

VOL. 4, NO. 1, 2022, 1-22

ETHICAL USE OF INFORMANT INTERNET DATA: SCHOLARLY CONCERNS AND CONFLICTS

Michael Nycyk^a

ABSTRACT

This article explores the scholarly concerns and conflicts debated by authors in the field of the use of informant internet data in research. The importance lies in informant protection and how to minimize harm to them, a long-standing cornerstone of research practice. It is also a public domain issue as increased calls for data privacy grew because of reported data breach scandals. Although not a new problem, academic researchers and university ethics boards struggle with concerns over data use and are in conflict about managing the problem. This article uses thematic analysis to identify, analyze and interpret patterns of concerns and conflicts over internet data use. Data was obtained from academic publications on these issues. Three themes from this data are discussed with examples demonstrating the types of, and complexity of, scholarly concerns and conflicts. These themes are: the problems of informant data use risks, gaining mass informed consent and the challenges ethics boards face, especially conflicts with researchers over internet data use on projects. This article contributes insights into a widely, and continuously, debated area which is constantly evolving as privacy laws and public awareness place pressure on researchers and ethics boards to address protecting informant public internet data.

Keywords: ethics boards, informed consent, informant protection, internet data ethics, researcher concerns and conflicts, thematic analysis

^a Queensland University of Technology, Brisbane, Australia.

1 INTRODUCTION

In 2014, Adam Kramer and his research colleagues' emotional contagion study was published in an open access journal (Kramer, Guillory and Hancock 2014). Their study used Facebook user's personal data without consent. Scholars, and the public, were vocal in condemning the study. Flick (2016) argued that it took advantage of peoples' autonomy to decide how their internet data should be used. Hallinan, Brubaker and Fiesler (2020) stated that the study violated social media users' expectations of data protection and had negative ramifications on the public's perceptions of the use of their data for intrusive research. User data is especially regarded by commercial organizations as being for monetization, not taking into account that it is the users who produce and create content and give their data in good faith to social media platforms and the internet (Johanssen, 2021). Koenig (2018) likens the harvesting and use of user internet data as a form of serfdom, where data is given to an overlord in exchange for using that part of the internet or a social media platform.

Social media and other large datasets, also called big data, have the potential to contribute to understanding and solving global research problems. The logistics of obtaining informed consent from thousands of people to use their data is often not viable. People are concerned that they do not know how their data will be used or if it will be shared with others (Hand 2018). Issues such as the Cambridge Analytica Data Breach Scandal, and Kramer, Guillory and Hancock's (2014) misuse of Facebook data have alerted the public to this issue, but also challenged researchers and universities on how to manage this significant problem. Internet and data privacy scholars argue more critical questions about who gets access to what data, how the analysis is undertaken and to what it will be used for, need further crucial scrutiny and continuous scholarly debate (Boyd and Crawford 2012).

This article contributes insights into the issue of researcher use of informant's internet data and the concerns and conflicts this causes. It is not a new problem as scholars have used peoples' online data for research since the public internet became available. Scholarly reflection on this problem, and how to solve it, has emerged as a large field with a growing body of literature. It is not necessarily a difficult to control or unmanageable dilemma. Large public internet datasets offer insights not seen before in academic history. Identifying and managing concerns and conflicts is an important problem to examine. Data use ethical issues are growing as privacy laws and new technological developments arise, further challenging how to balance using informant data without identifying or harming informants.

This article uses thematic analysis methods (Clarke, Braun and Hayfield 2015; Braun and Clarke 2006) to identify and analyze themes that represent the concerns and conflicts of informant internet data use. In identifying these themes, it gives an appreciation of the problem of using large public datasets without informant's consent that concerns researchers. Taking an exploratory and inductive research approach to this problem, this article's contribution is to examine and

explain the problems of public data use that arise, based on examining scholarly writing from researchers. The article closes with a discussion of, and implications of, the themes' findings and also give a brief evaluation of using thematic analysis in this study.

2 CONCERNS AND CONFLICTS OVER PUBLIC INTERNET DATA USE FOR RESEARCH

An important argument illustrating the concerns and conflicts over informant public internet data use is that the researcher and university must be bound by traditional ethical research practices. The protection of informants and getting informed consent to use their data is a primary concern of ethical research practice. The arrival of the internet encouraged a 'the data is public' type attitude among researchers, as they used text and image data from the internet without asking those that produced them. Metcalf and Crawford (2016) claim an emerging ethics divide has arisen with discontinuities between researcher practices and ethical practices. It is proven that business and government use peoples' personal data for analysis. Academia, by contrast, has been challenged by a number of ethical issues with using large datasets, including privacy, identification and surveillance by outsiders' issues.

How to regulate and adhere to ethical research practices using informant internet data is increasingly difficult. Using a metaphor, data to researchers is oxygen as a growing share of our lives are dedicated to its release (Svensson and Guillen 2020; Jarzombek, 2016). Koenig (2018) believes individual data is one's own property that one should have total control over. Practically, consent cannot be obtained from every person who creates and posts data (Sinclair, 2017). Consent is often not obtained because large dataset creates a valuable digital image of human reality wanted by researchers for their research. As Zwitter (2014) argues, even in a sea of unidentified data, this lack of consent has a possible potential to cause harm through the potential identification of the creator of the data.

It is argued that internet data is free for taking because although internet users expect privacy for their public data, they freely and frequently post it in the public internet domain (Bay 2018; Fiesler and Nicholas 2018). This is factual, but because of mass-media reporting of data misuse, researchers and their universities are being forced to rethink established ethical principles of informed consent, privacy and anonymity (Sugiura, Wiles and Pope 2017). Despite difficulty in practicing this, views have been expressed by researchers that the public should be the arbiters of deciding if their data should be used in research without consent. Ainscough et al. (2018, p. 505) and McKee (2013) argue that deciding what is public and private data, even if that data is posted publicly online, can only really be defined by the subjects themselves "as they are the authors of their own words."

An ethical use of data example is the use of microblogging platform Twitter's data, which is problematic despite Twitter allowing researchers to use its members' data without informed consent. This problem leads to confusion and ambiguity

about what data and details of that data should or should not be used, as Ahmed, Bath and Demartini (2017) state:

A justification often provided by those working in an academic context with Twitter data with regard to the ethical and legal implications of using data without informed consent is that the reuse of data is permitted by Twitter's Terms and service as well as within the privacy policy. However, it is important to note that the act of scraping tweets or downloading tweets from Twitter's Advance Search will contravene Twitter's Terms and Conditions, therefore voiding any protection these policies are likely to offer (Ahmed, Bath and Demartini 2017, p. 7).

However, although studies using informant public data continue unabated, lessons have not always been learned from over a decade of data ethics research (Markham and Buchanan, 2012) as practices such as identified by Ahmed, Bath and Demartini (2017) continue.

The fast-paced growth of using online datasets is also not keeping pace with ethical boundaries. Although studies using informant public data continue unabated, lessons have not always been learned from decades of data ethics research. For example, problems with consent were previously recognized in studies of virtual communities, newsgroups, Internet Relay Chat and Usenet. Researchers lurked in these communities, sometimes known, but often anonymously, not informing people they were collecting data. If the researcher became known or informants later found out their posts were collected, the researcher was seen as an intruder potentially damaging the confidential community relationships built up over time (Eysenbach and Till 2001).

Early studies using internet data obtained from newsgroups and other sites would use data without informed consent. Ethical concerns of this practice began to be debated. Potential participant harm was no longer localized as the internet grew and data could be obtained from anyone, anywhere (Ess 2006). Researchers would employ pseudonyms to protect informants. The problem can be that using text examples verbatim, such as a tweet, can possibly be traced back to the informant. This may not be problematic to the person, but harm is possible, hence concerns about ethical data use became consistently expressed by researchers.

Informant protection from harm is still a central concern in all research. Data collected from the internet may produce a sense of detachment for the researcher. This can foster an attitude that using any internet data without informed consent is acceptable. That is not to imply researchers do not care about their informants as they are obliged, legally and morally, to protect the informants they study. To illustrate this, a study by Samuel, Derrick and van Leeuwen (2019) identified attitudes towards the use of social media data in research concluding:

This article argues that this governance of ethical behaviour by individual researchers perpetuates a negative cycle of academic practice that is dependent on subjective judgements by researchers themselves, rather than governed by more

formalised academic institutions such as the research ethics committee and funding council guidelines (Samuel, Derrick and van Leeuwen 2019, p 317).

Their further finding was that in their sample of 324 United Kingdom research articles, 234 articles did not mention ethical approval or discussions of ethical implications of using social media data (Samuel, Derrick and van Leeuwen 2019). The suggestion from this is that there is a moral belief that more transparency in conducting studies using informant data should be a mandatory requirement.

Managing and solving these conflicts and concerns have been made in academia, but also in some country's legal systems to protect informants' internet data. Samuel and Buchanan (2020, p. 8) proposed a framework of ethical issues that needed to be addressed before doing an internet-mediated research study. Their call for best practice procedures involves the researcher having high ethical standards and respect for individuals and communities. Governments have also responded to public concerns for data privacy. For example, The General Data Protection Regulation (EU) (GDPR), part of European Law, clearly sets out what personal data is. Although applicable to corporations and governments, the regulations states researchers must be careful with an individual's identification:

If you cannot directly identify an individual from that information, then you need to consider whether the individual is still identifiable. You should take into account the information you are processing together with all the means reasonably likely to be used by either you or any other person to identify that individual. (Information Commissioner's Office 2018, p. 10).

This illustrates that some countries are taking seriously the protection of peoples' public and private data, urging researchers to take seriously the identification of people who researchers take data from without informed consent.

3 CONFLICTS BETWEEN RESEARCHERS AND ETHICS BOARDS

Conflicting beliefs about informed consent and data use, and how to protect informants' data, still occur as standards set by external laws and organizations are often still unclear. The literature suggests ethics boards are still lacking in understanding public data use. However, they and other organizations supporting internet researchers' projects have been proactive in developing ethical conduct policies. In the literature ethics boards are blamed for their lack of clarity on this issue. Conflicts between internet researchers and ethics boards, who are either over cautious or unclear on how to approve internet data collection without harming informants, do occur.

In one study, Stevens, O'Donnell and Williams (2015) wanted to find out about personal experiences of those living with chronic illness. The data in their ethics application was disclosed as public Facebook posts. Commenting on Stevens, O'Donnell and Williams project, Ainscough et al. (2018) stated that the initial

rejection of the Stevens led project, which was later overturned, was because of the panel's uncertainty whether informed consent from informants was required. This was an example of the uncertainty that exists in approving online data research projects, what is public and private, and if informed consent is needed or not (Ainscough et al. 2018).

Several authors have commented on the contested and hostile relationships between researchers and ethics boards. Examples include: deceiving ethics boards to rush through approvals, collecting data without approval from the board and not reporting changes for fear of having restrictions placed on their research (Vitak et al. 2017; Feeley 2007; Keith-Spiegel, Koocher and Tabachnick 2006). Internet researchers may succumb to the temptation, or necessity in large datasets, to collect data without informed consent. However, ethical boards in approving projects must understand the difficult logistics of obtaining consent and consider the level of risk to participants. Researchers and ethics boards need to be empathetic towards each other and work co-operatively.

The Association of Internet Researchers (AoIR) is an academic association that has been proactive in attempting to develop worldwide guidelines in internet data use ethics. They claim that ethics boards are now becoming overly cautious approving research projects after public data use scandals were reported. Their strategy to remedy this is to develop guidelines to mediate ethical relationships between boards and researchers. franzke et al. (2020) state their aim to address this:

This points to one of the founding justifications for and central subsequent uses of the AoIR IRE guidelines namely, the importance of being able to help inform ethical review boards of the distinctive characteristics of internet research, as well as to provide researchers with resources, beginning with the guidelines themselves, to help them in the processes of negotiating the process of seeking approval for their research with such boards (franzke et al. 2020, p. 13).

Managing internet research data use ethical questions requires mutual agreement and constant communication between researchers, research students' academic advisors and university ethics boards. One question identified in the literature that hinders this process is if social media data is counted as a human subject research or if it is merely text research (Samuel and Buchanan 2020). The issues though of the misuse of such data, combined with peoples' concerns over this, move this issue into the realm of not being detached text only data, but being part of a person's data ownership where informed consent, however impractical with large datasets, should be a part of any research project.

Academic human ethics committees are struggling with understanding potential ethical issues. Researchers may become frustrated at justifying why such research warrants taking peoples' social media data for analysis. To investigate this further, a thematic analysis of the scholarly concerns and conflicts in this field is offered that contributes to this growing, ongoing debate. The question guiding the

analysis is, what are the priority scholarly concerns and conflicts that the ethical, or not, use of internet data are of concern to researchers and university ethics boards?

4 METHODS

An inductive thematic analysis method was used to examine these scholarly concerns and conflicts. An overarching goal is to identify and search for themes appearing as patterns in data (Braun and Clarke 2006). Thematic analysis aims to identify or examine the underlying ideas, assumptions, conceptualizations and ideologies that are theorized as shaping or informing the semantic content of the data (Braun and Clarke 2006). Being inductive and exploratory, no existing theory is used in the data analysis. The method, though generically theoretically flexible, is not atheoretical (Braun and Clarke 2019; Clarke and Braun 2018). This means an inductive approach can generate theories, or explanations, that give insights as to what is going on in an examined problem.

Data collecting was done through searching for peer-reviewed academic articles and journal editorials using academic research databases and Google Scholar. Search term examples included: ‘social media data’, ‘privacy’, ‘big data’, ‘participant risk’ and ‘ethics boards’. A total of 50 publications, between years 2001 to 2019, were chosen, 4 of which were refereed conference papers (see Appendix for publications chosen and their classification). What is distinguished is, the total of 50 publications is a small dataset. Inductive research can have reduced datasets for smaller projects such as this (Clarke, Braun and Hayfield 2015). The text lifted from the publications that was analyzed was though substantive with many sentences coded for possible theme development. It was assessed that the data was reliable because the articles were all peer-reviewed, increasing the validity and truthfulness of the authors’ text taken from the publications.

The analysis was conducted according to Clarke, Braun and Hayfield’s (2015) methods as:

1. Data familiarization requiring constant reading of the text
2. Generating an initial set of codes across the dataset
3. Searching for themes and sub themes from the codes
4. Reviewing themes including merging them for accuracy
5. Final theme naming and definitions of themes produced

Although appearing formulaic, thematic analysis is rigorous and time-consuming. Themes are important to search for and justify because they describe a phenomenon (Ferday and Muir-Cochrane 2006; Daly, Kellehear and Gliksman 1997). During the process, recognizing and encoding a moment in the text was done as this describes types of observations that can interpret the phenomenon (Ferday and Muir-Cochrane 2006; Boyatzis 1998). Another advantage is, despite the straightforward research problem presented in this article, thematic analysis allows

for unexpected results to be found that may provide insights previously unknown or unclear (Nowell et al. 2017; Côté and Turgeon, 2005). To assist with validity, a constant reviewing of codes and themes was done. The time to review themes also assisted in ensuring the themes represented explanations of scholarly concerns and conflicts about internet data.

5 THEMES AND ANALYSIS OF SCHOLARLY CONCERNS AND CONFLICTS

The themes from the analysis demonstrated the conflicts that have been present since the early period of gathering data off the internet for research. Being a public sphere, people have been concerned if their internet data is made public or their posts are identified and linked to them. The dataset provided instances of growing concerns and conflicts as internet technologies evolved, especially with the arrival of social media. The concerns from the public are represented by commentators like Koenig (2018) who still believe internet data is the property of the person who creates it. The themes show this is often not respected by researchers who may hold the attitude of it is public data there for the taking. Yet a constant theme, being far larger than this study's sample, suggests by force or willingness, researchers want to do no harm to their informants who provide data.

Three themes were created from the inductive study showing the types of scholarly concerns and conflicts. Risk, consent and practices, and the researcher's and ethics board's perception of how to manage them, were consistent patterns found in the text and make up the three themes. It needs to be noted that although not a factor in the analysis, the texts in the academic publications came from many countries. Even though they had different laws and policies, the consistency of this problem transcends geographic location.

Table 1 shows the themes that thematic analysis study yielded to explain researcher concerns and conflicts:

Table 1. Themes of scholarly concerns and conflicts in using informant internet data defined

Theme Name	Definition
Informant's Data Use Risks	Captures and identifies the individual and structural risks that are a consequence of using internet and social media data for research.
Research Consent and Practice	Gaining consent from informants in some way despite the massive datasets and maximizing informant protection.
Ethics Boards' Practices	Challenges that boards that approve researcher studies must contend with but may also clash with researchers over.

The three themes showed three sets of interest groups, researchers, informants and ethics boards, often in conflict with each other on how to manage informant internet data protection. Informant protection and harm minimization remains a pillar of research practice, but clearly using and managing large internet datasets that may have identifying factors is still not resolved. The study's themes are now discussed with example text to support them.

6 SCHOLARLY CONCERNS AND CONFLICTS THEMES

To support the themes, evidence illustrating each of them is given with text examples. The quotes displayed are from journal and conference articles that exist in some form in a public or semi-private domain (such as a journal issue behind a pay wall). Therefore, these quotes attributed to the author/s do not need consent to be used, but are referenced for academic integrity.

6.1 Informant's Data Use Risks

Authors were overall very concerned with informant risks and potential negative consequences that may occur from identification of the informant. This was especially evident with authors advising caution when using informant data extracts, for example tweets, in scholarly and other publications. A second risk concern was the sharing of internet datasets with other researchers or organizations, as well as concerns about data storage where problems such as cyber criminals hacking databases might occur.

Buchanan (2017) illustrates this problem seen across the dataset and also in the theme of data use risks. Ethical dilemmas that pose significant risks to informant identification are appearing regularly as research datasets grow in volume:

Ethics and methods are interdependent, and the rise of mass data mining across social media and the Internet has presented ethical dilemmas surrounding privacy, rights and autonomy, and such social justice issues as discrimination (Buchanan 2017, p. 2).

Ethics boards have a reputational and moral responsibility to ensure informant research risks are reduced and managed in line with high university standards. Consistent standards to protect informant internet data still seem to be in their infancy. In the articles' data, calls for standards that can be flexible to change in time as privacy laws, new global problems requiring examining of large datasets and technological advances occur was considered important. Kumar and Nanda (2019) illustrate this common problem:

Development of ethical standards for the social media is not just a theoretical exercise, but a carefully planned strategy assuming responsibility and active participation of both the individuals and organizations (Kumar and Nanda 2019, p. 67).

As a generalization, it was not clear in the data if ethics boards were solely to blame for not keeping up with the use of large internet datasets and potential issues for informants these may cause. It is inaccurate to conclude that many university ethics boards are not addressing and managing this issue correctly. What was identified is that increased informant data risks can have potential serious consequences to informant privacy if a person's name or other personal details are shown in a dataset or in a publicly available publication. It is this potential, however unlikely, that must be addressed when setting out boundaries for researchers in their projects to protect all informants, especially where no informed consent is given.

Jouhki et al. (2016) also stated a similar concern seen across the sampled literature. Even though researchers may mask informant details, people are concerned at possible identification, especially with sophisticated internet tools able to find even small items of information about someone:

Especially when informed consent cannot be obtained in human-subject research, the benefits of the study should outweigh the harm of any invasion of privacy (Jouhki et al. 2016, p. 77).

Therefore, this theme demonstrates that protecting informant's internet data is crucial to reducing risk. Despite researcher's eagerness to use large internet datasets, and the acceptance that informed consent is almost impossible to obtain, care needs to be taken to mask personal details. The researcher and ethics boards must agree to all safeguards to protect informants and reduce potential harm. But researchers must not be dismissive that just because peoples' details can be masked in large datasets that it is acceptable to take and use the public data because it is there.

6.2 Research Consent and Practice

This theme correlates with the other two themes because it involves a cornerstone of any research; getting an informant's consent to use their data. An informant should always feel free to withdraw that data at will. Again, with large datasets that is often not possible. Researchers often decide to use the data without consent, but also use it for other reasons, such as storing and sharing that data in the future, showing the difficulty of always practicing ethical consent procedures, as Nunan and Yenicioğlu (2013) illustrate this conflict:

On the one hand they promise that ownership of content remains in the hands of the individuals who posted it. On the other, they grant themselves permission to use the content for a wide range of purposes, purposes which are largely unspecified (Nunan and Yenicioğlu 2013, p. 797).

Another part of this theme was deciding what details and data examples to publish in academic and other research publications coming out of a study using informant internet data. Researchers have a conflict that it was not just consent to use data, but also what to publish in terms of informant's details. Across the sample a consistent pattern was to strongly recommend researchers to think carefully about

what data examples to publish. Researchers do have a strong moral obligation to be careful with what informant details they publish. Fiesler and Proferes (2018) recommend:

Therefore, we suggest that publication of user identity should only occur when the benefits of doing so clearly outweigh the potential harms, or with user permission (Fiesler and Proferes 2018, p. 797).

This finding did not imply though that not putting real names or other informant details in publications was unacceptable. In some studies, an informant may actually request their real name and other details be published, but the ethics board must be made aware the informant allows this to happen and makes the university exempt from any reputational or other damage.

An expected and reoccurring pattern in the data was that the authors reported how and why it was important to respect the informant's personal internet data, especially if no consent was obtained to use it. It is important to be considerate to not to do harm when using data to which there is no consent given. That part of the theme was recognized in many of the publications. Gupa (2017) states that respecting participants' data means, regardless of the size of the dataset, the participants still hold the right to not give consent for data use:

Consent is considered to be knowledgeable when individuals know the purpose of the information collected about them and their right to give, withhold, or withdraw the consent anytime they wish (Gupa 2017, p. 7).

How this can be practiced remains a problem that researchers and ethics boards need to consistently work on. Again, the concern arises that it is unlikely informants will be contacted to have their internet data used due to the dataset's volume. Yet they still retain the right to ask to have their data, even if it is part of an anonymous large dataset, to be not used.

6.3 Ethics Boards' Practices

In the data, the authors were critical of the role of university ethics boards in the boards not understanding how serious an ethical issue it has become to protect participants' internet data. Boundaries of what constitutes unethical use of the data are unclear. Gustafson and Woodworth (2014) commented on this conflict in their study of using large amounts of patient internet medical data:

...differing assumptions about the social boundaries of public and private communication that occurs in virtual communities; and ... the ethical dilemmas that emerge from these assumptions (Gustafson and Woodworth 2014, p. 5).

Though these conflicts are being addressed, ethics boards are also criticized by researchers as not keeping pace with fast developments and emerging new ethical issues. Ethics boards may approve research projects without considering a website

or social media platforms' terms and condition's changes or passing of new data protection privacy laws. Kinder-Kurlanda et al. (2017) argue the following:

Sharing legally and ethically also means to follow the changes and updates in terms of services and policies and to participate in negotiations about data sharing for the sake of reproducibility with platform providers (Kinder-Kurlanda et al. 2017, p. 11).

It can be questioned from this theme why should ethics boards be totally responsible for this? Researchers do have an obligation to disclose potential risks. Ethics boards though need to be aware of informant and study risks as using internet data grows. That is the key solution in addressing scholarly concerns and conflicts that is being addressed by ethics boards and organizations like the AoIR.

This theme also urged greater co-operation between researchers and ethics boards should be a priority. There is not a suggestion that all ethics boards are not doing this. The issue lies in keeping informants' privacy and free from harm requires much work to maintain. In the sample, researchers wrote of how crucial this is. Shilton and Sayles (2016) for example urge greater and constant co-operation between researchers and ethics boards at all stages of the research:

...this work suggests that ethics review boards (or alternative institutional structures) might best be positioned as consultants to research design, rather than post-hoc enforcement mechanisms (Shilton and Sayles 2016, p. 9).

Examining the three themes, the concerns and conflicts of researchers must be consistently addressed in all academic institutions worldwide, so informant internet data can be ethically used. Themes, and the examples given in this section, suggest this has been not always handled with care and responsibility by all involved. Concurrently, it is understandable though the rush to use internet data for research has not kept up with protections for informants. The question drawn from this analysis is, if these concerns and conflicts have been identified, what will occur at a global level in academic institutions to urgently remedy these issues?

7 DISCUSSION AND CONCLUSIONS

This analysis gave insights into the scholarly concerns and conflicts when informant internet data is used, the harm and risks its use brings, as well as advising that researchers and ethics boards must pay greater attention to this problem. What is clear is informant internet data will now be a major part of the research landscape. Informed consent cannot be obtained in large datasets. Yet despite this, ways of protecting informant's data, especially through cooperation and communication between researchers and ethics boards, is vital to managing informant harm minimization. Unethical conduct in using data without consent can result in researcher and university reputational harm, withholding of government research funding and journal article retractions.

The study identifies, but does not place sole responsibility on, the role of ethics boards in understanding the consequences of informant internet data use. Ethics boards may meet infrequently, but mechanisms and structures should be in place to monitor the use of internet data over the life of a research project. It also needs to be decided the level of acceptability of informant protection such as, what should appear in public and academic publications or if certain factors, such as sexuality, ethnic background, or medical conditions, warrant extra care in data collection. Laws to protect participant data or changes to a platform's terms in giving permission to use data will also constantly change. Theme three demonstrated that the urgency to remedy this issue is crucial if new laws limit researchers obtaining online data for research. That can only take place with cooperation between researchers and ethics boards.

A comment on the use of thematic analysis is warranted here to reflect on its usefulness in identifying and analyzing scholarly concerns and conflicts. The small set of 50 scholarly articles could have produced minimal results. However, the amount of text that informed the themes was large. Clarke, Braun and Hayfield (2015) though state that for a smaller project this amount of dataset can still generate codes and themes to answer a research question. What has arisen from doing thematic analysis though is that the patterns of concerns and conflicts were turned into themes that explain the problems occurring in the field of data ethics. It is feasible to suggest another researcher may yield different themes and explanations. Yet insights presented in this article did provide one way of examining what is a serious and ever-changing issue, where the protection of informants remains the important pillar that research studies rest on.

Protecting informants is as crucial as it ever was; more so now that privacy has been eroded by increasing levels of personal data published online. People will increasingly expect their data to be theirs to do as they wish with and decide how it will be used. They want to feel their data is safe. Relationships between researchers and ethics boards must not be adversarial. Although talking about informant protection from a commercial, as well as academic, position, Johanssen (2021, p. 101) is accurate in stating beneath the surface, this feeling of security is broken and users are denied mastery over their data and their destiny. This debate will continue necessitating an urgent need for cooperation in protecting informants but continue vital global and local research that the richness of detail of internet data can provide.

FUNDING STATEMENT AND ACKNOWLEDGMENTS

There was no funding received for this study. I would like to acknowledge my former university and the Department of Internet Studies, Curtin, Perth, Australia for giving me the idea to conduct this study into the issues of using public internet data.

REFERENCES

- Ahmed, W., Bath, P., & Demartini, G. (2017). Chapter 4 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 volume 2* (pp. 79-107). Emerald Publishing Limited.
- Ainscough, E., Smith, S. N., Greenwell, K., & Hoare, D. J. (2018). Findings and ethical considerations from a thematic analysis of threads within tinnitus online support groups. *American Journal of Audiology*, 27(3), 503-512. https://doi.org/10.1044/2018_AJA-IMIA3-18-0013
- Bay, M. (2018). Social media ethics: A Rawlsian approach to hypertargeting and psychometrics in political and commercial campaigns. *ACM Transactions on Social Computing*, 1(4), 1-14. <https://doi.org/10.1145/3281450>
- Boyatzis, R. (1998). Transforming qualitative information: Thematic analysis and code development. Sage.
- Boyd, D., & Crawford, K. (2012). Critical questions for big data. *Information, Communication and Society*, 15(5), 662-679. <https://doi.org/10.1080/1369118X.2012.678878>
- Braun, V., & Clarke, V. (2019). Reflecting on reflexive thematic analysis. *Qualitative Research in Sport, Exercise and Health*, 11(4), 589-597. <https://doi.org/10.1080/2159676X.2019.1628806>
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77-101. <https://doi.org/10.1191/1478088706qp063oa>
- Buchanan, E. (2017). Considering the ethics of big data research: A case of Twitter and ISIS/ISIL. *PLoS ONE*, 12(12), 1-6. <https://doi.org/10.1371/journal.pone.0187155>
- Clarke, V., & Braun, V. (2018). Using thematic analysis in counselling and psychotherapy research: A critical reflection. *Counselling and Psychotherapy Research*, 18(2), 107-110. <https://doi.org/10.1002/capr.12165>
- Clarke, V., Braun, V., & Hayfield, N. (2015). *Thematic analysis*. In J. Smith (Ed.), *Qualitative psychology: A practical guide to research methods* (3rd ed., pp. 222-248). Sage Publications.
- Côté, L., & Turgeon, J. (2005). Appraising qualitative research articles in medicine and medical education. *Medical Teacher*, 27(1), 71-75. <https://doi.org/10.1080/01421590400016308>
- Daly, J., Kellehear, A., & Gliksman, M. (1997). *The public health researcher: A methodological approach*. Oxford University Press.
- Ess, C. (2006). Internet research ethics. In Joinson, A., McKenna, K., Postmes, T., and Reips, U (Eds.), *Oxford handbook of psychology*. (pp. 487-503). Oxford University Press.
- Eysenbach, G., & Till, J. (2001). Ethical issues in qualitative research on internet communities. *BMJ Clinical Research*, 323(7321), 1103-1105.

- <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC59687/>.
<https://doi.org/10.1136/bmj.323.7321.1103>
- Feeley, M. (2007). Legality, social research, and the challenge of institutional review boards. *Law and Society Review*, 41, 757-776.
<https://doi.org/10.1111/j.1540-5893.2007.00322.x>
- Ferday, J., & Muir-Cochrane, E. (2006). Demonstrating rigor using thematic analysis: A hybrid approach of inductive and deductive coding and theme development. *International Journal of Qualitative Methods*, 5(1), 80-92.
<https://doi.org/10.1177/160940690600500107>
- Fiesler, C., & Proferes, N. (2018). “Participant perceptions” of Twitter research ethics. *Social Media + Society*, 4(1), 1-14.
<https://doi.org/10.1177/2056305118763366>
- Flick, C. (2016). Informed consent and the Facebook emotional manipulation study. *Research Ethics*, 12(1), 14-28.
<https://doi.org/10.1177/1747016115599568>
- Franzke, A., Bechmann, A., Zimmer, M., Ess, C., & the Association of Internet Researchers. (2020). *Internet research: Ethical guidelines 3.0*.
<https://aoir.org/reports/ethics3.pdf>
- Gupa, S. (2017). Ethical issues in designing internet-based research: Recommendations for good practice. *Journal of Research Practice*, 13(2), 1-14. <http://jrp.icaap.org/index.php/jrp/article/view/576/476>
- Gustafson, D., & Woodworth, C. (2014). Methodological and ethical issues in research using social media: A metamethod of Human Papillomavirus Vaccine Studies. *BMC Medical Research Methodology*, 14(127), 1-11.
<https://doi.org/10.1186/1471-2288-14-127>
- Hallinan, B., Brubaker, J., & Fiesler, C. (2020). Unexpected expectations: Public reaction to the Facebook emotional contagion study. *New Media and Society*, 22(6), 1076-1094. <https://doi.org/10.1177/1461444819876944>
- Hand, D. (2018). Aspects of data ethics in a changing world: Where are we now? *Big Data*, 6(3), 176-190. <https://doi.org/10.1089/big.2018.0083>
- Information Commissioner’s Office. (2018). *Guide to the general data protection regulation (GDPR)*. <https://ico.org.uk/media/for-organisations/guide-to-the-general-data-protection-regulation-gdpr-1-0.pdf>
- Jarzombek, M. (2016). *Digital Stockholm Syndrome in the post-ontological age*. University of Minnesota Press.
- Johanssen, J. (2021). Data perversion: A psychoanalytic perspective on datafication. *Journal of Digital Social Research*, 3(1), 88-105.
<https://www.jdsr.io/articles/2021/3/26/data-perversion-a-psychoanalytic-perspective-on-datafication>. <https://doi.org/10.33621/jdsr.v3i1.57>
- Jouhki, J., Lauk, E., Penttinen, M., Sormanen, N., & Uskali, T. (2016). Facebook’s emotional contagion experiment as a challenge to research ethics. *Media and Communication*, 4(4), 75-85.
<https://doi.org/10.17645/mac.v4i4.579>

- Keith-Spiegel, P., Koocher, G., & Tabachnick, B. (2006). What scientists want from their research ethics committee. *Journal of Empirical Research on Human Research Ethics*, 1, 67-81. <https://doi.org/10.1525/jer.2006.1.1.67>
- Kinder-Kurlanda, K., Weller, K., Zenk-Möltgen, W., Pfeffer, J., & Morstatter, F. (2017). Archiving information from geotagged tweets to promote reproducibility and comparability in social media research. *Big Data & Society*, 4(2), 1-14. <https://doi.org/10.1177/2053951717736336>
- Koenig, G. (2018). Introduction: The Softian bargain. In I. Landreau, G. Peliks, N. Binctin, V. Pez-Pérard & L. Léger (Eds), *My data are mine: Why we should have ownership rights on our personal data* (10-17). Paris: GenerationLibre. <https://www.generationlibre.eu/wp-content/uploads/2018/01/Rapport-Data-2018-EN-v2.pdf>
- Kramer, A., Guillory, J., & Hancock, J. (2014). Experimental evidence of massive-scale emotional contagion through social networks. *PNAS*, 111(24), 8788-8790. <https://doi.org/10.1073/pnas.1320040111>
- Kumar, V., & Nanda, P. (2019). Social media to social media analytics: Ethical challenges. *International Journal of Technoethics (IJT)*, 10(2), 57-70. <https://doi.org/10.4018/IJT.2019070104>
- McKee, R. (2013). Ethical issues in using social media for health and health care research. *Health Policy*, 110(2-3), 298-301. <https://doi.org/10.1016/j.healthpol.2013.02.006>
- Markham, A., & Buchanan, E. (2012). Ethical decision-making and internet research: Recommendations from the AoIR Ethics Working Committee (Version 2.0). <https://www.aoir.org/reports/ethics2.pdf>
- Metcalfe, J., & Crawford, K. (2016). Where are human subjects in Big Data research? The emerging ethics divide. *Big Data & Society*, 3(1), 1-14. <https://doi.org/10.1177/2053951716650211>
- Nowell, L., Norris, J., White, D., & Moules, N. (2017). Thematic analysis: Striving to meet the trustworthiness criteria. *International Journal of Qualitative Methods*, 16(1), 1-13. <https://doi.org/10.1177/1609406917733847>
- Nunan, D., & Yencioğlu, V. (2013). Informed, uninformed and participative consent in social media research. *International Journal of Market Research*, 55(6), 791-808. <https://doi.org/10.2501/IJMR-2013-067>
- Samuel, G., & Buchanan, E. (2020). Guest editorial: Ethical issues in social media research. *Journal of Empirical Research on Human Research Ethics*, 15(1-2), 3-11. <https://doi.org/10.1177/1556264619901215>
- Samuel, G., Derrick, G., & van Leeuwen, T. (2019). The ethics ecosystem: Personal ethics, network governance and regulating actors governing the use of social media research data. *Minerva*, 57, 317-343. <https://doi.org/10.1007/s11024-019-09368-3>

- Sinclair, M. (2017). Using social media for research. *Evidence Based Midwifery*, 15(2), 39. Available at: <https://www.rcm.org.uk/media/2787/evidence-based-midwifery-june-2017.pdf>
- Stevens, G., O'Donnell, V., & Williams, L. (2015). Public domain or private data? Developing an ethical approach to social media research in an interdisciplinary project. *Educational Research and Evaluation*, 21(2), 154-167. <https://doi.org/10.1080/13803611.2015.1024010>
- Sugiura, L., Wiles, R., & Pope, C. (2017). Ethical challenges in online research: Public/private perceptions. *Research Ethics*, 13(3-4), 184-199. <https://doi.org/10.1177/1747016116650720>
- Svensson, J., & Guillen, O. P. (2020) What is data and what can it be used for? Key questions in the age of burgeoning data essentialism. *Journal of Digital Social Research*, 2(3), 65-83. <https://jdsr.se/ojs/index.php/jdsr/article/view/40/27>. <https://doi.org/10.33621/jdsr.v2i3.40>
- Vitak, J., Proferes, N., Shilton, K., & Ashktorab, Z. (2017). Ethics regulation in social computing research: Examining the role of institutional review boards. *Journal of Empirical Research on Human Research Ethics*, 12(5), 372-382. <https://doi.org/10.1177/1556264617725200>
- 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>
- Zwitter, A. (2014). Big data ethics. *Big Data & Society*, 1(2), 1-6. <https://doi.org/10.1177/2053951714559253>

APPENDIX

Dataset List: Journal Articles, Conference Papers and Journal Introductions

The following 50 academic articles were used in the dataset. The majority are peer-reviewed studies in academic journals. Also included were some conference papers and introduction editorial sections to journals. They are grouped according to type. Some publications listed in this appendix also appear in the reference list as they were used in data examples and in the body of this article.

Journal Articles - Empirical Studies and Editorials

- Adams, S., Veghel, D., & Dekker, L. (2015). Developing a research agenda on ethical issues related to using social media in healthcare. *Cambridge Quarterly of Healthcare Ethics*, 24, 293-302.
<https://doi.org/10.1017/S0963180114000619>
- Andryani, R., Negara, E., & Triadi, D. (2019). Social media analytics: Data utilization of social media for research. *Journal of Information Systems and Informatics*, 1(2), <https://doi.org/10.33557/journalisi.v1i2.23>
- Bay, M. (2018). Social media ethics: A Rawlsian approach to hypertargeting and psychometrics in political and commercial campaigns. *ACM Transactions on Social Computing*, 1(4), 1-14. <https://doi.org/10.1145/3281450>
- Berry, D. (2004). Internet research: privacy, ethics and alienation: An open source approach. *Internet Research*, 14(4), 323-332.
<https://doi.org/10.1108/10662240410555333>
- Boyd, D., & Crawford, K. (2012). Critical questions for big data. *Information, Communication and Society*, 15(5), 662-679.
<https://doi.org/10.1080/1369118X.2012.678878>
- Buchanan, E. (2017). Considering the ethics of big data research: A case of Twitter and ISIS/ISIL. *PLoS ONE*, 12(12), 1-6.
<https://doi.org/10.1371/journal.pone.0187155>
- Cesare, N., Lee, H., McCormack, T., Spiro, E., & Zagheni, E. (2018). Promises and pitfalls of using digital traces for demographic research. *Demography*, 55(5): 1979-1999. <https://doi.org/10.1007/s13524-018-0715-2>
- Clark, K., Duckham, M., Guillemain, M., Hunter, A., McVernon, J., O'Keefe, C., Pitkin, C., Praver, S., Sinnott, R., Warr, D., & Waycott, J. (2019). Advancing the ethical use of digital data in human research: challenges and strategies to promote ethical practice. *Ethics and Information Technology*, 21, 59-73. <https://doi.org/10.1007/s10676-018-9490-4>
- Côté, I. (2013). Fieldwork in the era of social media: Opportunities and challenges. *Assessment in Political Science*, 46(3), 615-619.
<https://doi.org/10.1017/S1049096513000565>

- Crawford, K., & Finn, M. (2015). The limits of crisis data: analytical and ethical challenges of using social and mobile data to understand disasters. *GeoJournal*, 80, 491-502. <https://doi.org/10.1007/s10708-014-9597-z>
- Eysenbach, G., & Till, J. (2001). Ethical issues in qualitative research on internet communities. *BMJ Clinical Research*, 323(7321), 1103-1105. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC59687/>. <https://doi.org/10.1136/bmj.323.7321.1103>
- Fiesler, C. (2019). Ethical considerations for research involving (speculative) public data. *Proceedings of the ACM on Human-Computer Interaction*, 3(249), 1-13. <https://doi.org/10.1145/3370271>
- Fiesler, C., & Nicholas, N. (2018). “Participant perceptions” of Twitter research ethics. *Social Media + Society*, 4(1), 1-14. <https://doi.org/10.1177/2056305118763366>
- Flick, C. (2016). Informed consent and the Facebook emotional manipulation study. *Research Ethics*, 12(1), 14-28. <https://doi.org/10.1177/1747016115599568>
- Golder, S., Ahmed, S., Norman, G., & Booth, A. (2017). Attitudes toward the ethics of research using social media: A systematic review. *Journal of Medical Internet Research*, 19(6), e195. <https://www.jmir.org/2017/6/e195/>. <https://doi.org/10.2196/jmir.7082>
- Gupa, S. (2017). Ethical issues in designing internet-based research: Recommendations for good practice. *Journal of Research Practice*, 13(2), 1-14. <http://jrp.icaap.org/index.php/jrp/article/view/576/476>
- Gustafson, D., & Woodworth, C. (2014). Methodological and ethical issues in research using social media: A metamethod of Human Papillomavirus vaccine studies. *BMC Medical Research Methodology*, 14(127), 1-11. <https://doi.org/10.1186/1471-2288-14-127>
- Hesse, A., Glenna, L., Hinrichs, C., Chiles, R., & Sachs, C. (2019). Qualitative research ethics in the big data era. *American Behavioral Scientist*, 63(5), 560-583. <https://doi.org/10.1177/0002764218805806>
- Jackson, D., Waine, M., & Hutchinson, M. (2015). Blogs as a way to elicit feedback on research and engage stakeholders. *Nurse Researcher*, 22(3), 41-47. <https://doi.org/10.7748/nr.22.3.41.e1300>
- Jouhki, J., Lauk, E., Penttinen, M., Sormanen, N., & Uskali, T. (2016). Facebook’s emotional contagion experiment as a challenge to research ethics. *Media and Communication*, 4(4), 75-85. <https://doi.org/10.17645/mac.v4i4.579>
- Kennedy, H., & Moss, G. (2015). Known or knowing publics? Social media data mining and the question of public agency. *Big Data & Society*, 2(2), 1-11. <https://doi.org/10.1177/2053951715611145>
- Kennedy, H., Elgesem, D., & Miguel, C. (2015). On fairness: User perspectives on social media data mining. *Convergence: The International Journal of*

- Research into New Media Technologies*, 23(3), 270-288.
<https://doi.org/10.1177/1354856515592507>
- Kennedy, H., & Moss, G. (2015). Known or knowing publics? Social media data mining and the question of public agency. *Big Data & Society*, 2(2), 1-11.
<https://doi.org/10.1177/2053951715611145>
- Kinder-Kurlanda, K., Weller, K., Zenk-Möltgen, W., Pfeffer, J., & Morstatter, F. (2017). Archiving information from geotagged tweets to promote reproducibility and comparability in social media research. *Big Data & Society*, 4(2), 1-14. <https://doi.org/10.1177/2053951717736336>
- Krotoski, A. (2012). Data-driven research: open data opportunities for growing knowledge, and ethical issues that arise. *Insights*, 24(1), 28-32. Available at: <http://eprints.lse.ac.uk/42923/>
- Kumar, V., & Nanda, P. (2019). Social media to social media analytics: Ethical challenges. *International Journal of Technoethics (IJT)*, 10(2), 57-70.
<https://doi.org/10.4018/IJT.2019070104>
- Lunnay, B., Borlagdan, J., McNaughton, D., & Ward, P. (2014). Ethical use of social media to facilitate qualitative research. *Qualitative Health Research*, 25(1), 99-109. <https://doi.org/10.1177/1049732314549031>
- Mackay, H. (2017). Social media analytics implications for journalism and democracy. *Journal of Information Ethics*, 26(1), 34-48.
- Mannheimer, S., & Hull, E. (2017). Sharing selves: Developing an ethical framework for curating social media data. *International Journal of Digital Curation*, 12(2), 196-209. <http://www.ijdc.net/article/view/518/494>.
<https://doi.org/10.2218/ijdc.v12i2.518>
- Moreno, M., Goni, N., Moreno, P., & Diekema, D. (2013). Ethics of social media research: Common concerns and practical considerations. *Cyberpsychology, Behavior and Social Networking*, 16(9), 708-713.
<https://doi.org/10.1089/cyber.2012.0334>
- Nunan, D., & Yencioğlu, V. (2013). Informed, uninformed and participative consent in social media research. *International Journal of Market Research*, 55(6), 791-808. <https://doi.org/10.2501/IJMR-2013-067>
- Nunan, D., & Domenico, M. (2013). Market research & the ethics of big data. *International Journal of Market Research*, 55(4), 505-520.
<https://doi.org/10.2501/IJMR-2013-015>
- Powell, T. (2019). Can quantitative research solve social problems? Pragmatism and the Ethics of Social Research. *Journal of Business Ethics*, 167, 1-8.
<https://doi.org/10.1007/s10551-019-04196-7>
- Samuel, G., & Buchanan, E. (2020). Guest editorial: Ethical issues in social media research. *Journal of Empirical Research on Human Research Ethics*, 15(1-2), 3-11. <https://doi.org/10.1177/1556264619901215>
- Samuel, G., Derrick, G., & van Leeuwen, T. (2019). The ethics ecosystem: Personal ethics, network governance and regulating actors governing the use

- of social media research data. *Minerva*, 57, 317-343.
<https://doi.org/10.1007/s11024-019-09368-3>
- Shaw, D. (2016). Facebook's flawed emotion experiment: Antisocial research on social network users. *Research Ethics*, 12(1), 29-34.
<https://doi.org/10.1177/1747016115579535>
- Sinclair, M. (2017). Evidence based midwifery. *Evidence Based Midwifery*, 15(2), 39. <https://www.rcm.org.uk/media/2787/evidence-based-midwifery-june-2017.pdf>
- Sormanen, N., & Lauk, E. (2016). Issues of ethics and methods in studying social media. *Media and Communication*, 4(4), 63-65.
<https://doi.org/10.17645/mac.v4i4.793>
- Stevens, G., O'Donnell, V., & Williams, L. (2015). Public domain or private data? Developing an ethical approach to social media research in an interdisciplinary project. *Educational Research and Evaluation*, 21(2), 154-167.
<https://doi.org/10.1080/13803611.2015.1024010>
- Taylor, J., & Pagliari, C. (2018). Comprehensive scoping review of health research using social media data. *British Medical Journal*, 8, 1-8. Available at: <https://bmjopen.bmj.com/content/8/12/e022931>. 10.1136/bmjopen-2018-022931
- Taylor, J., & Pagliari, C. (2018). Mining social media data: How are research sponsors and researchers addressing the ethical challenges? *Research Ethics*, 14(2), 1-39. <https://doi.org/10.1177/1747016117738559>
- Tinati, R., Halford, S., Carr, L., & Pope, C. (2014). Big data: Methodological challenges and approaches for sociological analysis. *Sociology*, 48(4), 663-681. <https://doi.org/10.1177/0038038513511561>
- Varley-Winter, O., & Shah, H. (2016). The opportunities and ethics of big data: practical priorities for a national Council of Data Ethics. *Philosophical Transactions: Mathematical, Physical and Engineering Sciences*, 374(2083), 1-7. <https://doi.org/10.1098/rsta.2016.0116>
- Warrell, J., & Jacobson, M. (2014). Internet research ethics and the policy gap for ethical practice in online research settings. *Canadian Journal of Higher Education*, 44(1), 22-37. <https://doi.org/10.47678/cjhe.v44i1.2594>
- Zimmer, M. (2010). "But the data is already public": On the ethics of research in Facebook. *Ethics and Information Technology*, 12, 313-325.
<https://doi.org/10.1007/s10676-010-9227-5>
- Zwitter, A. (2014). Big Data ethics. *Big Data & Society*, 1(2), 1-6.
<https://doi.org/10.1177/2053951714559253>

Conference Papers

- Bowser, A., & Tasi, J. (2015, May 2015). *Supporting ethical web research: A new research ethics review* (Paper presentation). WWW '15: Proceedings of the 24th International Conference on World Wide Web May 2015, pages 151-161, Florence, Italy. <https://doi.org/10.1145/2736277.2741654>

- Chancellor, S., Birnbaum, M., Caine, E., Silenzio, V., & Choudhury, M. (2019, January 29). *A taxonomy of ethical tensions in inferring mental health states from social media* (Paper presentation). Conference on Fairness, Accountability, and Transparency (FAT* '19) (79-88), January 2019, ACM, New York, United States. <https://doi.org/10.1145/3287560.3287587>
- Hutton, L., & Henderson, T. (2015, August 21). *Some challenges for ethics in social network research* (Paper presentation). NS Ethics' 15: NS Ethics '15: Proceedings of the 2015 ACM SIGCOMM Workshop on Ethics in Networked Systems Research, (13-15), London, England. <https://doi.org/10.1145/2793013.2793023>
- Shilton, K., & Sayles, S. (2016, January 5-8). "We aren't all going to be on the same page about ethics:" Ethical practices and challenges in research on digital and social media (Paper presentation). The 49th Hawaii International Conference on System Sciences (HICSS), (1909-1918), Kōloa, Hawaii. <https://doi.org/10.1109/HICSS.2016.242>