videos were under active threat of removal by platforms and users. The archive contents have since been used in criminal investigations into the rioters. While similar projects pursued punitive ends, the r/DataHoarder archive project maintained focus on the archival aspects of their task. The archive project effectively brought together the technical expertise of long-time "data hoarders" with the energy and motivations of participants from across the Reddit platform. Participants demonstrated an awareness of the historical value not just of the insurrection archive project, but of archiving as a whole. Through language such as "highest calling for hoarders like myself" and "we must continue to archive," these Redditors express a belief in the inherent value of archival work. This project serves as an example of the efficacy of crowdsourced, participatory archiving work during a crisis event.

## References

- [1] David M. Douglas. 2016. Doxing: a conceptual analysis. *Ethics Inf Technol* 18, 3 (September 2016), 199–210. DOI:<https://doi.org/10.1007/s10676-016-9406-0>.
- [2] Home Grown Terrorists. 2021. homegrownterrorists. *Instagram*. Retrieved August 11, 2021 from <https://www.instagram.com/homegrownterrorists/>.
- [3] Mat Honan. 2014. What Is Doxing? *Wired*. Retrieved September 29, 2021 from <https://www.wired.com/2014/03/doxing/>.
- [4] Johnny Nhan, Laura Huey, and Ryan Broll. 2017. Digilantism: An Analysis of Crowdsourcing and the Boston Marathon Bombings. *The British Journal of Criminology* 57, 2 (March 2017), 341–361. <https://doi.org/10.1093/bjc/azv118>
- [5] u/AdamLynch. 2021. r/DataHoarder - MEGATHREAD: Archiving the Capitol Hill Riots. *Reddit*. Retrieved February 17, 2021 from [https://www.reddit.com/r/DataHoarder/comments/krx449/megathread\\_archiving\\_the\\_capitol\\_hill\\_riots/](https://www.reddit.com/r/DataHoarder/comments/krx449/megathread_archiving_the_capitol_hill_riots/)
- [6] 2021. Faces of the Riot. *Faces of the Riot*. Retrieved February 17, 2021 from <https://facesoftheriot.com/>.

## Appendix A: Code List

Code	Description
<i>evidence</i>	URL links to insurrection evidence
<i>strategy</i>	Coordinating archival efforts; technical support
<i>saving</i>	Downloading <i>evidence</i> content; uploading to archive
<i>analysis</i>	Analyzing insurrection media
<i>doxing</i>	Proposals to target rioters via release of personal information
<i>arguing</i>	Political arguments
<i>seeding</i>	Contributing internet bandwidth to help distribute the archive
<i>encouragement</i>	Words of encouragement
<i>motivation</i>	Personal motive for participating in the archive
<i>fbi</i>	Proposals to share archive with FBI
<i>removal</i>	Flagging <i>evidence</i> as removed by user or platform
<i>caution</i>	Expressing ethical hesitation regarding archive project