Just Because the Data Is There, It Doesn't Mean It's Yours to Take: Exploring User, Researcher and Review Board Perceptions in Twitter Data Research

Kate McCandless

Word Count (excluding references): 6617

Abstract

In research conducted using Twitter data, informed consent has taken the back seat. This literature review examines the perspectives of users, researchers and research ethics boards to provide nuance and context to the issue. Users are generally unaware that their data can be taken for research purposes and that they have agreed to be studied within the platform's terms of service. This is concerning for both researchers and users alike, as it continues to blur the line of public and private information. Users want to be informed when they are being studied. When informed consent is not obtained, researchers are not respecting the data and the humans who created it. If researchers were required to obtain informed consent when engaging with Twitter data, the resulting research would be more ethical and protect everyone involved: the researcher, the user, and the university.

Keywords: information ethics; ethical research; Twitter research; informed consent; social media research

McCandless, K. (2021). Just because the data is there, it doesn't mean it's yours to take: Exploring user, researcher and reviewer board perceptions in social media research. *Emerging Library & Information Perspectives*, 4, 34-61.
<https://doi.org/10.5206/elip.v4i1.13554>

Introduction

The practice of conducting research via Twitter is increasing in the academic community. A general search for "twitter", "research" and "ethics" in Western Libraries' OMNI Catalogue yields over forty-one thousand results, including articles, books, conference proceedings, and more. Twitter-based data from this research is readily available and can yield highly informative results that are applicable to a range of disciplines. There is, however, concern amongst experts regarding a lack of ethical considerations from researchers. The first concern is a lack of *informed* consent from the research participants. There is a general unawareness amongst users that data can be taken from Twitter and used for research purposes (Fiesler and Profores, 2018; Kennedy, 2017; Parsons, 2019; Williams et al., 2017). Some users do not know that they have agreed to this in the terms of service established by the platform. A second issue is what constitutes public or private information. While users may have a public facing account, they may not intend every tweet to reach a public audience. There is a difference between having a tweet out in the Twitter-verse amongst millions of other tweets and having that tweet quoted in a news article or in research. Using these public-facing data without seeking express, informed consent from the author/user could be understood as lack of respect for that person's experience, data, and privacy. This may constitute using these people and their data as a mere means to an end. Twitter-based research conducted without informed consent is within an ethical grey zone. These ethical considerations need to be carefully examined when undertaking research that relies upon Twitter data.

The following research question guided my inquiry into recent literature on the subject: how do researchers, Research Ethics Boards and users view the current practices of conducting research on Twitter? I begin by contextualizing human research ethics in Canada, followed by a brief examination of Twitter's policies. I then conduct a literature review focusing on the researcher, review board and user perspectives on Twitter research. I conclude my article with a discussion on the implications of this mode of research and provide some recommendations that could be implemented to increase its overall efficacy and validity within the academic research community.

Contextualizing the Exploration

Human Research Ethics

Following the Second World War, there was a need to establish ethical guidelines for human research and medical practice. In 1948, the Declaration of Geneva was established which outlined the ethical practices of medical professionals. This later influenced the Declaration of Helsinki, which focused specifically on clinical trials and required consent from participants. Revisions in 2000 broadened the scope of the document beyond clinical trials and addressed the need to weigh the benefits and risks to the participant. This document has guided individual policy statements from institutions and governments and thus jurisdictional variations exist.

In Canada, the Tri-Council Policy Statement (TCPS) outlines the value of research and human research ethics. All research, regardless of field or future impact, needs to be conducted responsibly and transparently. The TCPS (2018) notes that respect for human dignity is the core motivation of the document. In order to conduct and produce research at a university, one must comply with the TCPS document and

meet specific standards to protect the university's interests and funding. This includes completing training on the ethical standards in research and projects being approved by a Research Ethics Board (REB). REBs may also be termed Institutional Review Boards (IRBs); the title varies depending on the organization. As stated in the TCPS, "researchers' commitment to the advancement of knowledge also implies duties of honest and thoughtful inquiry, rigorous analysis, commitment to the dissemination of research results, and adherence to the use of professional standards" (Canadian Institutes of Health Research, Natural Sciences and Engineering Research Council of Canada, and Social Sciences and Humanities Research Council [hereafter TCPS], Chapter 1A). If one does not adhere to these standards, their research may be considered inadequate or unethical.

A core component of the document is the Respect for Persons principle. It "recognizes the intrinsic value of human beings and the respect and consideration that they are due" (TCPS, Article 1.1, Respect for Persons). The key element is the necessity for consent. Consent must be freely provided, informed, and ongoing, meaning that participants are able to remove themselves from a study whenever they wish. This informed choice to participate in research is based on an understanding of "the purpose of the research, what it entails, and its foreseeable risks and potential benefits, both to the participant and to others" (TCPS, Article 1.1, Respect for Persons). Informed consent ensures that the subject is protected and respected at all times. Consent, however, can be difficult to procure when working with mass amounts of social media data.

Twitter's Policies

Twitter's privacy policy makes it explicit to users that published tweets are public data. The policy notes that "you are responsible for your Tweets and other information you provide through our services, and you should think carefully about what you make public, especially if it is sensitive information" (section 1.2). Further, "By publicly posting content when you Tweet, you are directing us to disclose that information as broadly as possible, including through our APIs¹, and directing those accessing the information through our APIs to do the same" (section 1.2). Users can alter their settings to determine how public their tweets and Twitter account may be, but some information must remain collectable to participate on the platform. Section 4 of the policy document states that a user can "access or rectify this data at any time" and that a user can submit forms to "object, restrict, or withdraw consent where applicable to the use of data." These policies indicate that if a user wants to join Twitter, they must play by Twitter's rules. Some data can be restricted to the broad public, and some data can be removed if a user so chooses. But how many Twitter users have read this policy and truly understand it?

There are also documents available regarding the use of the API tool ("Developer Agreement" (2020) and a "Commercial and non-commercial use of the Twitter API" (n.d.)). The Developer Agreement is largely concerned with how to legally use and alter the API tools for various purposes. The second document outlines the differences in Twitter's distinction between commercial and non-commercial uses. They provide links to information regarding the products they have available for academic research and

¹ API stands for "automated programming interface" and allows researchers and companies to interact with Twitter and individual tweets.

some examples of current research that uses their tools. There are also tutorials available on how to use these products and forums to provide feedback. There is, however, no mention of ethics, advice for ethical reviews, or notes of consideration for researchers on these webpage documents.

Literature Review

In this discussion of conducting research ethically through Twitter, it is important to evaluate the perspectives of the various players involved. These groups are the REBs, researchers, and research subjects.

Research Ethics Board Perspectives

Henderson et al. (2013) explored how often ethics are discussed in studies that report social media as their data source. The authors concluded that “articles about research involving social media [were] largely silent about ethical issues” (p. 548). They also note that a third of these articles “did not even disclose whether they had ethics approval, let alone describe the ethical issues they may have considered” (p. 548). This points to a long-standing issue regarding the value of sharing ethical considerations in scholarly communication. In the literature, there has been a continuous thread indicating that REBs do not have members knowledgeable in social media or the workings of big data research (Buchanan and Hvizdak, 2009; Henderson et. al., 2013; Hibbin et. al., 2018; Sellers et. al., 2020). This lack of knowledge in social media is not inherently limiting; these REBs still have extensive experience in evaluating research proposals. Not every member of a REB needs to be an expert in social media or the Internet at large. They do, however, need to exercise mental flexibility and consider the various implications the research in question could have. Buchanan and Hvizdak (2009) in particular note

that Internet research is “qualitatively different from traditional forms of research” and encourage boards to “consider how best to translate ethics from one environment to another” (p. 41).

Flexibility and improvisation should not be the only tools upon which a review board should rely. Many organizations do have *some* guidelines in place that discuss social media research, but they are often vague and point to other external sources for clarity. In a 2018 article by UK researchers Taylor and Pagliari, the current guidelines for social media research were put to the test. The authors evaluated the guidelines supported by the seven subsidiaries of Research Councils United Kingdom (RCUK) and evaluated their compliance with a theoretical framework established by Conway (2014). The RCUK guidelines relied heavily on previous documents from external organizations, such as the National Institute for Health Research and the Association of Internet Researchers (AoIR). Taylor and Pagliari then initiated a search through PubMed to source articles that engage in social media research and measured their compliance with those RCUK recommended guidelines. From the 156 eligible articles, only 50 mentioned “one or more of the ethical concepts, procedures or approval processes specified in Conway’s taxonomy” (Taylor and Pagliari, 2018, p. 11). Furthermore, only thirteen of the eligible articles discussed “other relevant ethical considerations” beyond that of REB approval (p. 11). Through this study, it is established that researchers are aware of the “*procedures and requirements* necessary to obtain [REB] approval”, however, there is a lack of consideration for the “*moral or societal* implications of repurposing information that people have shared for reasons other than research” [emphasis original] (p. 24-25). The authors conclude their article by asking peer

reviewers and journal editors to “seek authors’ explanations of the ethical challenges they faced and how these were managed” (p. 28). They assert that ethical guidelines have been poorly integrated in current literature and it is important to continue to demand more transparency from the researchers regarding ethical dilemmas.

While Taylor and Pagliari examined issues at the researcher level, there are studies pointing to issues within the review boards themselves. This is echoed in Brooke Foucault Welles’ (2015) account of the issues involved in her PhD dissertation concerning online friendships. Welles describes examining a data set pulled from *Second Life* users to examine the connections these users formed within the game. Her initial REB application to examine a dataset of over six million users was denied as “[she] did not have consent forms from [her] users” (p. 225). There was no feasible way to obtain consent. The data from millions of users were encrypted so that no identifying information could be found or used. Welles notes that the REB did eventually give her the approval for the project and that the “initial denial stemmed largely from the [REB] reviewer’s misunderstanding about big-data research” (p. 226). She further notes that this review board had never examined a request for big-data research and the “staff did not understand how such extensive data could be collected from users without interacting with them directly” (p. 226).

Welles (2015) discusses how she approached the reviewer in a one-to-one conversation to break down the research project so that the board could understand its aims and goals. When she resubmitted the proposal, with a letter from Linden Lab (the owner of *Second Life*) detailing permission to use the data, the project was approved. This was approved under the assumption that the users “consented to having data

collected about their actions when they created their accounts” (p. 227). While these users *had* consented to creating an account and to having their data used in research via the terms of service, there is still the remaining grey area that Welles describes. Despite not having a direct interaction with her “participants,” studying friendships requires the assessment of intimate, human data. There is also “no way to demonstrate that subjects could reasonably be aware that they might be participating in research, or that they had the opportunity to opt out if they did not want to participate” (p. 227). According to the TCPS and Welles, this is a critical component of human research ethics.

Researcher Perspectives

Researchers have established that social media data is extremely important for understanding a variety of human phenomena. However, the lack of definitive guidelines for social media research adds another complicating factor. The following section will recount a variety of researcher-based perspectives on Twitter research ethics.

Parsons’ (2019) book “Ethical Challenges in Digital Psychology and Cyberpsychology” discusses social media research ethics. Parsons draws on Beninger (2017) to note that “ethical considerations in social media research are not ones that can be boiled down to a checklist” and thus no definitive guideline for researchers currently exists (p. 193). He also notes that “researchers and institutional review boards may under- or overestimate the potential risks for participants” and thus “cyberpsychologists may be discouraged from performing online research or even submitting studies for review” (Kosinski et al., 2015 in Parsons, 2019, p. 193). He

recounts the issues of identifying public and private spaces online, the context of the data or intention of the user, and informed consent. Parsons is clear that many of these social media users are not sharing their information on these platforms with the intention of creating research data, “they disclosed the personal information because they wanted to connect socially with others” (p. 195).

Parsons notes that a website’s policies or terms and conditions are not a sufficient demonstration of informed consent. The sheer volume of legal forms a user may interact with on the Internet desensitizes the user to their importance. “As a result of their length and ubiquity, users have grown accustomed to simply clicking ‘Agree’ without first reading the fine print” and therefore researchers cannot assume that every user is adequately informed (Böhme and Köpsell, 2010 in Parsons, 2019, p. 196). Similarly, if a user is not informed that research is being conducted on their data, they cannot be debriefed or pointed to supports where required. This should be worrying for researchers, especially when working with sensitive and personal topics. Parsons emphasizes that “just because data is publicly available online does not mean that it was obtained legally or that it conforms to the standards of scholarly ethical practice” (p. 203). Ultimately “researchers are responsible for establishing a data set’s origins and the nature of consent given by users” (p. 203). He then provides several charts that researchers and REBs can reference when evaluating the need for informed consent in their research projects.

Moreno et al. (2013) examined the issues of social media data mining practices through the lens of health information. These authors sought to identify if there are unique risks or benefits to the participant when conducting research via social media

sites. A notable issue identified by the authors related to confirming the identity of account holders (pp. 710-711). This can add a complicating element for researchers, especially as alias accounts are quite common. While there is no clear answer for how to handle this complication, Moreno et al. conclude that researchers need to be selective and be aware of the context of the data they are using. They additionally note that researchers need to “consider applying these [ethical] principles by only collecting the essential data needed to answer the research question, and presenting those data carefully to avoid participant identification” (p. 712). This highlights two important factors: first, the collect-all and sift-later approach that many big data researchers use is not appropriate, and second, researchers must take steps to mitigate identification.

Fuchs (2018) discusses the unique challenges for researchers probing “taboo” or controversial topics. He, like Parsons, notes that a firm guideline for ethical procedures in social media does not exist. He suggests that in studying what guidelines are available, a researcher can devise a plan that suits their research (section 2). Fuchs also identifies the two prominent trains of thought in the debate of Internet research: “while privacy fetishism is one extreme, another its opposite pole is the complete ignorance of research ethics” (section 3). He thus calls for the development of “critical-realist digital media research guidelines that go beyond research ethics fundamentalism and big data positivism.... The approach has to both engage with research ethics and enable the conduction of actual online research” (section 3).

Fuchs emphasizes that it is important to continue to study controversial topics in an Internet context, such as far-right movements, but notes that topics of this nature can be precarious. The researcher themselves can easily become a target for backlash.

Fuchs justifies his opinion through two examples of his own research involving fascism and white supremacy and explains why explicit consent was not sought. In many cases users are engaging with popular hashtags, thus taking their tweets to wider, public audience. He further notes that “obtaining informed content [*sic*] when working with a large online dataset is for the most part practically impossible due to time restrictions” (section 2). This suggests that when conducting research on more delicate topics, Twitter is an ideal site of investigation due the open availability of the data.

These debates among researchers regarding informed consent have carried on into more informal channels, such as personal blogs and Facebook posts. In August 2020, Twitter expanded their Developer API to better suit the needs of academic researchers and businesses. It appears as though these changes were considered and implemented after sourcing feedback in their forums. In one conversation from mid-July 2019, the issue of informed consent was raised and acknowledged (*Informed Consent API*, July 17, 2019). The researcher’s initial post suggested that there should be a tool implemented from Twitter itself that can notify users when their data is examined. As I could not confirm the identity of each of these users, nor did I have the opportunity to communicate with them about reproducing their posts, I have identified each respondent as “R” with their corresponding order in the forum conversation. The first researcher to post in this forum asks Twitter to

“Implement an API / Bulk access call to notify users that their tweets were / are used in a study, with the option to opt-out, or opt-in to more information from the researcher. Maybe a DM with links to IRB / Home Institution / Researcher contact details?” (*Informed Consent API*, R1, July 17, 2019).

This researcher's sentiments were echoed in comments from others using Twitter data in their research. One notes: "This would increase so much the ethical standing of our research" (R6, July 25, 2019). Another wrote that: "Local university IRB is generally not helpful with these questions and Twitter users are put into potentially vulnerable positions. Building in an informed consent feature would be a positive, honestly overdue step in the right direction" (R4, July 18, 2019). There is a desire from the users to be informed when one's data is being used. Informing these "subjects" that their data has been examined through a feature built-in to Twitter's API could rectify this. It also respects the person's privacy and alleviates the concern that they could be used as a mere means for the research.

In two separate blogs authored by researchers, a Facebook conversation on the subject was reproduced (Fish, 2010; Zimmer, 2010). The discussion began in a workshop entitled "Revisiting Research Ethics in the Facebook Era: Challenges in Engineering [Computer Supported Cooperative Work] Research" and continued on Michael Zimmer's personal Facebook page (Zimmer, 2010). The question and discussion revolved around exactly what is being discussed in this paper: "[is it] ethical for researchers to follow and systematically capture public Twitter streams without first obtaining specific, informed consent by the subjects?" (Zimmer, 2010). He argues that while setting a Twitter stream to "public" does entail others to search, follow, and read your data, there is the "reasonable expectation that one's tweet stream will be 'partially obscure' within the thousands (if not millions) of tweets similarly publicly viewable" (Zimmer, 2010). Zimmer then edited and condensed the most salient points from the Facebook conversation in his blog post.

Adam Fish was also involved in this Facebook debate and reproduced the conversation in full on his blog. Zimmer took steps to request permission from each person involved in the conversation before allowing Fish to reproduce it on his blog. As such, I have retained the names as they were published; permission has been sought already and a search for these quotes will yield these names. One respondent, Siva Vaidhyanathan, wrote: “Tweets are publications. I think it’s absurd to even consider IRB review for anything dealing with things people have published.” Zimmer responds to this: “does this extend to any utterance? Is public speaking a publication?” Siva then responds in two separate comments:

“A publication is fixed in a tangible medium of exchange” and “IRB processes impede social science, humanities, history, and journalism research and academic freedom when misapplied to adults in public roles or forums. There is no reason why IRB review should be required for oral histories of adults. Rules about anonymity etc. stifle history and journalism. We need a complete overhaul of the system which treats everyone like children” (Vaidhyanathan in Fish, 2010).

Various other participants in the debate note that there is an apparent gap between REB standards, the participant’s expectations regarding ethical treatment, and “what each ethical theory can justify as ethical” (Hunsinger in Fish, 2010). This conversation does not settle on a collective, definitive answer. Fish notes in the introduction of his post that “Exchanges like this are instrumental both because they address a serious and misunderstood facet of new media social research and as a sample of a reflective public working out theory in real time” (Fish, 2010). This issue is

still being addressed and “worked out” in scholarly writings, news articles and personal correspondence today.

Participant and User Perspectives

While the researchers and REB perspectives have been examined, opinions from the participants themselves should be considered as well. User perspectives are mixed: most want to support academic research and are happy to share their data, but there still remains serious concern regarding the original context of a tweet as well as the possible identification of users in published research.

Fiesler and Proferes (2018) conducted a study interviewing 286 Twitter users who participate in crowdsourcing on the Amazon platform, *Mechanical Turk* (or MTurk). This is a platform where users complete small tasks, such as responding to a survey, for monetary compensation. The participants were asked to reflect on a number of scenarios and context factors related to sourcing data from Twitter for scholarly research. One question posed was if the participants were aware that researchers sometimes use tweets as data in academic work. Nearly two thirds of the respondents (61.2%) were not aware that this occurs (p. 5). Close to half of respondents (42.7%) did not think researchers were allowed to use this data without permission from the user (p. 6). Some of those who believed that researchers were not allowed to use this data thought that the Twitter terms of service provided protection. In reality, Twitter allows this kind of research to occur. Fiesler and Proferes ultimately asked the participants if they would want their data included in research, and again, the sample was divided; 64.9% of participants did not think researchers should be able to use this data without express permission (p. 6). Context is everything became the underlying comment

across the study. There is concern from respondents that if the text of the tweet is removed from its context the original meaning will not stay fixed to the message (p. 6).

This study points to several issues. First, participants are not aware that these public data can be used, or perhaps are not aware the data are considered “public” at all. Second, there is a clear desire to be informed when one’s data is being observed. Although this observation may just be mining the metadata from tweets, there can still be identifiable information located within. Third, context is everything. Users want to ensure that their message is not being skewed by the researcher; they want the use of their data to be integrous and transparent.

A similar study was conducted in the UK regarding predicting users’ perceptions of Twitter research (Williams et al., 2017). The findings of this study are also echoed in Beninger (2017) as they rely on the same data. These researchers examined the current laws and regulations on the subject from “learned societies” such as the British Educational Research Association (BERA) and the AoIR (Williams et al., 2017, p. 30). The authors note that most guidelines adopt the “situational ethics principle: that each research situation is unique, and it is not possible to apply a standard template in order to guarantee ethical practice” (p. 30). The guidelines that are developed by the AoIR acknowledge the complex issues surrounding questions of what constitutes human subject research, while BERA focuses more on the problem of informed consent and privacy with Internet research.

These researchers employed qualitative and quantitative research strategies to gauge user perceptions. They first investigated the questions through focus groups and interviews with thirty-four diverse participants. Next, a survey was administered through

social media to a sample of Twitter users that reflect the platform's population. Steps were taken to account for bias and the authors clearly state that they "made efforts to establish informed consent via the introduction page to the online survey" (Williams et al., 2017, p. 37).

In the qualitative interviews, three kinds of users were established: creators, who post original content; sharers who share or retweet other's content; and observers, who tend not to pass on information, but do read and view social media (p. 37). From these users, three concerns were raised: "maintaining privacy, protecting the reputation and identity of themselves, friends and family, and ensuring safety" (p. 37). There was a view shared across many participants that it is hard to stay up to date with privacy policies and settings of various social media spaces. Some were unaware of how to alter privacy settings or to what the settings actually applied (p. 38). These participants then were asked questions pertaining to research on social media. The results were once again compiled into three categories: acceptance, skepticism, and ambiguity (p. 39). The users that accepted this kind of research were the same ones who were generally aware of how much content can be viewed and traced online. The skeptics expressed many of the views echoed in other literature (p. 39). This included a concern over the lack of informed consent, a lack of transparency regarding who is conducting the research, and that information may be taken out of context. Across the group, there was the perception that researchers should ask for permission when publishing text alongside any identifiers, such as a username, and that researchers should do everything within their power to secure consent when working with more sensitive or emotional information (p. 40). Participants suggested that "no matter how open or public

a site is considered to be, the *user's* expectation about how the post should be used is what should be considered" [emphasis added] (p. 40). If a user intended for a post to be widely shared, then the general consensus was that researchers could examine it.

In the quantitative survey, 564 responses were garnered and evaluated (Williams et al., 2017, p. 41). Nearly all respondents were aware of Twitter's terms of service and two-thirds had read at least half of the document; "Seventy-six percent knew that when accepting the terms of service, they were providing consent for some of their information to be accessed by third parties" (p. 41). In general, users were not concerned if their data was used in research from a university but were less enthused when the research was conducted by a government organization or private company. Interestingly, "Just under 80 percent of respondents agreed that they would expect to be asked for their consent before their Twitter posts were published in academic outputs" and "Over 90 percent of respondents agreed that they would want to remain anonymous in publications stemming from Twitter research based in university settings" (p. 41). Participants who were quite familiar with terms of service did not share these views; they expressed very little concern over the use of their data (p. 44).

The findings from the study from Williams et al. (2017) and Beninger (2017) note that there is a wide range in user views on this subject. Many of the respondents want to be informed about research and believe researchers should secure consent. They also feel that when this data is reported in academic publications, the subjects need to be protected through anonymization techniques in order to avoid abuse (Williams et al., 2017, p. 47). There are some participants who feel that these concerns should be mitigated by the user themselves through taking steps to ensure their social media

accounts are using the appropriate privacy settings. When the results of the survey and interviews are examined together, “these findings show that there may be a disjuncture between the current practices of researchers ... and users’ views of the appropriate use of their online information and their rights as research subjects” (p. 45). As Beninger (2017) bluntly states, “thinking is not optional when it comes to applying ethical frameworks to changing online environments” (p. 58). This research points to a need to continue to examine these practices and the views of all parties involved.

Hutton and Henderson also examine the issue of informed consent in their 2015 article. They ultimately propose a new method for ensuring ongoing, informed consent in research social media data. “Contextual Integrity,” stemming from Nissenbaum’s 2004 work, finds the balance between continuous (sustained) and single-instance (secured) consent. Secured consent is obtained at the beginning of a study and encompasses all of the data that could be probed and collected. Sustained consent relies on continuous requests to collect data (Hutton and Henderson, 2015, p. 181). The Contextual Integrity model asks a number of questions of the participants, seeking their approval or disapproval, in collecting a piece of data (p. 181). Once enough information has been collected, a person’s “norm conformity” level is evaluated, and it can be assumed that the rest of the data will be approved by the participant. This kind of questioning has been proven to return a high accuracy in the results, reduce the burden on both the participant and the researcher, and is comprehensive enough in its coverage (Hutton and Henderson, 2015, pp. 182-183). These authors were able to confirm that seven questions can potentially determine a participant’s level of norm

conformity. This could be an approach for researchers to follow in the future, however, it does not fully address the issue of informed consent from participants.

Kennedy et al. (2017) similarly examine participants' opinion of social media research. They interviewed 65 diverse users from England, Norway and Spain (p. 275). These authors also support an approach based on Nissenbaum's Contextual Integrity model (p. 274), and thus used examples from "real-world social media data mining practices" in the scenarios discussed with participants during focus group interviews (p. 271). Their findings echo much of what has been reported by other authors; context is an important factor in users' minds. They further explain that "for many respondents, attitudes seem to be influenced by a number of factors related to the fairness of data mining practices, such as the sensitivity of the information, the purpose for which the information is used and whether or not the user is aware of the monitoring" (p. 280). Additionally, the issue of being informed on research goals and giving express consent to be involved in the research was important to participants. "One participant asked whether the scenarios involved monitoring with or without users' consent, and when told that *all* were without consent, stated 'that's my sticking point'" (emphasis added, p. 283). These authors also identified that users want plain-language and transparent explanations for what happens with their social media data (pp. 283-284). Often this information is hidden in "five billion words" or the "fine print" of terms and conditions agreements (p. 284). As such, there was an assumption that "users frequently do not read the terms and conditions to which they agree on social networking sites" (p. 284).

From the literature reviewed, it is clear that participants want to be informed and involved in research that benefits greater society. Participants are concerned about the

potential for harassment when identified in research, but are also conscious of the fact that paraphrasing may skew the user's original intent. The common denominators in these studies highlight a desire for the participant and their data to be treated fairly and respectfully.

Discussion

While much of this article has focused on the ethical pitfalls regarding social media research, it is important to acknowledge that this mode of research is valuable to the broader community. The benefits and possibilities of this big data research are explored in a "news feature" article from *Nature*. Heidi Ledford highlights a number of studies that have used social media data to find patterns in human behaviour and societal norms (2020). Many computational social scientists argue that big data research can help us to "eventually start solving ... the major problems that affect the well-being of human populations" (King, 2007 in Ledford, 2020).

Ledford recounts other studies that have used big data to draw sweeping conclusions on society and how it functions. One study examined cell-phone records and compared them with results from a traditional survey to infer wealth and economic status in Rwanda. The conclusion was that the continual narrowing of the divide between qualitative and quantitative research of this nature is ultimately beneficial for everyone (Blumenstock et al., 2015 in Ledford, 2020). Ledford does highlight that this research needs to be grounded in social science theory for it to be truly successful (2020), which can be achieved through a continual push for interdisciplinary research from the researchers themselves. Collaboration across disciplines is the primary way to

ensure that current and future research is ethical, complete, and ready to be implemented in real-world applications and developments.

Through the literature examined, it is clear that there is no consensus on the twin issues of consent and intention. There are academics who view tweets and accounts as documents of the Internet: completely fair game for research. Fuchs (2018), for example, goes so far as to suggest that researchers do not need to seek consent from Twitter users. This both protects the researcher from potential harassment and keeps the research process convenient. It is difficult to square this with ethical research practice as it seems to disregard the author of the content completely. There are other researchers who urge that this kind of data cannot be examined without express approval from the REB and the users themselves. If users have had the purpose of the research articulated to them, and understand that the research benefits people as a whole, then there is no issue. When informed consent is not obtained there is concern from users regarding their data being used to promote causes they may not personally support.

These concerns fall under the umbrella of the TCPS' Respect for Persons principle. This ensures that people are treated with respect and that they expressly consent to be observed or studied as participant in a research study. Consent is not given when a tweet is published, and there is no aspect of being *informed* in this instance. The potential risks or benefits need to be communicated to the author of the tweet. If informed consent is not sought, and a person's data is examined without regard to potential impacts, the research cannot be considered fair, equitable, and just. Although anonymization of research data can attempt to ensure participants will not be

identified, improvement in Internet searching, and ongoing web archiving may lead to unwanted participant identification.

There is no simple answer to this complex problem. Conventions that govern human subject research, like the TCPS, have been established, justified and followed for nearly a century. As demonstrated in this paper, informed consent in Internet research has been grappled with by researchers and their institutions for over a decade. I feel that we need to treat any data produced by humans as human subject research. With our lives becoming more public through sharing in social media spaces, we have to be conscious of how using that data may impact users. What may appear as a fully public utterance may not have been intended to be singled out in a sea of millions of other tweets.

Perhaps Terms of Service need to be updated to better reflect the fact that some people do not read them in full. There is a general lack of awareness that Twitter's Terms of Service permit researchers access to all data available on the platform, at least in North American settings. It appears that users situated in Europe and the UK may be more aware of the clauses they agree to when creating accounts. Perhaps there is a way to notify users when their data has been used, just as the researchers in the Twitter Developer API forum have articulated. Perhaps a "Twitter-bot" can be developed that automatically replies to a Tweet when it is included in a dataset. This reply could include a link to the letter of information the researcher develops as well as contain an opt-out form to remove the tweet from the dataset. This would decrease the burden on the researcher in obtaining consent and would ensure that the participant could be informed on what the research project entails. A possibility for the social media

companies to consider is a clear and plain-language summary of the most salient clauses near the “I agree” checkbox. Perhaps it would be better to have an option for the user to opt out of specific clauses, like allowing researchers to access their data, just as there are opt in/out options for receiving promotional materials from companies when shopping online. We cannot ensure that users are fully informed about how their data may be used on these platforms, therefore, they are not automatically informed participants. We need to find a way to keep respect for persons and their data at the forefront of research. Covert research via mass data collection cannot continue in this manner. People need to be informed when they, and their data, are being observed for research purposes.

Conclusion

People have the right to be informed about the research in which they are participating. Complications arise when trying to accomplish informed consent with big data. There are suggestions that I feel are possible to implement and will make a difference in the research landscape. Researchers, ethics boards and users are all concerned over how data may be used in studies. This debate has extended for over a decade and there needs to be some tangible steps on how to raise the level of respect for all involved. By requiring informed consent from researchers, the resulting research will be more ethical. Having the REBs approve research as ethical protects not only the funding and reputations of researchers and their associated institutions but the data’s author’s interests as well. There is no reason to not try to adopt better practices, other than it being cumbersome or placing another step on the researcher. Respecting the people who have created this data should be the primary objective of all involved.

References

Beninger, K. (2017). Social media users' views on the ethics of social media research.

In L. Sloan, & A. Quan-Haase (Ed.s), *The SAGE Handbook of Social Media Research Methods*, (pp. 57-73). SAGE Publications.

<https://doi.org/10.4135/9781473983847>

Buchanan, E. A., & Hvizdak, E. E. (2009). Online survey tools: Ethical and methodological concerns of human research ethics committees. *Journal of Empirical Research on Human Research Ethics*, 4(2), pp. 37–48.

<https://doi.org/10.1525/jer.2009.4.2.37>

Fiesler, C., & Proferes, N. (2018). "Participant" perceptions of Twitter research ethics. *Social Media + Society*, 4(1), pp. 1-14.

<https://doi.org/10.1177/2056305118763366>

Fish, A. (2010, February 12). Mining Twitter and Informed Consent. *Savage Minds*.

Retrieved December 4, 2020, from <https://savageminds.org/2010/02/12/mining-twitter-and-informed-consent/>

Fuchs, C. (2018). "Dear Mr. Neo-Nazi, can you please give me your informed consent so that I can quote your fascist tweet?": Questions of social media research ethics in online ideology critique. In G. Meikle (Ed.), *The Routledge Companion to Media and Activism*. Routledge. <http://www.routledge.com/9781138202030>

Henderson, M., Johnson, N. F., & Auld, G. (2013). Silences of ethical practice:

Dilemmas for researchers using social media. *Educational Research and*

Evaluation: An International Journal on Theory and Practice, 19(6), pp. 546-560.

<https://doi.org/10.1080/13803611.2013.805656>

Hibbin, R. A., Samuel, G., & Derrick, G. E. (2018). From “a fair game” to “a form of covert research”: Research ethics committee members’ differing notions of consent and potential risk to participants within social media research. *Journal of Empirical Research on Human Research Ethics*, 13(2), pp. 149–159.

<https://doi.org/10.1177/1556264617751510>

Hutton, L. & Henderson, T. (2015). “I didn’t sign up for this!”: Informed consent in social network research. In *Proceedings of the 9th International AAI Conference on Web and Social Media (ICWSM)* (pp.178-187). AAAI Press.

<https://www.aaai.org/ocs/index.php/ICWSM/ICWSM15/paper/view/10493/10501>

Informed Consent API. (2019, July 17). Twitter Developer Feedback. Retrieved December 4, 2020, from

<https://twitterdevfeedback.uservoice.com/forums/922918-twitter-for-academic-research/suggestions/38168167-informed-consent-api>

Kennedy, H., Elgesem, D. & Miguel, C. (2017). On fairness: User perspectives on social media data mining. *Convergence: The International Journal of Research into New Media Technologies*, 23(3), pp. 270-288.

<https://doi.org/10.1177/1354856515592507>

Ledford, H. (2020, June 17). How Facebook, Twitter and other data troves are revolutionizing social science. *Nature*, 582(7812), pp. 328–330.

<https://doi.org/10.1038/d41586-020-01747-1>

- Moreno, M. A., Goniou, N., Moreno, P. S. & Diekema, D. (2013). Ethics of social media research: Common concerns and practical considerations. *Cyberpsychology, Behavior and Social Networking*, 16(9), pp. 708-713.
<https://doi.org/10.1089/cyber.2012.0334>
- Parsons, T. D. (2019). Social Media Ethics Section 2: Ethical Research with Social Media. In *Ethical Challenges in Digital Psychology and Cyberpsychology* (pp. 192-207). Cambridge University Press.
<https://doi.org/10.1017/9781108553384.011>
- Sellers, C., Samuel, G., & Derrick, G. (2020). Reasoning “uncharted territory”: Notions of expertise within ethics review panels assessing research use of social media. *Journal of Empirical Research on Human Research Ethics*, 15(1-2), pp. 28-39.
<https://doi.org/10.1177/15562664619837088>
- Taylor, J., & Pagliari, C. (2018). Mining social media data: How are research sponsors and researchers addressing the ethical challenges?. *Research Ethics*, 14(2), pp. 1-39. <https://doi.org/10.1177/1747016117738559>
- Twitter. (n.d.). *Commercial and non-commercial use of the Twitter API*.
<https://developer.twitter.com/en/developer-terms/commercial-terms>
- Twitter. (2020, March 10). *Developer Terms: Developer Agreement*.
<https://developer.twitter.com/en/developer-terms/agreement>
- Twitter. (n.d.). *Twitter Privacy Policy*. <https://twitter.com/en/privacy>

- Welles, B. F. (2015). Big data, big problems, big opportunities: Using internet log data to conduct social network analysis research. In Hargittai, E., & Sandvig, C. (Eds.), *Digital Research Confidential: The Secrets of Studying Behaviour Online* (pp. 223-242). MIT Press.
- Williams, M. L., Burnap, P., Sloan, L., Jessop, C., & Lepps, H. (2017). Chapter 2: Users' views of ethics in social media research: Informed consent, anonymity, and harm. In K. Woodfield (Ed.), *The Ethics of Online Research (Advances in Research Ethics and Integrity, Vol. 2)*, pp. 27–52. Emerald Publishing Limited.
<https://doi.org/10.1108/S2398-601820180000002002>
- Zimmer, M. (2010, February 12). Is it ethical to harvest public Twitter accounts without consent? *MichaelZimmer.Org*. <https://michaelzimmer.org/2010/02/12/is-it-ethical-to-harvest-public-twitter-accounts-without-consent/>