



Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.



When is public private? Tweets, privacy and consent in health research

ARTICLE INFO

Keywords

Twitter
Privacy
Health research
Consent

1. Introduction

Online social media platforms provide opportunities for the global community to share and express their views, opinions, reactions, and feelings openly. The use of social media for the purpose of information sharing surged during the COVID-19 pandemic due to mandated physical distancing requirements. This is a seemingly consensual catharsis at a time of heightened need for alternative social activity and critical information sharing.

Communication on open social media platforms has created opportunities for researchers to access and analyse rich, publicly available data to study a range of topics and issues. The creation of this abundant public data has also led to fundamental methodological and ethical challenges for social science researchers. Namely, is the use of this public data for research a breach of privacy and confidentiality? Are social media users becoming involuntary research participants as their communications and personal information are mined and published on without participant insight and informed consent? In this discussion, we aim to highlight some of the critical methodological and ethical issues that researchers must consider while using Twitter as a data source to publish from.

2. Twitter

One such social media platform is Twitter, a micro-blogging site for online news and social networking. Launched in 2006 and ranked as one of the most used social media platforms, Twitter is widely used to follow trending current issues, share personal views, and significant information regarding political debate, sensitive social issues, natural disasters, pandemics and other crisis events. This communication is at times conducted to seek attention and responses of government and other stakeholders. Twitter data are publicly accessible and can contain insightful demographics, such as the location of the user. Such metadata are valuable information in the exploration of the relationship between the geo-socio-political context and the specific phenomena of research

interest.

However, the use of Twitter posts for research needs further consideration to protect participants from additional risk of harm relating to their identity and information disclosure (Webb et al., 2017). Does this risk justify the benefits gained by publishing this analysed data, in conjunction with the professional benefits to the researchers themselves by publishing? Is it an acceptable level of risk? Williams et al. (2017) state that there is a need to examine the use of these posts via a social science lens that utilises a reflexive ethical view as opposed to a legal stance that allows the use of data within research publications.

The use of social media data for research purposes and associated methodological issues has been discussed in the research literature (Ahmed et al., 2017), but attention is yet to be focused on the incorporation of practical strategies to address critical ethical issues and reduce risk. There is no agreed approach for how researchers can responsibly analyse Twitter data and publish results to protect participants. This creates further complexities for researchers to act ethically to disseminate the results of the analysis of public tweets.

3. Why do researchers use Twitter data?

The popularity of using Twitter data for research in different disciplines such as social sciences, communications, political sciences, psychology, and health sciences, has increased over the years. There are various reasons why researchers like to analyse Twitter data. The most salient reason likely being access to metadata including the tweets, numbers of followers, favourites, language, and geographic location give richness to the data (Ruiz Soler, 2017).

Data extraction is enabled in several ways. First, the Twitter Application Programming Interface (API) is open and accessible, enabling the easy extraction of large amounts of data. Researchers can search specific conversations using hashtags which enables following topic-based conversations by location. Second, the hashtag culture of Twitter makes data collection easier to explore stories and incidents of interest for analysis. Third, Twitter is widely used for media purposes which can

<https://doi.org/10.1016/j.nepr.2022.103396>

Received 14 June 2022; Accepted 27 June 2022

Available online 30 June 2022

1471-5953/© 2022 Elsevier Ltd. All rights reserved.

attract attention to secure research opportunities, funding or to raise the profile of the researchers who themselves are active on Twitter.

As Twitter data are publicly available, researchers are not always required to apply for ethical approval to analyse and publish from the data. Researchers are therefore able to avoid this often-time-consuming process which further incentivises the use of social media data. Researchers can investigate multiple issues in much less time, often with equally insightful data. Further, with no direct contact with participants, the complexities of preparing for fieldwork are removed. The associated time and cost are greatly reduced, and as such is a cost-effective approach.

Twitter has also recently released a new 'academic research product track' allowing researchers to pull in 10 million tweets per month for free with access to a full archive search (Ahmed, 2021). This new access is advertised as an opportunity to 'get more precise, complete, and unbiased data from the public conversation for free' (Twitter API, 2022). Researchers are encouraged and enabled to analyse Twitter data; a valuable opportunity for those with a low budget and limited time to access other historical data or conduct fieldwork research. With the company encouraging and enabling the analysis of Twitter data for research purposes, their motivation for doing so and the complexities associated with the use of publicly available Twitter data need closer attention.

4. What are the complexities of using Twitter data?

Consumers of Twitter differ in how cognizant they are of the privacy options available to them, and many do not read or comprehend terms of service or conditions that govern consent and data privacy (Williams et al., 2017). Research conducted by the Consumer Policy Research Centre in 2020 surveyed 1000 Australians, finding that 88% of participants reported not having a sound comprehension of how their personal data is gathered and shared by companies. Of participants, 94% reported being uncomfortable with how their personal data were used online and expected the government to provide protection against the collection and sharing of their personal information; 94% of participants reported not reading all the privacy policies or Terms and Conditions applicable to them in the last year. One third of participants reported never having read a privacy policy. Of those who had read privacy policies, 69% reported accepting terms despite not being comfortable in doing so. The primary reason for doing so was to have access to the product or service (75%) (Consumer Policy Research Centre, 2020). Reading privacy policies is disincentivised due to the length and complexity of the documents. McDonald and Cranor (2008) estimate that it would take 244 h per year (on average 40 min per day) for a consumer to read all the privacy policies that apply to them. This is because privacy policies are unreasonably long and contain thousands of words of difficult to comprehend cryptic jargon (Roderick, 2020). As our lives move increasingly online, it could be assumed that the time investment is likely to have also increased since this research was published in 2008.

The ethical implications of these findings regarding the use of Twitter data are reasonably clear. Many Twitter users are likely not providing informed consent for their data to be used for research purposes. Many consumers do not read, or do not understand privacy policies, and are also coerced into accepting the privacy agreement due to wanting to access the platform. As such, consumers are not protected from risk. Gold (2020, p. 3) provides guidance for researchers and ethics reviewers outlining a range of factors that need to be considered when using publicly available data via Twitter: primarily Tweets and public account information. Gold (2020) highlights "that the Twitter 'dataset' is dynamic... This means that the ethical arguments below around consent and privacy must be applied at every use and regularly during retention... [with] implications for research design since data retrieval and ongoing management must account for this, and particular care taken around consent for publication" (p.5). Gold (2020) further suggests viewing 'public' Twitter data as private data which is displayed

publicly based on continued consent under contract, instead of public data as a consequence of publication. Certainly, not all Twitter users are aware of the public visibility and access to their posts by username and location, as well as other demographic information. Users can opt to not post tweets publicly, however the Twitter API provides summary statistics of all accounts.

There was a peak in using Twitter to share COVID-19 related information and evidence during the recent pandemic which had associated benefits and risks. For example, Twitter played a vital role in the rapid research response with over 80% of publications included in the World Health Organisation (WHO) COVID-19 database cited on Twitter (Patel et al., 2021). In addition, Twitter is used to keep the public updated on natural disasters. For example, Willson et al. (2021) used Twitter posts during the Australian bushfires disaster 2019–2020 to analyse data to facilitate the future of Australia's tourism industry.

The practicality of conducting research without seeking informed consent to access personal data from Twitter has been criticised by some social researchers demanding a more reflexive ethical approach (Williams et al., 2017). The arguments around 'public data' and the 'informed consent from users' are diverse to manage any potential harms that may arise from accessing and reporting Twitter data (Williams et al., 2017, p. 1151). Due to the open and public nature of data, researchers often do not inform users that their interactions are being scrutinised for research and reverse identification issues have been noted.

5. Ethical issues of Twitter based research

Twitter has the potential to optimise the research uptake in the current digital age. Twitter provides an opportunity to reach the global participants for greater representations within a short period of time which makes recruitment easier (Wasilewski et al., 2019). Researchers have options to do targeted tweets of 280 characters (Boot et al., 2019) with the link of research information by mentioning specific users to enhance the reach of research activities. The human research ethics processes involve informed consent with participants who are aware of their involvement in the research; this does not always occur on Twitter (Fiesler and Proferes, 2018).

Research that does not comply with certain set principles and recommended guidelines is considered unethical. In the United States, for example, the National Institutes of Health provide seven main principles to guide the conduct of ethical research: social and clinical value, scientific validity, fair subject selection, favourable risk benefit ratio, independent review, informed consent and respect for potential and enrolled subjects (National Institutes of Health, 2016). Recruitment of participants via Twitter does not necessarily abide with the principle of informed consent for research engagement.

Similarly, the National Health, Medical and Research Council (NHMRC) in Australia has set standards for ethical practice for consent to participate in research clearly stating that: "consent should be a voluntary choice and should be based on sufficient information and adequate understanding of both the proposed research and the implications of participation in it" (The National Health and Medical Research Council, 2018, p. 16). The research conducted in Twitter bypasses the 'consent requirement' and participants do not always know about their involvement and the use of their tweets for research purpose. Avoiding the consent process risks potential risks and harms to the participants. The All European Academies (ALLEA) *European Code of Conduct for Research Integrity* highlights the importance of researchers mastering: .

"Failing to follow good research practices violates professional responsibilities. It damages the research processes, degrades relationships among researchers, undermines trust in and the credibility of research, wastes resources and may expose research subjects, users, society or the environment to unnecessary harm." (All European Academies, 2017, p. 8)

Data privacy is another debatable issue. A study among 368 Twitter users revealed that almost two-thirds (61.2%) of participants were not aware of the use of their tweets for research purposes (Fiesler and Proferes, 2018). In the same study, two-thirds of participants (64.9%) agreed that researchers should not be able to use tweets without the user's permission. This indicates that users feel strongly that informed consent is required for their tweets to be used for research. There are also sentiments that if permission was requested then users would permit use of their tweets for research purpose which will add to the credibility of research findings (Fiesler and Proferes, 2018) and ensure that there are no potential risks or harms to the participants that researchers are not aware of to consider while conducting study.

6. Strategies to protect harms and risks

Twitter provides an effective platform for recruitment of participants if used appropriately. Twitter feeds can be set to be publicly viewable, or to be only viewable by specific followers as members of the desired target participant population. A careful selection of words to draft tweets can result in effective reach with minimal risks to the users if they are allowed to make decisions on whether to participate in the research or not. Some strategies to manage recruitment on Twitter have been identified (Arigo et al., 2018) including identifying high profile users who act as potential influencers of the desired participant population and using relevant hashtags.

Twitter provides engagement opportunities for users to share their perspectives on matters that are often potentially of research interest. Nevertheless, it is hard to control the comments that could be offensive or disrespectful to the other users engaged in the discussion which researchers are not able to control in current features of Twitter. A careful selection of users with permission to participate in discussion would help to minimise the emotional harm and social risks to the participants. These complexities could be resolved by reviewing the ethical guidelines in terms of privacy, confidentiality, and ownership of information in social media research and whether the participants are sufficiently protected from the potential harms (Taylor and Pagliari, 2018). Researchers can provide participants with the opportunity to opt out from the research when they feel not safe in the online environment.

Arigo et al. (2018) provide the following solutions to mitigate the potential risks and harms to the users who are being scrutinized for research: consider differential risks to users; customise communication to users about risks and expectations; and collaborate with regulatory bodies to establish standards and update them as required. The authors note that the features of privacy policies are modified often, frequently without public announcement. Researchers who intend to use Twitter data should be familiar with the terms of service and closely monitor if there are any changes to the privacy policies. Despite the popularity and the benefits of using Twitter for research, concerns around the safety of users remain critical in the current digital environment. While there is no easy answer regarding how the ethical challenges can be addressed, a cross-collaboration between the social media regulatory bodies, research communities and the government to develop appropriate research guidelines would be useful.

Finally, while there is undoubtedly value in using Twitter data and metadata for research purposes it should be noted that, while Twitter claims to be a free speech platform and not a publisher, it has exercised censorship more akin to that of a publisher raising the potential for bias. Therefore, the possibility of publication bias on the platform should be considered.

To conclude, the use of Twitter in research is increasing. While Twitter provides quick and easy access to large and diverse sets of data, there are ethical issues that need to be considered and addressed by researchers. Guidelines for researchers to support data privacy and mitigate associated risks are also required.

Editorial note

Editorials in NEP are not reviewed and are published at the discretion of the Editor-in-Chief. We welcome a constructive rejoinder on this editorial provided it is not offensive or personal.

Conflicts of interest

No conflict of interest has been declared by the authors.

References

- Ahmed, W., 2021. Using Twitter as a data source an overview of social media research tools (2021). *Impact Soc. Sci.* (<https://blogs.lse.ac.uk/impactofsocialsciences/2021/05/18/using-twitter-as-a-data-source-an-overview-of-social-media-research-tools-2021/>).
- Ahmed, W., Bath, P.A., Demartini, G., 2017. Using Twitter as a data source: an overview of ethical, legal, and methodological challenges. In: Woodfield, K. (Ed.), *The Ethics of Online Research (Advances in Research Ethics and Integrity)*, 2. Emerald Publishing Limited, Bingley, pp. 79–107. <https://doi.org/10.1108/S2398-60182018000002004>.
- All European Academies, 2017. *The European Code of Conduct for Research Integrity Revised Edition*. Berlin. Retrieved from (<https://www.alllea.org/wp-content/uploads/2017/05/ALLEA-European-Code-of-Conduct-for-Research-Integrity-2017.pdf>).
- Arigo, D., Pagoto, S., Carter-Harris, L., Lillie, S.E., Nebeker, C., 2018. Using social media for health research: methodological and ethical considerations for recruitment and intervention delivery. *Digit Health* 4, 1–15. <https://doi.org/10.1177/2055207618771757>.
- Boot, A.B., Tjong Kim Sang, E., Dijkstra, K., Zwaan, R.A., 2019. How character limit affects language usage in tweets. *Palgrave Commun.* 5 (1), 76. <https://doi.org/10.1057/s41599-019-0280-3>.
- Consumer Policy Research Centre, 2020. *CPRC 2020 Data and technology consumer survey*. Retrieved from (<https://cprc.org.au/cprc-2020-data-and-technology-consumer-survey/>).
- Fiesler, C., Proferes, N., 2018. "Participant" perceptions of Twitter research ethics. *Soc. Media Soc.* 4 (1), 1–14. <https://doi.org/10.1177/2056305118763366>.
- Gold, N., 2020. Using twitter data in research: Guidance for researchers and ethics reviewers. University College London, UK. <https://www.ucl.ac.uk/data-protection/sites/data-protection/files/using-twitter-research-v1.0.pdf>.
- McDonald, A.M., Cranor, L.F., 2008. The cost of reading privacy policies. *ISJLP* 4, 543–568.
- National Institutes of Health, 2016. *Guiding principles for ethical research pursuing potential research participants protections*. Maryland National Institutes of Health Retrieved from (<https://www.nih.gov/health-information/nih-clinical-research-trials-you/guiding-principles-ethical-research>).
- Patel, V.M., Haunschild, R., Bornmann, L., Garas, G., 2021. A call for governments to pause Twitter censorship: using Twitter data as social-spatial sensors of COVID-19/SARS-CoV-2 research diffusion. *Scientometrics* 1–15. <https://doi.org/10.1007/s11192-020-03843-5>.
- Roderick, M., 2020. The average social media privacy policy takes an hour to read. Retrieved from (<https://therealtime.com/2020/09/02/the-average-social-media-privacy-policy-takes-an-hour-to-read/>).
- Ruiz Soler, J., 2017. Twitter research for social scientists: a brief introduction to the benefits, limitations and tools for analysing Twitter data. *Digitos. Rev. De Comun. Digit.* 3, 17–31.
- Taylor, J., Pagliari, C., 2018. Mining social media data: how are research sponsors and researchers addressing the ethical challenges? *Res. Ethics* 14 (2), 1–39. <https://doi.org/10.1177/1747016117738559>.
- The National Health and Medical Research Council, 2018. *National Statement on Ethical Conduct in Human Research 2007 (Updated 2018)*. The National Health and Medical Research Council, the Australian Research Council and Universities Australia, Canberra. (<https://www.nhmrc.gov.au/about-us/publications/national-statement-ethical-conduct-human-research-2007-updated-2018#block-views-block-file-attachments-content-block-1>).
- Twitter API, 2022. *Academic Research access: Advance your research objectives with public data on nearly any topic*. Retrieved from (<https://developer.twitter.com/en/products/twitter-api/academic-research>).
- Wasilewski, M.B., Stinson, J.N., Webster, F., Cameron, J.I., 2019. Using Twitter to recruit participants for health research: an example from a caregiving study. *Health Inform. J.* 25 (4), 1485–1497. <https://doi.org/10.1177/1460458218775158>.
- Webb, H., Jirotko, M., Stahl, B.C., Housley, W., Edwards, A., Williams, M., Procter, R., Rana, O., & Burnap, P., 2017. The ethical challenges of publishing Twitter data for research dissemination. Paper presented at the Proceedings of the 2017 ACM on Web Science Conference. June 25–28, 2017, Troy, NY, USA. (<https://doi.org/10.1145/3091478.3091489>).
- Williams, M.L., Burnap, P., Sloan, L., 2017. Towards an ethical framework for publishing Twitter data in social research: taking into account users' views, online context and algorithmic estimation. *Sociology* 51 (6), 1149–1168. <https://doi.org/10.1177/0038038517708140>.
- Willson, G., Wilk, V., Sibson, R., Morgan, A., 2021. Twitter content analysis of the Australian bushfires disaster 2019-2020: futures implications. *J. Tour. Futures* 7 (3), 350–355. <https://doi.org/10.1108/JTF-10-2020-0183>.

Sabitra Kaphle^{a,1}, Rachel Kornhaber^{b,2}, Susan Hunt^c, Roger Watson^{d,3},
Michelle Cleary^{e,*}

^e School of Nursing, Midwifery & Social Sciences, Central Queensland
University, Sydney, NSW, Australia

^a School of Health, Medical and Applied Sciences, Central Queensland
University, Melbourne, VIC, Australia

^b School of Nursing, University of Tasmania, Sydney, NSW, Australia

^c School of Nursing, Midwifery and Social Sciences, Central Queensland
University, Melbourne, VIC, Australia

^d Southwest Medical University, Luzhou, Sichuan Province, China

* Corresponding author.

E-mail addresses: s.kaphle@cqu.edu.au (S. Kaphle), s.hunt@cqu.edu.au
(S. Hunt), rwatson1955@gmail.com (R. Watson), m.cleary@cqu.edu.au
(M. Cleary).

¹ <https://orcid.org/0000-0003-2350-6568>

² <https://orcid.org/0000-0001-6556-6775>

³ <https://orcid.org/0000-0001-8040-7625>

⁴ <http://orcid.org/0000-0002-1453-4850>