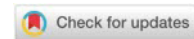


ETHICAL CONSIDERATIONS IN USAGE OF TWITTER DATA

Ivan Blazhevski^{1*}

¹Institute for Sociological, Political and Juridical Research, Ss. Cyril and Methodius University Skopje,
Republic of North Macedonia, e-mail: ivan.blazevski@isppi.ukim.edu.mk



Abstract: Internet-based data, especially the ability to extract and analyze data from social media such as Twitter, is becoming more prevalent in providing data in empirical research. The public nature of Twitter and its more open access to data, compared to other social media, as well as the approval to use that data in accordance with the Twitter's Terms of Service were considered a sufficiently ethical justification for the use of Twitter data in research. This tendency to collect data from the Internet, that is from social networks, in social research, has prompted a number of scientific societies to develop ethical guidelines for Internet research.

This paper provides an overview of the recommendations contained in the ethical guidelines, compared to the requirements set out in the Twitter's Terms of Service. Additionally, research on social media users' perceptions regarding the use and publication of their data is analyzed. The tendency to apply the principle of situational ethics is evident in the ethical guidelines, starting from the existing collision between the established ethical principles for scientific research and the privacy policy on Twitter. However, there is a lack of consensus among ethical guidelines in determining the appropriate way of collecting, processing and presenting data in research and scientific publications. Also, in research on the perception of social media users, there is a significantly lower concern about the use of data by Twitter users, compared to users of other platforms. But in spite of that, the prevailing opinion among them is the need for prior consent for the use of their data in research and scientific publications, with a special emphasis on the request for anonymity. Given the complexity of this debate, which ultimately aims to preserve the academic integrity of research, the conclusion points to the need to summarize the various ethical aspects in establishing a methodological approach in studies that use Twitter data.

Key words: ethics, research, data, social media

INTRODUCTION

Modern technological advances, which enabled wide availability of information, as well as the possibility to conduct certain research in a shorter time and with less financial resources, have led to the widespread use of the Internet. Given that the traditional survey methods have long faced difficulties from increasing costs and declining responses rates, the great interest in web surveys was quite expected. For the first time, an alternative was offered, which made data collection much easier, especially the collection of more sensitive data (Woodfield and Iphofen, 2018). But with increasing public interest and access to the Internet during the early 1990s, ethical questions about practices and outcomes also arose. Thereby, the public was rightfully concerned about the practices such as the widespread collection, archiving, and selling of one's personal information; the loss of privacy from surveillance technologies; the rise in cut and paste plagiarism in education; questions of authorship and credibility in the publication of information and layperson journalism; and copyright issues raised by the easy sharing of music over the Internet (Markham, 2006).

After the first social media appeared on the Internet, 25 years ago, the interest of scientific researchers was immediately attracted by the unprecedented opportunities for research endeavors. Their interest was particularly driven by the opportunities for recruiting participants and collecting data. At that stage of social media development, the activity of the participants in the communication through social media was significantly lower, while the research role was not clearly defined. The first social media surveys took place during this period, dominated by online interviews, participant observations, online surveys and focus groups. In doing so, the first ethical issues were raised, whether social media data should be considered human subjects research or text, how these deliberations influence practices of consent, and whether social media users have perceived expectations of privacy when using social media platforms (Samuel and Buchanan, 2020).

Twitter as one of the most widespread social media platform across the world, has gained popularity among researchers due to its open access. Additionally, researchers can relatively easily collect

*Corresponding author: ivan.blazevski@isppi.ukim.edu.mk



large amounts of data in a short period of time as well as process it through the platform's Application Programming Interface (API). When collecting data, in many cases, the researcher is not required to their oversight or informed consent practices. In doing this, most researchers who collect tweets do not gain consent from each user whose tweet is collected, nor do researchers inform those users. However, it should be noted that Twitter's Privacy Policy mentions that academics may use tweets as part of research, but this update was not included until revisions were made to the policy in 2014. Moreover, Internet users rarely read or could fully understand website terms and conditions (Webb et al., 2017).

RESEARCH ETHICS IN SOCIAL AND e-RESEARCH

Discussions on ethical issues are often perceived as pre-imposed constraints and activities. But in academic research contexts, ethical debates focus on how researchers treat humans within their studies. That is, according to Markham, in the context of: "policy-making, laws and regulations - which have probably developed in particularly cultural contexts over long periods of time - predetermine ethical parameters" (Markham, 2015). According to Bryman (2012), discussions about ethical principles in social research, or more precisely – the ethical transgressions, tend to revolve around certain issues that recur in different guises, which Diener and Crandall (1978) have broken down into four main areas: whether there is harm to participants; whether there is a lack of informed consent; whether there is an invasion of privacy; and whether deception is involved. Although these four principles overlap in part, Bryman points out that there is no doubt that these four areas form an appropriate classification of ethical principles in and for social research.

The advent of the Internet and its use in social research has raised new dimensions of ethical decision-making for social researchers. The biggest challenge in conducting research in the internet community is the issue of informed consent, that is the risk of violating ethical principles of privacy and confidentiality. As a starting point in such an ethical dilemma, Essaybach and Till (2001) suggest that it should be the determination of the material, posted by the internet community, as public or private. In addition, Bryman (2012) points out that the more the website is recognized as public, the less the researcher's obligation to protect the confidentiality and anonymity of the website users, or the less the respondent's obligation to seek informed consent. But very often, the difference between the public and private sphere of the Internet is vague and disputable. Also, complete protection of anonymity is almost impossible in internet surveys, which arises from the difficulty of knowing who has access to the information. This, according to Bryman, places special ethical concerns on quality researchers, because a message posted on the social networking websites may be available for anyone with a computer and internet connection. Here I would list Pace and Livingston's (2005) point of view, according to whom such electronic communications should be used for research only if: the information is publically archived and readily available; no password is required to access the information; the material is not sensitive in nature; and no stated site policy prohibits the use of the materia. Moreover, they suggest that if these circumstances do not pertain, informed consent should be provided and it should be provided without disrupting the ongoing online activity. They also point out that identities and confidentiality must be protected.

According to another point of view (Jouhki et al., 2015) research ethics in any academic research can be seen as balancing between two classic moral philosophical stances – utilitarianism and deontology. Utilitarianism attempts to calculate the morality of an act by estimating the total amount of happiness or suffering produced by the act, while deontology views certain actions as immoral or moral per se, regardless of their consequences. In this regard, the Jouhki et al. (2015) indicate that: "in social media research when scholars contemplate the effect of their study on the subjects' privacy: the utilitarian view of privacy might allow certain incursions into privacy if the result is the greater good, whereas deontological ethics holds that a certain level of privacy is a right that should not be violated by, for example, conducting a study without receiving the informed consent of the subjects of the study". Ultimately, they emphasize that the utilitarian emphasis on avoidance of harm and the more deontological value of receiving informed consent from research subjects are the two most crucial imperatives of research ethics in studies with human participants.

EMERGE OF ETHICAL ISSUES IN SOCIAL MEDIA SURVEYS

The representation and importance of social media can best be shown through the statistics according to which in April 2022 there were five billion internet users worldwide, which is 63 percent of

the world's population. At the same time, 4.65 billion of them were social media users (Statista, 2022). Social networking is the most common online activity, which resulted in global average penetration rate of 58.4 percent (Statista, 2022). This presence of social media resulted in their positioning as the most influential media on public opinion. The ability to access via smartphones has increased the speed of initiating any kind of public debate, thus establishing the undisputed position of social media on public opinion. This technological advancement has introduced social media platforms in the areas of Facebook, Twitter, YouTube and others, and thus their presence on the Internet. First ethical issues that emerged at the beginning of usage of social media in scientific surveys, like issues of identifiability of participants, vulnerability, potential harm, intrusiveness, and confidentiality, were developed into the first and second Association of Internet Researchers ethics guidance for researchers using social media data (Ess and the AoIR ethics working committee, 2002; Markham and Buchanan, 2012). But with the further technological advancement of social media platforms, social media research methodologies have also evolved, as they have progressed to more analytical and autonomous processes based on large quantitative analyzes and "data models". Consequently, debates, which initially revolved around consent and privacy for social media data use as an isolated source, were reshaped by large-scale "big data" surveillance, and scraping and mining research employing artificial intelligence and data modeling methods (Samuel and Buchanan, 2020).

In addition to the research benefits of social media development, in recent years, there have also been a number of research ethics controversies that gained public attention such as Facebook Emotional Contagion study that studied Facebook users without their consent; OkCupid, a social media dating service, public release of non-anonymized data scraped by a researcher against the site's Terms of Service; and perhaps the most media-covered case of this kind – Cambridge Analytica scandal – misuse of data of an estimated 87 million Facebook users by data firm Cambridge Analytica to build voter profiles (Markham, 2015). Therefore, Whiteman (2012) points out that the public/private distinction has specific implications for social science research. According to her: "it plays a key role in academic discussion of the legitimacy of certain methods and, in particular, the contested acceptability of covert methods of data collection. This is because, despite a 'nearly universal' understanding of the need to protect the private sphere, there is a tradition of observational research in which 'it has been accepted that behaviour that is performed within the public domain may be observed and researched without consent'. In raising questions of privacy, the public/private distinction therefore raises fundamental 'questions about the necessity of consent.'" Finally, Williams, Burnap and Sloan (2017) emphasize that a principal ethical consideration in most learned society guidelines for social research is to ensure the maximum benefit from findings while minimising the risk of actual or potential harm during data collection, analysis and publication. They also point out that potential for harm in social media research increases when sensitive data are estimated and published along with the content of identifiable communications without consent. These data usually include sensitive personal demographic information (ethnicity and sexual orientation), information on associations (memberships to particular groups or links to other individuals known to belong to such groups) and communications of an overly personal or harmful nature (details on morally ambiguous or illegal activity and expressions of extreme opinion).

ETHICAL GUIDELINES AND PRIVACY POLICY (LEGAL OBLIGATIONS)

Twitter was launched in 2006, as a online social networking and micro blogging service, as of January 2022 were registered 436 million monthly active users (Statista, 2022). Communication on Twitter is enabled through statements called 'tweets', which can contain up to 140 characters. Users are not obliged to state their real names and location, and are usually identified by their unique usernames. When creating an account, each user is referred to the terms and conditions of Twitter, where the open character of this platform is especially emphasized. The Privacy policy also highlights the privacy settings in which the character of the user account can be determined, that is the setting up of an open access account, or closed account, whose activities are visible only to the users that are registered as 'followers'. Additionally, due to the open nature of Twitter, users are informed that according to the Twitter's Terms of Service and Privacy Policy, their information may be collected and used by third parties (Twitter, 2022).

The growing popularity of Twitter as the primary online space for public expression of citizens' reactions to certain events has made it one of the most widely used data sources in social science research. One of the most significant features that drives such popularity of Twitter, as well as the influence of the Twitter audience on various topics of social life, according to Stephen Dann (2015), arises from the default public nature of Twitter versus the default private nature of other social networks. According to him, this

perception stems from the fact that Twitter users post content on the public timeline, which can be viewed from a public website, thus creating a source of secondary published data. This perception has also been reflected in the view of a number of researchers who believe that the ethical and legal justification for using data without informed consent stems precisely from Twitter's privacy policy. Specifically, the Twitter's Developer policy states that: "academic researchers are permitted to distribute an unlimited number of Tweet IDs and / or User IDs if they are doing so on behalf of an academic institution and for the sole purpose of non-commercial research" (Developer policy, 2022). However, Twitter imposes certain additional conditions on this permission of the researchers, so that when distributing the tweets, the user ID must be provided. Also, if certain content has been deleted, or given protected status, or otherwise been suspended or removed from Twitter, the developers should delete or modify that content as soon as possible (Developer policy, 2022). This Twitter privacy policy allows researchers to collect tweets, which they use for analysis through the API (Application Programming Interface) platform, without the prior informed consent of the users whose tweets have been collected. Based on this, users, in most cases, are not approached directly to obtain their informed consent to participate in a specific survey, as Twitter's Developer policy is considered to allow this. But social media users' interest in using their posts is also complementing this trend. This situation was determined in a survey conducted in 2011, regarding the legal implications of the posts of users of various social media, and it was found that only 18 percent of respondents read the terms and conditions for posting on the social media they use (Zimmer and Proferes, 2012). However, in most cases, the research is conducted within universities or is funded by institutions that impose certain ethical standards. The principles and rules deriving from these standards often go beyond just binding legal standards. Also, violations of the ethical principles of universities, that is institutions entail sanctions which in certain cases may result in job loss or further and future funding of research.

Facing a growing number of ethical dilemmas that have arisen from the use of the benefits of digital technology and the Internet in social research, a number of scientific societies have developed specific ethical guidelines for research in digital environment. So in the past years such guides have been introduced by the British Sociological Association, the British Psychological Society, the European Society for Opinion and Market Research, the International Visual Sociology Association and others. However, the Association of Internet Researchers (AoIR) was the first to developed ethical guidelines for internet research in 2002, designed as recommendations for researchers, students, developers, that is all those who face ethical concerns during their Internet research (Ess and the AoIR ethics working committee, 2002). Following this initial version, two more ethical guidelines were introduced in 2012 and 2020, which are upgraded versions of this guideline (Markham and Buchanan, 2012; Franzke et al., 2020). What these guidelines have in common, according to Williams et al. (2017), is that they generally adopt the principle of situational ethics: "that each research situation is unique and it is not possible simply to apply a standard template in order to guarantee ethical practice". Furthermore, these ethical guidelines on the issue of informed consent indicate that in cases where it is not possible to obtain, the analysis must be conducted on depersonalized data. This position, to a large extent, prevails in the later acts and documents introduced by various institutions and universities, so in the US a 2015 amendment to the 'Common Rule' Federal Policy for the Protection of Human Subjects it is stated that certain forms of online behaviors can be classed as public behavior, hence there is no need to require additional ethical review. Whereas it is stated that: "Any research involving standardized testing, surveys, interviews, or observations, including audio and video recording, of public behavior, including behavior online, will be able to proceed without further review" (HHS, 2015). The British Psychological Association and the Association of Internet Researchers also point to the need for careful consideration of ethical issues when using social media as a source of research data, particularly in relation to privacy breaches. It is also important to cite the University of Aberdeen recommendations from a project funded by the Economic and Social Research Council, which states that in the case of sensitive social media content, researchers should either consider using paraphrased / composite data instead of reproducing the actual posts or use an informed consent approach (Webb et al., 2017). Webb et al. (2017) also cites the updated 2016 Oxford University web-based research guidance, which states that when providing data from Twitter, researchers do not need to seek informed consent from participants/users, unless they plan to publish individual posts in a specific report or scientific paper.

PUBLIC ATTITUDES ON TWITTER DATA USAGE

The lack of consensus in the academic community, as well as some controversy regarding the use of data from social networks, initiated research on users' perceptions regarding the use of data. In one of the first such studies, conducted by Beninger et al. (2014), it was found that for social media users, the need for informed consent or anonymity guarantees depends on the following factors: mode and content of the posts; social media website being used; the expectations the user had when posting, and; the nature of the research. In addition, the participants in this research clearly stated that they are particularly concerned about posting their pictures, the publication of sensitive or personal posts, as well as the publication of their usernames in research papers. What is particularly noticeable is the perception of content posted on social media websites for fun purpose as much more "personal" compared to content posted on social media websites for "professional" purpose. Respondents also expressed less concern about the use of their tweets in research without prior consent, compared to their Facebook posts, which the authors said stemmed from Twitter being understood as a public network, as well as its more open privacy settings. In 2015, Evans et al. (2015) conducted research in the UK on users' attitudes regarding ethical principles for social media research. According to their findings, 38% of the respondents believe that the sharing of their data from social media with third parties for research purposes is carried out according to the terms and conditions that were indicated to them on the social media sites. Nevertheless, 60% of them think that their data on social media sites should not be shared with third parties for research purposes. However, this attitude varies among the respondents depending on whether that data was already publicly available (as is the case with Twitter posts), the level of anonymity, as well as the purpose of the research and how much personal data is contained in the data used in the research.

A 2017 study by Williams et al. (2017), conducted on Twitter users in the UK, found that respondents had the lowest level of concern about using their data in scientific research (84% of respondents were not at all or only slightly concerned). While, users' concern about the use of their data by other actors is significantly higher, with 49% being quite or very concerned about the use of data by the government, and 51% being quite or very concerned about commercial settings. However, despite the low level of concern of the respondents about the use of their data in scientific research, still almost 80% of them expect their consent to be requested before publishing their data in scientific research, while 90% of the respondents expect to remain anonymous in scientific publications.

CONCLUSION

The wide range of possibilities that have emerged with the advent of the Internet have also been reflected in scientific research. Online surveys have made it possible to reduce the cost of research, which was one of the main difficulties, as well as to increase the range of respondents. Additionally, with the advent of social media, scientific researchers have recognized the potential for simpler provision of participants, as well as significantly facilitated communication and data collection. However, the increased use of social media, that is Internet platforms and applications in data collection in research, has resulted in increased confrontation of researchers with the conditions of use, that is ethical principles that dictate the conditions under which such research activities can take place.

Twitter data is more openly accessible than other social media platforms, and in web research, it is standard practice to collect tweets for analysis through the platform's API. In doing so, users are usually not approached to seek their informed consent to participate in the research, instead it is assumed that the approval arises from the Twitter's Terms of Service and Privacy Policy, that is their acceptance by the user. However, the emergence of numerous ethical controversies regarding the use of data from social media, as well as the manipulation of users for various researches, has attracted the attention of the public. Public pressure has sparked debate over the applicability of established ethical principles in the use of social media data in scientific research. This has resulted in the development of specific ethical guidelines for scientific research in the digital environment by a number of scientific societies. In the analysis of such ethical guides, that is the use of data from Twitter in scientific research, it was observed that there is no consensus among the academic community in determining the appropriate way of collecting and using data from social media, including Twitter. The most obvious are the disagreements about relying solely on the Terms of Use of Twitter when collecting data. According to certain guidelines and many studies it is a standard practice, but it is problematic, especially in terms of the already established ethical principles in scientific research. In such a debate, researchers are brought into a position of constantly negotiating ethical positions, as well as their interpretation and introduction of alternative ethical positions.

According to certain authors, the views of the public, that the users whose data are used in research, must also be taken into account. Analyzing their findings, it was found that although, compared to other platforms, Twitter users are much less concerned about the use of their data in research, they are still particularly concerned about the amount of personal data contained in data that will be used in the research. In doing so, users emphasize the publication of their usernames, as well as the publication of their posts with sensitive content in research papers. Also, most of the respondents, that is users expect their consent to be previously requested for the use and publication of their data in research, especially emphasizing the request for their anonymity in research publications.

Given all this, we could conclude that from an ethical perspective, the practices established in the use of Twitter data are to some extent in collision with the ethical guidelines and attitudes of users. Hence, in order to maintain scientific credibility, it is necessary to summarize the different views and aspects when conceptualizing the methodological approach for research that uses Twitter data.

REFERENCES

- Beninger, K., Fry, A., Jago, N., Lepps, H., Nass, L. & Silvester, H. (2014). Research using Social Media; Users' Views. NatCen Social Research Retrieved from: <https://www.bl.uk/collection-items/research-using-social-media-users-views> (October 30, 2022)
- Bryman, A. (2012). *Social Research Methods*, 4th edn. Oxford University Press
- Dann, S. (2015). Twitter Data Acquisition and Analysis: Methodology and Best Practice. In Burhalter, J. & Wood, N. (eds) *Maximizing Commerce and Marketing Strategies through Micro-Blogging*. Business Science Reference
- Developer policy [website]. (2022, August 27) Content redistribution Retrieved from: <https://developer.twitter.com/en/developer-terms/policy>
- Diener, E. & Crandall, R. (1978). *Ethics in Social and Behavioral Research*. Chicago: University of Chicago Press
- Ess, C. & the AoIR ethics working committee (2002). Ethical decision-making and Internet research: Recommendations from the aoir ethics working committee. Retrieved from: <https://www.aoir.org/reports/ethics.pdf> (September 12, 2022)
- Evans, H., Ginnis, S. & Bartlett, J. (2015). #SocialEthics: A Guide to Embedding Ethics in Social Media Research. London: IpsosMORI Retrieved from: <https://www.ipsos.com/sites/default/files/migrations/en-uk/files/Assets/Docs/Publications/im-demos-social-ethics-in-social-media-research-summary.pdf> (September 10, 2022)
- Eysenbach, G. & Till, J. E. (2001). Ethical issues in qualitative research on internet communities. *British Medical Journal*, 323 (7321)
- franzke, a. s., Bechmann, A., Zimmer, M., Ess, C. & the Association of Internet Researchers (2020). Internet Research: Ethical Guidelines 3.0. Retrieved from: <https://www.aoir.org/reports/ethics3.pdf> (September 12, 2022)
- Jouhki, J., Lauk, E., Penttinen, M., Rohila, J., Sormanen, N. & Uskali, T. (2015). Social media personhood as a challenge to research ethics: Exploring the case of the Facebook experiment. Social Media Research Symposium "Succeeded and failures in studying social media: issues of methods and ethics", University of Jyväskylä, Finland
- Markham, A. (2006). Ethic as Method, Method as Ethic: A Case for Reflexivity in Qualitative ICT Research. *Journal of Information Ethics* 15 (2)
- Markham, A. & Buchanan, E. (2012). Ethical decision-making and Internet research 2.0: Recommendations from the AoIR ethics working committee. Retrieved from: <https://www.aoir.org/reports/ethics2.pdf> (September 12, 2022)
- Markham, A. (2015). Producing ethics [for the digital near future]. In Lind, R. (Ed.). *Producing theory in a digital world 2.0: The intersection of audiences and production in contemporary theory* Vol.2. Peter Lang Inc.
- Pace, L. A. & Livingston, M. M. (2005). Protecting Human Subjects in Internet Research, *International Journal of Business Ethics and Organization Studies*, Vol.10
- Samuel, G. & Buchanan, E. (2020). Ethical Issues in Social Media Research. *Journal of Empirical Research on Human Research Ethics* 15(1-2)
- Statista.com [website]. (2022, May 20) Worldwide digital population as of April 2022, Retrieved from: <https://www.statista.com/statistics/617136/digital-population-worldwide/>
- Statista.com [website]. (2022, May 20) Social media: global penetration rate 2022, Retrieved from: <https://www.statista.com/statistics/269615/social-network-penetration-by-region/>
- Statista.com [website]. (2022, May 20) Most popular social networks worldwide as of January 2022, ranked by number of monthly active users Retrieved from: <https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/>
- Twitter Terms of Service [website]. (2022, August 25) Retrieved from: <https://twitter.com/en/tos>
- U.S. Department of Health and Human Services: Office for Human Research Protections. (2022, October 03). Code of Federal Regulations (Pre-2018 Requirements). Retrieved from: <http://www.hhs.gov/ohrp/regulations-and-policy/regulations/nprmhome/index.html>
- Webb, H., Jirotko, M., Procter, R., Stahl, B.C., Rana, O., Burnap, P., Housley, W., Edwards, A. & Williams, M. (2017). The Ethical Challenges of Publishing Twitter Data for Research Dissemination. *ACM Web Science*
- Whiteman, N. (2012). *Undoing Ethics: Rethinking Practice in Online Research*. Springer
- 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)
- Woodfield, K. & Iphofen, R. (2018). The Ethics of Online Research. In Woodfield, K. (ed.), *The Ethics of Online Research (Advances in Research Ethics and Integrity, Vol. 2)*. Emerald Publishing Limited
- Zimmer, M. & Proferes, N. (2012). Privacy on Twitter, Twitter on Privacy. In Weller et al. (Eds.) *Twitter and Society*. Peter Lang Publishing