Northumbria Research Link

Northumbria
University
NEWCASTLE

University**Library**

Title: Using Twitter as a data source: An overview of ethical, legal and methodological challenges

Book Chapter: '*Advances in Research Ethics and Integrity' published by Emerald Books.*

Authors: Wasim Ahmed[1], Peter A. Bath[1], Gianluca Demartini[2]

[1] Health Informatics Research Group Information School, University of Sheffield

[2] Information Retrieval Research Group, Information School, University of Sheffield

Running Head: Twitter as a data source: ethical challenges

Abstract

Purpose: This chapter provides an overview of the specific legal, ethical, and privacy issues that can arise when conducting research using Twitter data.

Approach: We review existing literature to inform those whom may be undertaking social media research. We also present a number of industry and academic case studies in order to highlight the challenges that may arise in research projects using social media data. Finally, the chapter provides an overview of the process that was followed to gain ethics approval for a PhD project using Twitter as a primary source of data.

Practical Implications: By outlining a number of Twitter-specific research case studies the chapter is a valuable resource to those considering the ethical implications of their own research projects utilizing social media data. Moreover, the chapter outlines existing work looking at the ethical practicalities of social media data, and relates their applicability to researching Twitter.

Value: This chapter is a potentially useful resource to those conducting social media research, or those who wish to gain an understanding of the specific legal, ethical, and privacy issues that can face social media researchers.

Keywords: Ethics, Privacy, Twitter, Social Media, Research Integrity

Background

The use of social media and social networking websites has increased rapidly in recent years with more households, organisations and individuals having access to the Internet (OECD, 2016). There are more social media platforms, and more members of the public, businesses, charitable and other organisations that are using these platforms (Chaffey, 2016). Online interaction, therefore, is now a regular part of daily life for a demographically diverse population of billons of people worldwide (Golder and Macy, 2014). Those who use social media may post their thoughts, feelings, and/ or opinions on almost every aspect of life (Chew and Eysenbach, 2010). Social media content, therefore, presents academic researchers with important new opportunities to study a range of topics in a naturally occurring setting. There are a number of ethical issues associated with undertaking this research, which will be discussed in this chapter; however, there are enormous benefits that can be derived from this research, in understanding what, and how, people communicate in particular situations.

We have a special interest in how Twitter is used by citizens during extreme circumstances, and the authors of this chapter form part of a PhD supervisory team which looks at the use of Twitter during Infectious disease outbreaks. The overall aim of the PhD project is to gain a better understanding of the types of information that was shared on Twitter during the 2009 Swine Flu outbreak and the 2014 outbreak. This comparison will allow for the two outbreaks to be compared in regards to any similarities, differences, and trends in how Twitter users respond to infectious disease outbreaks. As the PhD project makes use of Twitter data we are able to report on the ethical challenges, and wider methodological issues that were faced across the conception and design of the study.

Social media are changing the way people communicate, both in their day-to-day lives, but also during extreme circumstances, for example, disasters that may threaten individuals, groups of people and overall public health in local and regional areas (Merchant, Elmer, and Lurie, 2011). Merchant, Elmer, and Lurie (2011) suggested that engaging with, and using, social media platforms such as Twitter may place the emergency-management community in a better position to be able to respond to emerging disasters. As the use of social media has changed the way in which people communicate (Cameron, Power, Robinson, & Yin, 2012), e.g., during emergency situations, information is now available from the public and it can be used to inform the situational awareness of emergencies and to help crises coordinators respond appropriately.

Research on Twitter ranges from analysing tweets related to riots (Procter, Vis, & Voss, 2013), natural disasters (Mendoza, Poblete, & Castillo, 2010; Lachlan, Spence, Lin, Najarian & Greco, 2015), and crisis events (Gupta, Joshi, & Kumaraguru, 2012; Simon, Goldberg, Aharonson-Daniel, Leykin, & Adini, 2014). The studies in relation to natural disasters have found that Twitter offers a decisive channel of communication between government, emergency responders and the public during crises (Cameron, Power, Robinson, & Yin, 2012; Simon, Goldberg, Aharonson-Daniel, Leykin, & Adini, 2014).

Although these new information sources will not replace existing sources of information, they can provide a new source of data that potentially could have many applications within emergency management and crisis coordination. Social media can play a role in the pre-incident activity, near real-time notification of an incident occurring, first hand reports of the impact of an incident occurring, and gauging the community responses to emergency warnings (Merchant, Elmer, & Lurie, 2011).

Twitter has become a datasource which can be utilized by emergency services during disasters (Tomer, Avishay, and Bruria, 2015). Twitter data, in comparison with other social media platforms such as Facebook, are more openly accessible and, for a proportion of tweets, can contain valuable meta-data, including geospatial data, such as the precise latitude and longitude co-ordinates from which a Tweet was posted. These data can be used to provide important aid to those in need during a natural disaster. Moreover, it is now also possible to use these data to monitor political events, disasters, health problems in real time and provide support to people in the location and at the time it is most needed. However, the immediacy of social media research can potentially place participants at risk of greater harm after an event has occurred, because of the potential to link Tweets to specific geolocations and individuals. Not all Twitter users are aware that all their posts are public, or that they are available for analysis and scrutiny. Tweets and posts may be quoted by newspapers and other news organisations soon after the occurrence of an event and potentially linked back to individuals. In addition, data from Twitter, and other platforms, are being used by academic researchers to develop a better understanding of how people are using social media in specific circumstances. The issue around whether tweets are public is open to debate, and will be further discussed in this chapter.

There are legal and ethical implications to using social media data posted by people who may have been sending tweets while in a vulnerable state of mind, e.g., during a disaster, or health outbreak. For instance, someone tweeting during an emergency may not necessarily realise that their tweet may be being collected and analysed, either to help co-ordinate relief activities, or to be reported in in a research article. Therefore, it is important for those undertaking social media research to critically reflect on the possible implications of a research project involving social media data to the persons involved in creating or being mentioned in such content.

Instagram, Facebook, and Twitter are among some the most popular places online interactions take place (Chaffey, 2016; Macy, Mejova, & Weber, 2015). Though, Twitter is one of the most researched platforms in regards to academic research (Weller, and Kinder-Kurlanda, 2015)

In terms of the overall structure of this chapter, we first look at the popularity for using Twitter within a research context, and some of the key issues that can arise in social media research. Previous work by NatCen and Ipsos Mori is, will be outlined and related specifically to Twitter. The chapter then outlines a number of academic and industry case studies and highlights issues that can arise in research which uses Twitter data. Finally, the steps and arguments that were made in order to obtain research ethics approval for a PhD project which is using Twitter as a primary source of data are outlined as well as some of the issues that have arisen as the project is underway.

Twitter

Twitter reports having 316 million monthly active users, there being 500 million tweets posted per day, and 80% of active Twitter users use a mobile device (About Twitter, n.d.). Tweets contain a wealth of data, and mining this data can provide insight into public opinion and behaviour responses in particular situations (Chew and Eysenbach, 2010). Twitter was described by Purohit et al (2013) as a microblogging platform that acts as a medium for the flow of information where users can post updates and subscribe to other users, known as 'following', in order to receive updates or microblogs from other users.

It is important to understand the features of a social media platform fully before a research project commences, or is even considered, as these features may have ethical implications that should be considered. For example, although, as we indicated in the previous section, people may not be fully aware that their Tweets are publicly viewable, some researchers (Townsend & Wallace, 2016) argue that if a tweet contains a hashtag, then the user tweeting this has intended for their tweet to be visible to a broader audience, and therefore informed consent is not necessary when reproducing the tweet in an academic article. Purohit et al (2013) described the key features of Twitter:

- A tweet is a short message also known as a post, status or microblog from a user on Twitter and which consists of a <140 characters, these tweets may contain updates about user activities, or share useful information.

- Tweets can contain links to web-pages, blogs etc., and, to avoid lengthy URLs, Twitter users will use condensed versions of URLs which are shortened by external services such as http://bit.ly/

- A hashtag, is denoted by a word preceding with the '#' symbol, (e.g., #EbolaOutbreakAlert). The hashtag is a platform convention for user-defined topics, and which was intended to identify a topic of communication, e.g. #Brexit

- The reply feature is platform provided to communicate with the author of a tweet by clicking on Twitter's 'Reply' button in response to a tweet.

- The retweet function forwards a tweet from a user to their followers and this is similar to forwarding an email to one's email contacts, for example. The 'mention' feature acknowledges a user with the symbolic '@' sign, but this does not use the reply platform feature e.g., 'Thanks @userhandle.

- A new feature implemented after the paper by Purohit et al (2013) was published is that Twitter allows users to retweet with a comment. Users can now quote a tweet and attach a comment to it e.g., users tweet '[Original tweet]' as @userhandle I agree [@userhandle1 today is a good day]

- A trend also known as 'trending' on Twitter refers to when a topic (a keyword or hashtag) is popular at a specific time. Twitter provides a list of topics that are currently trending for users, based on the frequency of particular hashtag.

There are a number of existing software applications that researchers can use to retrieve data from Twitter such as NodeXL (Smith, Milic-Frayling, Shneiderman, Mendes Rodrigues, 2010) or Discover Text (n.d.). There are also a number of software applications that are available at no cost such as Mozdeh (n.d.), Chorus (n.d.), or TAGs (n.d.). Data from Twitter can be retrieved at either no cost via the Search *Application Programming Interface (API)*, or at a fee via the Firehose *Application Programming Interface (API)*. The Search API is where the majority of Twitter research has focused, and for those undertaking social media research it is important to have some understanding of APIs. In simple terms, APIs are "…sets of requirements that govern how one application can talk to another" and they govern how applications can talk to one another by "'exposing' a program's internal functions to the outside world in a limited fashion. This makes it possible for applications to share data and take actions on one another's behalf without requiring developers to share all of their software's code"

(*What APIs Are And Why They're Important*, 2013). So for Twitter, this means that it will allow members of the public to create tools that can be used to download data. This effectively enables anyone with an Internet connection the ability to obtain Twitter data. The difference between using a software application that retrieves data for free via the Search API is that the data will not be a complete record of tweets, whereas via a paid API such as the Firehose API, researchers will have a complete record of tweets. The differences, therefore, relate to the amount of data that is retrieved as the Firehose API retrieves almost a complete record of tweets whereas the Search API provides a sample of tweets.

Twitter is known to attract more research in comparison to other social media platforms such as Facebook. The next section explores some possible reasons for the popularity of Twitter and compares these to that of Facebook.

Popularity of Twitter for Research

Researchers among the New Social Media New Social Science (NSMNSS), an online peer led community, have hypothesized the reasons for why Twitter has attracted more academic research compared with other social media platforms. The NSMNSS network is an online peer led community which was established in 2011 with a small grant from National Centre for Research Ethics (NCRM) in order to provide a space for reflective discussions about how working with new forms of data, including social media data, was likely to challenge conventional approaches to social science research (Woodfield et al, 2013).

There are at least five possible reasons for the popularity of Twitter in academic research (Ahmed, 2015a):

1. The Twitter API is more open and accessible compared with other social media platforms. This makes Twitter more favourable to developers creating tools to access data. This consequently increases the availability of software and online tools to researchers. Facebook data, in comparison, are very difficult to obtain, and are only available on an aggregate level for marketing purposes.
2. Twitter makes it easier to find and follow conversations as Twitter has a search feature which allows users to look up tweets, and tweets also appear within Google search results, which makes it easier to locate tweets. Facebook can be considered more of a private platform and not all public posts appear in Google Search results. Facebook also provides users with more privacy controls.
3. Twitter has a strong hashtag culture which makes it easier gathering, sorting, and expanding searches when collecting data. Therefore, Twitter data is easier to retrieve as major incidents, news stories and events on Twitter tend to be centred on a hashtag. Facebook does have hashtag capability, however, the use of hashtags does not appear to be as widespread as on Twitter.
4. Twitter may be a popular platform due to the attention it can receive from the mainstream media, and can attract more research due to its cultural status. Twitter is also widely used by journalists, both to identify newsworthy events as well as to distribute breaking news. In comparison with Facebook, it could be argued that Twitter receives much more media attention because celebrities, politicians, and sports starts often tweet about current events, and some tweets may be controversial and are therefore reported in the news.

5. Many researchers use Twitter themselves and, due to their favourable personal experiences, they may feel more comfortable when they research a more familiar platform. Facebook may be used by academics, however, it is rarely used during conferences in comparison with Twitter.

This list is not intended to be exhaustive but to offer a suggestion of some of the reasons Twitter has surged in popularity.

Key Ethical Issues for Social Media

In this section we explore some of the key issues within social media research, such as whether social media spaces are private or public spaces, and will explore some of the challenges of obtaining informed consent on Twitter. For example, questions may arise such as: How much weight should assign to the views of social media users as they may not fully comprehend the Terms and Conditions of a social media platform? Is ignorance really a justification for a researcher to override the privacy rights of a user?

Public vs Private Spaces

The British Psychological Society (2013) in their ethics guidelines for Internet-mediated Research have written that:

> "In an IMR context, the distinction between public and private space becomes increasingly blurred, however. For one thing, much internet communication is conducted in both a private (.e.g., the home) and public (.e.g., open discussion forum) location simultaneously. Secondly, in this new medium it is not always easy to determine which online spaces people perceive as 'private' or 'public'; where they might be happy to be observed, or otherwise. To complicate things further, a communication perceived as private at the time (e.g., a posting to a password-protected online discussion group) may become public at a much later date, should the archived information become publicly accessible (as has happened on occasion in the past) (British Psychological Association, 2013.pp.6-7).

The passage above highlights the lack of clarity over whether an online space is public or private. Certain social media platforms are seen as inherently private spaces, for instance, Facebook whereas others are seen as public spaces for online communication to take place, for example, Twitter. It is important to note that a key difference between platforms, such as Facebook and Twitter, is that most content that is shared on Twitter is publicly accessible via the Twitter API and/or via data resellers, whereas the majority of Facebook is considered private and that data from Facebook are normally only made available at an aggregate level. The public and private distinctions are important when researching online spaces as in addition to ethical implications the public may react negatively if they feel that researchers are intruding on their privacy. It is also important to note that Twitter profiles and tweets are, by default, set to public visibility and, consequently, Twitter could be considered more of a public space compared to Facebook. However, the extent to which individual users of Twitter are aware of this or moderate their behaviour on Twitter to account for this is debatable.

Informed Consent

Traditional conceptions of informed consent may be challenged when using social media data. For example, in our own project looking at infectious disease outbreaks on Twitter, it has not been possible to obtain informed content from users. When researching social media platforms, such as

Twitter, researchers may be working with large datasets, in which it would be difficult to obtain informed consent from all users that are part of a dataset. Additionally, it may not be possible to reach Twitter users, because when they are approached for informed consent via a tweet or a direct message, they simply may not reply or may no longer be maintaining their account. Some of the earliest health-based research on Twitter (Chew & Eysenbach, 2010) analysed tweets at an aggregate level, due to the difficulty of obtaining informed consent. However, it is important to consider thoroughly possible issues around consent as a study on Facebook (Kramer, Guillory, and Hancock, 2014) was criticised for a lack of informed consent (Arthur, 2014; Panger, 2014).

In our own PhD project we found that it would be very difficult to obtain informed consent from Twitter users to use their tweets. For instance in one of the case studies that forms part of the PhD which looks at tweets related to Swine Flu from 2009 there are over 7 thousand tweets which were analysed. The first issue was related feasibility, i.e., that it would be very labour intensive to individually ask each user whether or not they would to take part in the research. The second issue is that, considering the outbreak occurred in 2009, the rate of potential responses was likely to be low. The users who tweeted in 2009 might have left the platform, or may take a long time to reply which would fall beyond the scale of the project. Issues over informed consent highlight the need for researchers to work alongside social media companies, for instance, asking users at the sign up phase whether they are OK with their content being used for research purposes. Or more generally, displaying a pop up that allows social media users to opt of research projects.

Legal Concerns

Twitter's Terms of Service and Privacy Policy are documents which govern how users may access and use the Twitter platform (Zimmer & Proferes, 2014; Weller, Bruns, Burgess, Mahrt, & Puschmann, 2014). By agreeing to Twitter's terms and service agreement, users will consent for their information to be collected and used by third parties (Twitter, 2016A). For example, the privacy policy notes that:

> 'What you say on the Twitter Services may be viewed all around the world instantly. You are what you Tweet!' (Twitter, 2016A).

The Twitter terms of service notes that:

> "You agree that this license includes the right for Twitter to provide, promote, and improve the Services and to make Content submitted to or through the Services available to other companies, organizations or individuals who partner with Twitter for the syndication, broadcast, distribution or publication of such Content on other media and services, subject to our terms and conditions for such Content use. (Twitter, 2016B).

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. This procedure would bypass retrieving data from Twitter's APIs and would allow Twitter to see who has retrieved data from the platform. As a consequence this practice is expressively discouraged by Twitter:

> …scraping the Services without the prior consent of Twitter is expressly prohibited (Twitter, 2016B)

Additionally, reproducing tweets but removing user IDs, or altering tweets significantly will contravene Twitter's User Development Policy which requires tweets to be published in full. However, academic researchers could argue that the policy is frequently breached and that Twitter has never taken any action due to the breaches. This is not to say that Twitter may never take action, nor that they will not take retrospective action, or that doing this is ethical. It is important for researchers to take the time to read user agreements for social media platforms as they govern what practices are permissible and provide guidance on publishing posts. Twitter users may be concerned about of who owns tweets and whether users have the right of ownership and copyright of tweets that they post. However, in practice, many tweets would not be considered under copyright law because, in most cases, tweets are not original messages (Shinen Law Corporation, n.d.). Therefore, researchers may need to act 'more ethically' than other investigators for professional reasons. Moreover, this also highlights the power of social media platforms whom not only control access to data, but whom also dictate how results of research projects are presented. For that reason, there is a definite need for researchers to engage with social companies for academic use of data.

Previous Work

NatCen Report

A report published by NatCen Social Research examined users' views towards research using social media, this was split across four focus groups, two paired, and two one-to-one in-depth interviews. The report found that the views of participants fell into three categories: scepticism, acceptance, and ambiguity. Moreover, they found that the views varied greatly depending on the context within which the research was taking place, and also of the participant's knowledge and awareness of social media websites (Beninger et al, 2014).

Beninger et al.'s study also found that participants expressed concern about the quality of social media research and that these concerns were grouped under the research principles of validity and representativeness (Beninger et al, 2014). More specifically, the concerns related to four key points:

1. Those who post online may behave differently when they engage with social media platforms compared with how they behave in real life.
2. Views on social media platforms may be more exaggerated due to the anonymity afforded by online communication. Therefore, any research findings which use social media data may lead to inaccurate conclusions of a topic of study.
3. Comments which are posted on social media websites may be impulsive and may not necessarily reflect a participant's viewpoint when in a more measured state of mind.
4. Social media profiles may not always be accurate and without any further context they may lead to inaccurate information and findings

In their report, Beninger et al. (2014) suggested that traditional ethical principles such as consent, anonymity, and avoiding undue harm should also be applied to social media research. Moreover, participants provided reasons for and against upholding the principles of informed consent and anonymity. With regard to those who felt that consent and anonymity are unnecessary, two chief

reasons were provided. Firstly, it was noted that the responsibility falls on end users to decide on how privately to post and whether to post at all. Secondly, it was noted that, as long as the site owners make it clear that social media posts may be public, as well as the level of their being public, then consent and anonymity are unnecessary. Twitter is a platform where, by default, posts are set as public and it could be argued that the responsibility falls on the end users to alter the settings of their accounts if they wish them to be more private.

With regard to those participants who felt that informed consent was necessary there were nine reasons, as described below (Beninger et al, 2014):

- It is morally and legally required
- To promote trust between the researcher and participants
- When researchers are quoting a username alongside a social media post
- When a post is no longer recent it was noted that it would be important to ascertain whether the participant still holds the same view
- When researchers seek to publish photographs or other images
- If a social media post is considered particularly sensitive and/or personal
- In order to ascertain whether a user intended to post publicly
- If the social media post would be used to generate a profit
- In order for users to determine both the quality and purpose of the research

In terms of the users who thought that anonymity is required on social media platforms, there were at least three key reasons. Firstly, participants felt that this was particularly necessary when informed consent was not gained. Secondly, some felt that anonymity is needed in order to avoid harm including judgements and/or potential ridicule. A third reason that participants felt that anonymity is needed was to preserve and/or protect personal or professional reputations.

Overall, it was found that there were at least four factors that would influence participants' views and expectations related to informed consent and anonymity, depending on the context of the research. The four factors identified consisted of:

- The mode and the content of the post which would also include written content, any photographs appended to social media posts, and also the sensitivity of the content
- The platform being researched, because certain platforms are seen as more private (Facebook) than others (Twitter)
- The expectations users may have when posting to a particular social media platform
- The nature of the research including the organisation the researcher is affiliated to along with the purpose of the research

Beninger et al (2014) summarised a number of suggestions for improving research practices by drawing on discussions with participants. They noted that these suggestions should not be taken as rules and may not always be appropriate in all circumstances, but that they can be considered in the design stage of studies that use social media data.

Table 1 – Suggestions for improving research practices: from sampling to reporting

Table 1. Suggestions for improving research practices: from sampling to reporting		
Research Stage	**Aim**	**Activity**
Recruitment	To appear legitimate, accommodate different user types and be transparent in your purpose and aims in order to ethically recruit participants to online and social media research.	• Ask the preferred mode of communication once a participant is recruited • Approach possible participants over the platform being used in the research (rather than email) • Be transparent in recruitment materials. Consider including your affiliation, web link to verify your idea (e.g., biography on organisational website), aims • Explicitly state the security and privacy terms in recruitment materials of the platform the research will involve • Explain where you obtained a participants contact details (i.e., Searched Facebook for public profiles) • Include a link to a company webpage; examples of previous work; transparent about research aims
Collecting or generating data	To improve representativeness of findings, and to understand the privacy risks of a platform used in a study in order to uphold protection and trust of participants.	• Recognising differing views on what is legally permitted to be collected compared to what some may consider their intellectual property • Take time to consider the openness of a platform you are using and whether steps can be taken to gain trust of users (i.e., if a closed chatroom consider introducing yourself and state your research purposes and ask participants to opt into your research) • Acknowledge the different ways users engage online, how they create, share and observe, and

		how your data may include a specific view or type of user
Reporting results	To protect the identity and reputation of participants, maintain their trust in the value of the research and contribute to the progression of the field by being open and honest in reporting.	• Testing the traceability of a tweet or post and taking responsible steps to inform the user and to protect their identity, if desired. Options include paraphrasing instead of verbatim or using quote but no handle/user name. • Where reasonable, seek informed consent to use verbatim quotes, images or video such as through direct tweets • Acknowledging limitations of the representativeness and validity of your findings • Explicitly stating the platform used (i.e., from Facebook rather than generally saying social media)

Recreated in Microsoft Word, from: Beninger, A. K., Fry, A., Jago, N., Lepps, H., Nass, L., Silvester, H. (2014). Research using social media; users' views. NatCen Social Research. [Online] Available at: http://www.natcen.ac.uk/media/282288/p0639-research-using-social-media-report-final-190214.pdf\nhttp://www.natcen.ac.uk.[Last accessed 28/11/15]

Looking more specifically at examining the recommendations above in relation to Twitter, for recruitment, researchers could be more transparent by indicating that they are academic researchers on their Twitter biographies. Researchers could also include a link to their academic page, and examples of previous work. It may also be a good idea to ask whether users are happy to be contacted via Twitter, or whether they would like to be contacted using a different communication platform. In regards to collecting or generating data it will be important to carefully consider whether any tweets contain content that has copyright restrictions and/or which has implications for intellectual property. Moreover, on Twitter academics may wish to consider how Twitter users engage with the platform, for instance, whether this is in a public or private capacity. Researchers may also wish to see whether there are methods of Twitter users opting into the research before it begins. In regards to reporting results researchers may wish to anonymise tweets and to search for the tweet in Twitter's advance search to ensure the user is not findable. It would also be advisable to state the social media platform used, and to report on the limitations in regards to the representativeness of the data.

Ipsos Mori Report

As a part of the Wisdom of the Crowd project, an ethical review was conducted on large-scale, aggregated analysis of social media data sometimes termed as 'social listening' (Ranco, Aleksovski,

Caldarelli, Grčar, and Mozetičl, 2015). In order to develop the report, three types of research were conducted. Firstly, the authors conducted a survey online of 1,250 adults aged between 16 and 75 which sought to develop a better understanding of how people perceived the use of social media data for research, as well as how useful social media research can be. Secondly, three qualitative workshops were conducted in which participants discussed how social media content could be used and "the principles of ethical social media research" (Ranco, Aleksovski, Caldarelli, Grčar, and Mozetičl, 2015. p.9). The third type of analysis was a statistical analysis of the survey.

The report itself focused on three key stages:

1. Secondary research– a review of the current literature related to social media research was conducted along with the ethical, legal, and regulatory implications
2. Primary research – interviews and discussions were held with experts and users in order to better understand issues raised at stage one
3. The final stage of the project sought to develop a set of 'best practice' recommendations by combining findings from stages 1 and 2

The first two recommendations related to increasing awareness of social media research in order to build trust with the public. The first recommendation was that researchers should aim for transparency when doing research that involves social media analysis. When it is possible, details of the research project should be provided online with an outline of whose data is going to be collected, as well the purposes for which it will be used. The second recommendation stated that social media companies should ensure that they continue to review their terms and conditions in order to simplify the possible uses of the data. Twitter makes its data open and accessible to anyone with an Internet connection, and anyone is able to build large datasets of tweets for marketing purposes. However, this may not be known to Twitter users, and this information may also be forgotten. A further recommendation could be for social media companies to remind Twitter users that their data may be accessed and used by third parties.

Recommendations three to five were based on the option of opting out (Ranco, Aleksovski, Caldarelli, Grčar, and Mozetičl, 2015) of social media research. This recommendation argues that researchers should ensure that there are systems in place that allow users to opt-out of specific social media research. For instance, having the ability for members of the public to submit their email address, or user id, in order to be excluded from social media research which is being conducted by an organisation. They recommended that research organisations would need to work with analytics platforms. This recommendation could be difficult to achieve in practice, because social media companies may be unwilling to give an option to opt out of social media research.

Recommendations six to seven were based on "minimising unnecessary personal data collection" (Ranco, Aleksovski, Caldarelli, Grčar, and Mozetičl, 2015 p.14), that is, researchers should question whether the extent to which the data they are collecting are necessary for a research project. The recommendation also suggested that researchers should remove names, or user handles from sight, strip out any other identifiable data, remove metadata which are not relevant for a project, create generalised groupings, and to develop specific metrics to analyse data rather to rely on standard algorithms. This is a particularly valid recommendation towards research on Twitter because, often, social media analytics tools are capable of retrieving vast quantities of meta-data. A further suggestion

would be for software developers to develop mechanisms of selecting the metadata to retrieve and metadata should be left out.

Recommendations eight to ten were based on not including anyone under the age of 16 in their research social media (Ranco, Aleksovski, Caldarelli, Grčar, and Mozetičl, 2015). This is because, currently, there is no method of removing minors, i.e., those under the age of 18, from social media research. This recommendation stated that researchers should derive the age of social media users from the content that they post. This is a perfectly valid recommendation, however there would be clear issues over the reliability and accuracy of such an approach, and an approach which would be very labour intensive.

Recommendations eleven to sixteen were based around "permission for publication" (Ranco, Aleksovski, Caldarelli, Grčar, and Mozetičl, 2015 p.16). Specifically these recommendations concerned some of the following:

- Projects on social media must consider whether there is really a need to publish posts verbatim and projects which desire to display verbatim text should gain approval from an internal ethics review.
- Researchers should contact individual social media users if they would like to cite their posts in their original format

Additionally, it was recommended that regulatory bodies should try to formulate clearer definitions of brands on social media. For example, some people may reasonably expect that those who have relatively large social media followings would expect to have less privacy than those who have a smaller following. Currently, some social media platforms prohibit the altering of posts at the publication stage. Here, it was also recommended that social media organisations should alter their developer guidelines in order to provide researchers with the flexibility of altering the social media content, in order to anonymise Tweets and prevent their being used in search engines to identify users.

Recommendations seventeen to eighteen were based on the definition of 'private' and note that that it is the responsibility of researchers to ensure that they have a good understanding of whether data that have been collected are public or private. Here, it was noted that analysis of private content should be approved via an internal ethics review. This recommendation is only likely to apply to Twitter research if private accounts or Direct Messages were to be studied.

Recommendation nineteen was based on "establishing ethics reviews for social media research" (Ranco, Aleksovski, Caldarelli, Grčar, and Mozetičl, 2015 p.17). It was noted that social media companies should provide clarification on whether users are happy with privately shared data being used for research. Moreover, it was noted that researchers should try to understand the potential harm to participants, and to identify possible steps that could put in place in order to meet user expectations and to also protect users from harm.

Academic and Industry Perspectives

In an academic setting, it is widely considered a cornerstone of research integrity and research quality to have considered the ethical implications of research, especially within the fields of social research. Bryman (2008) noted at least three areas of ethical concern that must be considered: informed

consent, invasion of privacy, and the risk of deception (Bryman, 2008). He also noted that, within social research, consideration must also be given to legal aspects of holding personal data on others, such as the UK Data Protection Act (1998) or, for that matter, European or international directives. Higher educational organisations may well have a dedicated research ethics committee or Institutional review board which could offer guidelines regarding research ethics. Research may need to pass through a research ethics committee, whose role it is to protect research participants from potential harm, institutions from potential negative attention and reputational risk, as well as the researchers themselves. All of these principles apply to social media research because, essentially, the majority of content on online spaces such as Twitter is curated by people, with the exception of organisational, news, and automated Twitter accounts.

Broadly speaking, industry perspectives on social media research and research ethics may differ from that of academic research. In industry, it is rare to find research ethics committees, and there is more of a focus around the legality of the research being undertaken, and whether the users have shared posts publicly. It is important to note that while an individual researcher or organisation may have met their legal obligations, the ethical concerns may still persist, and these should not be considered in isolation of each other. One of Twitter's revenue streams is via the reselling of its data to third parties, and these data are often used by advertisers to target users with products. Therefore, public posts may be considered by some to be 'fair game', because analytical tools for analysing social media data are readily available. To provide an example of some of the differences between academic and industry research, we can examine the #SpeakBeautiful campaign launched by Dove. An advertising campaign, called #SpeakBeautiful, saw Dove enter into a partnership with Twitter in order to send non-automated responses to negative tweets which were sent by women with the aim of empowering the users in order for them to speak with "more confidence, optimism and kindness about beauty online" (Nudd, 2015). Dove and Twitter may argue that Twitter data are publicly available, and that users check a terms and conditions tick-box when registering for Twitter. However, boyd & Crawford (2012) noted that, for academic researchers, it is not sufficient to state that using social media data is ethical just because the data are accessible.

The next section outlines two further case studies based on anonymised accounts of real research projects. These case studies hope to highlight that it is not possible to take a fixed position in relation to research on Twitter as different projects will have different aims and study different phenomena. However, it will be possible to identify patterns and principles that cut-across social media research and so will be able to cut-across a number of potential studies.

Case Studies

Case Study 1 – Analysing Mentions of a Terrorist Group

Imagine a study wished to understand better how users talk about terrorist organisations by analysing hashtags in support, and against, the organisation. It might be decided that the topic of interest is of a highly sensitive nature because if the tweeters were identified, it could place them at risk of serious harm, e.g., from people or organisations with opposing views. There are issues over protecting the anonymity of the participants, and issues over informed consent, and whether there is likely to be any under 18s that would be tweeting. Due to the immediacy of the research, the result of the research could also pose a security threat. It could be argued that the research should not take place at all, as the risks of the research might outweigh any benefits. If a research project of this nature were to go

ahead, then the anonymity of the participants must be protected, the analysis performed should be of an aggregate nature, and data must be stored securely. In such cases, direct quotes should not be utilized as these could compromise the anonymity of Twitter users. Additionally, the method of analysis should aim to uphold the anonymity of Twitter users, and examine the themes that emerge from the data and report on aggregated trends without drawing attention to individual Twitter users. There would also be issues over publishing the dataset of tweets; however, according to Twitter's terms of service agreement, it is not possible to share datasets. However, sharing tweet Identification Numbers (IDs) is permissible, and this is particularly useful for academics as it allows research to be reproduced. However, the researchers in this case may not wish to release tweet IDs due to the potential risk of harm.

Case Study 2 – Examining correlations of followers

Imagine another proposed study that seeks to examine the correlations between the number of Twitter followers a company has and its worth on the stock exchange. The project would not collect Twitter data in itself, but would examine the number of followers a particular organisation had. This case study is distinct from the previous three case studies in that the researchers are not retrieving data generated by human participants. This is one of the major ethical debates in social media research i.e., identifying what counts as a human subject (Metcalf & Crawford, 2016).

It is important to note that not all research on Twitter will analyse the content of tweets, because Twitter has over 20 metadata fields. Research on Twitter, for instance, has been used to predict when to buy or sell stock corresponding to whether there has been a peak in tweets (Anco, Aleksovski, Caldarelli, Grčar, and Mozetič, 2015). There may be data that has recorded how a user engages with Twitter, for instance, their likes, shares, retweets, followers and so forth. Regardless of whether a research study would make use of tweet content, there would still be issues over informed consent, and how the results were likely to be used and perceived by Twitter users. In this particular case study, there would be ethical issues on how the result would impact the organisation.

Case Study 3 – Using Network Visualizations to Analyse a Hashtag relating to sensitive topics

Imagine a further study that seeks to use a popular network visualization tool in order to analyse a hashtag dedicated to particular views on sensitive topics, for example, abortion, sexuality. There are two versions of the visualization, an online and an offline capability. In the online visualization, Twitter users would be fully identifiable and their tweets would be visible via an interactive interface. The tweets and users would be clustered based upon their opinions, which may draw attention to individuals who may otherwise be lost in a crowd. Furthermore, the online version of the network visualization would display the Twitter handles of the individuals, whereas the offline network visualization would produce a network graph where users would be identifiable, but which would have the functionality of anonymising Twitter users. In both the online and offline capabilities, there are issues over protecting the anonymity of the participants, and issues over informed consent, and whether there is likely to be any under 18s that would be tweeting. In an academic context, it would be advisable to use an offline version to produce the network visualization and to only identify users who would have consented for their Twitter ID, and tweet, to be identified as holding a position within the network. This is because although the users tweeting using a particular hashtag would understand

that their tweets may reach a larger audience, this is a sensitive area and could have a potentially negative impact on participants.

There have been some real life examples of controversies regarding the use of social media data, as illustrated below.

Samaritans Radar Application

The Samaritans Radar Application was an online application which was designed to monitor tweets on Twitter matching a specific set of keywords such as 'kill myself' or 'end it all' (Lee, 2014). The application was designed to then send an alert if it was to spot anyone who may have been struggling to cope. The application marketed itself by suggesting that it gave Twitter users a second chance in seeing tweets which may have been missed (Samaritans Radar, 2016; Lee, 2014).

When news broke of the release of the application on Twitter, there was a large negative reaction with Twitter users being concerned about their privacy. Samaritans released a number of announcements noting that there was a Whitelist which could prevent tweets from being monitored. Another announcement noted that the app was in development for over a year and had been tested with a number of user groups from young people with mental health issues, Samaritans' volunteers, social media platforms and other organisations. They also noted that they had worked with academic experts from the Universities of Glasgow and Cardiff on the project.

Further issues were raised over data protection laws, and whether the app would act as a data controller. Samaritans sought legal advice and argued that Samaritans would not act as a data controller, or processor of information. Moreover, they argued that, even if they were judged to be a data controller, they would be exempt from data protection laws. Moreover, this was also related to how an online community felt that their safe space to share honest thoughts and feeling had been breached.

Despite the reassurances offered by the Samaritans Radar App, shortly after its release (nine days), the application was suspended, and Samaritans offered a full apology alongside a support page for anyone whom had been negatively affected by the launch of the application. Ultimately, Twitter users believed that collecting and sharing their data without their express permission infringed their right to privacy. This case study highlights the importance of understanding how users view their privacy on online spaces. Lee (2014) noted that a lesson could be learnt by those whom may develop similar apps in the future to ensure that their testing groups will reflect the wider Twitter ecosystem. Moreover, this case study highlighted the issues that, although tweets can be accessed publicly, Twitter users could expect a certain level of privacy when engaging on the platform.

Facebook Emotion Study

An experimental study sought to exclude certain elements from 689,003 peoples' news feed which were around 0.04% of all Facebook users during a week in 2012 (Arthur, 2014). The experiment manipulated two groups of Facebook users to assess whether their emotional states, which were measured by posting behaviours, would be affected by the emotional expressions of others (Arthur, 2014; Kramer, Guillory, & Hancock, 2014). Kramer (2014), an author of the paper wrote in a Facebook post, in defence of the paper, that:

"We felt that it was important to investigate the common worry that seeing friends post positive content leads to people feeling negative or left out. At the same time, we were concerned that exposure to friends' negativity might lead people to avoid visiting Facebook"

The study received a lot of negative publicity because of the methods that were used by the researchers. In particular, this concerned the decision related to not approaching participants for informed consent. Some researchers argued that consent was not necessary as the study was unlikely to cause significant harm to participants, and that the study was important contribution to developing new knowledge (Kleinsman & Buckley, 2015).

Facebook displays advertising on its platform which attempts to alter people's buying habits by trying to make them buy a product and/or service from advertisers; this is part of Facebook's business model. However, users may know that the adverts that they see are targeted and come to expect this from a free service. Conversely, in this context, users were affected without having prior knowledge or giving informed consent. This case study, at the very least, highlighted the need to consider whether a particular type of social media research will require informed consent.

Common Themes across Case Studies

There are a number of cross-cutting themes across the case studies that have been outlined, and these issues may need to be considered when undertaking social media research. For example, in case study 1 and case study 3 the central issue was based on the risk of harm to social media users. This is because if users were identified mentioning a terrorist group and/or a sensitive topic then they may be at an increased risk of harm from those who may hold opposing views. In case study 2 an issue that was highlighted was related to the difficulty of identifying whether a study contained human participants. The final two case studies related to the Samaritans Radar Application and the Facebook Emotion study which highlight how issues of informed consent can arise in social media research. The purpose of outlining these distinct cases studies was to highlight how even among different topics it would be possible to identify patterns and principles that cut-across social media research.

Ethical Practicalities for a PhD Project

Research Ethics Application for the University of Sheffield

The authors of this chapter form part of a team which is undertaking a novel PhD study that is using Twitter data related to infectious disease outbreaks as a primary data source. The research team obtained research ethics approval in accordance with the research ethics policy at the University of Sheffield (n.d.) (https://www.sheffield.ac.uk/polopoly_fs/1.112642!/file/Full-Ethics-Policy.pdf). This section outlines the steps and arguments that were made in order to obtain research ethics approval.

Initial Steps

The team proposed in the ethics application that data on Twitter could be considered to be public because anyone with Internet access can access content on Twitter. There is no need to subscribe, enter a password, or pay to access the data. One of the first questions raised was whether ethics approval was required for the project because, seemingly, Twitter data were in the public domain.

It was noted that ethical approval was required, because the data are generated by people, and individuals may be identifiable from the data. In addition, when the data related to infectious disease

outbreaks are analysed issues may emerge from the data that could draw attention to groups, individuals, and trends. This would be beyond what would normally be expected from engagement on these platforms. For example, someone who was in conversation with another user on Twitter may tell a joke about Swine Flu believing that it would eventually disappear. In addition, people may post tweets when they are under distress, for instance while experiencing an infectious disease, believing for their post to have disappeared, or simply having forgot about it.

Potential Participants

The application stated that all captured and relevant tweets would be analysed, that is, those relating to relevant epidemics and pandemics or other health related topics. These tweets could be posted by any user with a Twitter account, e.g., anyone from the general public or from an organisation. It was decided that tweets from public figures with Twitter accounts would not be examined intentionally; however, their tweets may co-incidentally be a part of the data captured. Tweets with geographical locations (geotag data) would only be analysed at an aggregate level, i.e., at the country level, i.e., from where people were tweeting. It was also made clear in the application that Tweets with geographical locations would not be used to identify individual users.

Informed Consent

In the project, it was decided that it would not be practicable to gain informed consent to analyse tweets as a sample of tweets may contain in excess of a hundred thousand items. However, during the analysis of tweets, it may become apparent that tweets from a user or a set of users were of particular interest. In this case, if it would be useful to quote the content of the user(s) verbatim, e.g., for the purpose of reporting and substantiating the results, or for the user IDs to be indicated, e.g., in the PhD thesis or in a publication, then informed consent would be sought retrospectively. In the case of Twitter this would involve sending a tweet to the user with details of the study and requesting permission to quote their Tweet.

However, it was noted that it may be difficult to obtain consent via participant information sheets and consent forms as users may not wish to reveal their email address or click on links to participant information sheets and consent forms. Therefore, in this situation, it was argued that it may be necessary to accept a Tweet or saying 'Yes' or similarly suggesting that quotes from the Tweet could be used. In some instances, the researcher may have to gain consent via a Tweet, email, etc. The application noted that one of the researchers had a public Twitter account which could be used for this. However, in order to gain consent electronically, it was argued that the researcher would ensure that the person giving consent was the rightful owner of a Twitter account.

The research project, although having procedures in place for quoting tweets in research, as outlined above, took the ethical standpoint of not quoting tweets or disclosing non-public usernames, unless with the permission of the user. The reason for taking this decision is that those users tweeting, although they may be doing so in a public space, may not be aware that their tweets are being used for academic research. Although Twitter's terms and conditions state that user data may be redistributed or used for other purposes, a survey once found as few as 18 per cent of users may actually read terms conditions agreements (Zimmer & Proferes, 2014).

Potential Harm to Participants and Data Confidentiality

With regard to data confidentiality and back-up, it was noted that data would be kept on two secure password-protected laptops alongside a University backed-up secure research server. Individual

tweets would not be published without informed consent. In the case of Twitter's hash tag(s), generated by users for an infectious disease outbreak, there was the possibility of identifying participants through de-identification techniques, that is, by searching for the hashtag using a search engine and locating participants. It was argued that, if this was to occur, the risk to the end users would be low as the captured data did generally not fall under the category of being a highly-sensitive topic, unless it related to the health of an individual, e.g., suffering with the Ebola Virus Disease (EVD).

Data Storage

Related to the above, the ethics application outlined who would have access to the data and how they would be stored. It was decided that the lead researcher would have control of, and act as the custodian for, the data captured from Twitter (note this is not the same as being the data controller, as defined under English law). The analysis of the data would be conducted by the lead researcher and would take place in the researcher's place of study and home. It was argued that the data would not be analysed in places deemed as public. The data would be stored on two password protected laptops, which themselves would be stored securely when not in use. The ethics application also noted that certain data might need to be shared with the supervisors for marking or for administrative use.

Issues Encountered

In addition to a number of ethical issues that were faced, there were also legal considerations, issues surrounding the retrieval of datasets, issues around cost of data, and dealing with tweets which were spam or fictitious accounts. It is important to understand the limitations of Twitter data, and the issues below also have implications for conducting, these challenges are described below (Ahmed, 2015b):

1. As highlighted in this chapter there were a number of ethical issues that we encountered around obtaining informed consent due to the large volume of tweets that were retrieved for the project. Moreover, as mentioned earlier, we also faced specific moral challenges related to whether we could publish tweets in publications that refer to user handles that can identify Twitter users. We decided not to gain informed consent from Twitter users, but with the proviso of not reporting user-handles and/or tweet verbatim. Tweets that were reported in the results were reworded.

2. We had to consider legal issues, that is, the sharing of Twitter datasets is prohibited by Twitter's API Terms of Service (Twitter, 2012b). However, we found that we could release tweet *Identification Numbers*, when required, so that other researchers could retrieve a similar dataset. We were also transparent with the date of retrieval for the data that we collected. As to allow other researchers the ability to retrieve data from a similar time period.

3. We were aware that there was likely to be missing data at the point data retrieval, for instance, when retrieving data using the keyword 'Swine Flu' this would only retrieve data from users whom had used that specific keyword. If a user sent out a tweet that mentioned Swine Flu, without using the keyword, our data retrieval system would miss this. Issues around data retrieval have potential to lead to a biased sample, and this can also occur by the choice of language which is used to create a list of queries.

4. We also faced issues over the cost of data, as historical Twitter data can cost a lot of money depending on the query terms used to retrieve data as well and the time period of data retrieval. This is because if it has not been possible to set up a data retrieval system it will only be possible to retrieve Twitter data going

5. There is also the issue of representivity as Twitter users are not representative of the national offline population. Further than this Twitter users are not representative of Internet users, and Twitter data is not representative of Twitter users. This is because not all users will tweet about a topic of interest, such as, Swine Flu and those that do will belong to a very specific group of users.

6. On Twitter we found that there was a lot of spam such as link-baiting which occurred in popular hashtags. Moreover, we also found that there were a lot of fictitious accounts on Twitter, and we are aware that celebrities, politicians among others are likely to purchase retweet, favourite or user packages to increase their brand presence. This type of content and accounts are likely to affect the validity and reliability of results.

Conclusion

This chapter has provided an overview of existing literature in order to inform those whom may be undertaking social media research. It then outlined a number of industry and academic case studies in order to highlight the challenges that may arise, and finally it provided an overview of the steps that were described to gain ethics approval for a PhD project using Twitter as a primary source of data. An advantage of using Twitter data for academic research is that it may be possible to retrieve data at a faster rate than it might take to run a survey or a series of interviews.

Emerging news stories, crisis events, and political discourses, for example, can all now be studied almost as soon as they occur. Researchers are now able to examine reactions or sentiments to most major events via social media data without the need to worry about the accuracy of participant's memories in recalling how they felt during an event. Therefore, social media research can be said to avoid one of the central problems facing qualitative research: chiefly that of interviewer bias (McKee, 2013).

However, although social media as a source of data may have benefits over traditional qualitative research data, as highlighted throughout this chapter, it brings with it its own challenges. There are a number of overlapping ethical and methodological issues to social media research that must be carefully considered, and researchers must reflect on whether data from social media platforms can sufficiently address the research question of a project.

References

About, (n.d.). *About Twitter, Inc*. [Online]. Retrieved from: https://about.twitter.com/company [Last Accessed 29/03/16].

Ahmed, W. (2015a). [Online]. Using Twitter as a data source: An overview of current social media research tools LSE Impact of Social Sciences blog Retrieved from: http://blogs.lse.ac.uk/impactofsocialsciences/2012/04/19/blog-tweeting-papers-worth-it/ 10 July 2015 (Last Accessed 11/07/2015).r

Ahmed, W. (2015b). [Online]. Challenges of using Twitter as a data source: An overview of current resources LSE Impact of Social Sciences blog Retrieved from: http://blogs.lse.ac.uk/impactofsocialsciences/2015/09/28/challenges-of-using-twitter-as-a-data-source-resources/ 28 September 2015 (Last Accessed 11/07/2015).

Arthur, C. (2014). [Online]. Facebook emotion study breached ethical guidelines, researchers say. *The Guardian*. Retrieved from: https://www.theguardian.com/technology/2014/jun/30/facebook-emotion-study-breached-ethical-guidelines-researchers-say [Last Accessed 11/08/2016].

Beninger, A. K., Fry, A., Jago, N., Lepps, H., Nass, L., Silvester, H. (2014). Research using social media; users' views. *NatCen Social Research*. [Online] Available at: http://www.natcen.ac.uk/media/282288/p0639-research-using-social-media-report-final-190214.pdf.[Last Accessed 28/11/15].

Boyd, D. & Crawford, K (2012). Critical questions for big data. Information, Communication and Society. 15 (5). 662-679.

Boynton, G. R. (2014). Co-Motion—Twitter Enters the Public Domain. *Open Journal of Political Science*, *04*(02), 76–88. doi:10.4236/ojps.2014.42009.

Bryman, A. (2008). *Social Research Methods*. *Social Research* (Vol. 3rd). doi:10.4135/9781849209939.

British Psychological Society (2013). [Online]. Ethics Guidelines for internet-mediated research. Available at: http://www.bps.org.uk/system/files/Public%20files/inf206- guidelines-for-internet-mediated-research.pdf [Last Accessed 02/10/16].

Cameron, M. A., Power, R., Robinson, B., & Yin, J. (2012). Emergency situation awareness from twitter for crisis management. *In Proceedings of the 21st international conference companion on World Wide Web - WWW '12 Companion* (p. 695). New York, New York, USA: ACM Press. doi:10.1145/2187980.2188183

Chorus. (n.d.). Project site for the Chorus Twitter analytics tool suite. [Online]. Retrieved from: http://chorusanalytics.co.uk/ [Last Accessed 11/12/16].

Chaffey, D. (2016). Global social media research summary. *Smart Insights: Social Media Marketing*.

Chew, C., & Eysenbach, G. (2010). Pandemics in the age of Twitter: Content analysis of tweets during the 2009 H1N1 outbreak. PLOS ONE, 5(11).

Dugan, L. (2011). [Online]. 52 percent of Twitter users do not consider legal implications of their tweets. All Twitter. Retrieved from: http://www.mediabistro.com/alltwitter/twitter-legal_b15407. [Last Accessed 02/10/16]

Discovertext.com, (n.d.). [Online] Retrieved from: https://www.discovertext.com/ [Last Accessed 15/12/2016].

Evans, H., Ginnis, S., & Bartlett, B (2015). [Online]. #SocailEthics a guide to embedding ethics in social media research. [Online] Retrieved from: https://www.ipsos-mori.com/Assets/Docs/Publications/im-demos-social-ethics-in-social-media-research.pdf [Last Accessed 22/09/16].

Google Scholar. (2016). [Online]. Scholar.google.com. Retrieved 22 August 2016, from https://scholar.google.com/schhp?hl=en [Last Accessed 22/08/16]

Gupta, A., Joshi, A., & Kumaraguru, P. (2012). Identifying and Characterizing User Communities on Twitter During Crisis Events. *Proceedings of the 2012 Workshop on Data-Driven User Behavioral Modelling and Mining from Social Media*, 23–26. doi:10.1145/2390131.2390142

Hawksey, M. (n.d.). TAGS. [Online]. Retrieved from: https://tags.hawksey.info/. [Last Accessed 13/12/2016].

Kleinsman, J., & Buckley, S. (2015). Facebook study: a little bit unethical but worth it? *Journal of Bioethical Inquiry.*

Woodfield, K., Morrell, G., Metzler, K., Blank, G., Salmons, J., Finnegan, J., & Lucraft, M. (2013). Blurring the Boundaries? New social media, new social research: Developing a network to explore the issues faced by researchers negotiating the new research landscape of online social media platforms. NCRM Paper. NCRM.

Kramer, A.D.I. (2014). [Online]. [Untitled Facebook post, June 29, 2014.] Retrieved from: https://www.facebook.com/akramer/posts/10152987150867796. [Last Accessed 08/14/16]

Kramer, A.D.I., J.E. Guillory, and J.T. Hancock. (2014). Experimental evidence of massive-scale emotional contagion through social networks. *Proceedings of the National Academy of Sciences* 111(24): 8788–8790.

Lee, N. (2014). The Lancet Technology: November, 2014. *The Lancet,384*(9958), 1917.

Lachlan, K. A., Spence, P. R., Lin, X., Najarian, K., & Del Greco, M. (2015). Social media and crisis management: CERC, search strategies, and Twitter content. *Computers in Human Behavior*. doi:10.1016/j.chb.2015.05.027

Metcalf, J & Crawford, K. (2016). Where are human subjects in Big Data research? The emerging ethics divide. *Big Data and Society.*

Mozdeh (n.d.). Mozdeh Twitter Time Series Analysis. [Online] Retrieved from: http://mozdeh.wlv.ac.uk/ [Last Accessed 14/01/2016].

Mendoza, M., Poblete, B., & Castillo, C. (2010). Twitter under crisis. In Proceedings of the First Workshop on Social Media Analytics - SOMA '10 (pp. 71–79). New York, New York, USA: ACM Press. doi:10.1145/1964858.1964869

Macy, M., Mejova, Y. & Weber, I. (2015). *Twitter : a digital socioscope*. New York, NY: Cambridge University Press.

McKee, R. (2013). 'Ethical issues in using social media for health and health care Research'. Health Policy.110.2: 298-301.

Nudd, T. (2015). [Online]. *Dove and Twitter Team Up to Address Hateful Tweets About Beauty on Oscar Night*. AdWeek., Retrieved from: http://www.adweek.com/news/advertising-branding/dove-and-twitter-team-address-hateful-tweets-about-beauty-oscar-night-163040 [Last Accessed 02/10/16].

OECD (2016). [Online]. Internet access (indicator). Available at: https://data.oecd.org/ict/internet-access.htm.doi: 10.1787/69c2b997-en [Last accessed 14/08/16].

Procter, R., Vis, F., & Voss, A. (2013). Reading the riots on Twitter: methodological innovation for the analysis of big data. *International Journal of Social Research Methodology*, *16*(3), 197–214. doi:10.1080/13645579.2013.774172.

Panger, G. (2014). Why the Facebook Experiment is Lousy Social Science. Medium. Retrieved from: https://medium.com/@gpanger/why-the-facebook-experiment-is-lousy-social-science-8083cbef3aee#.xxqym5ltd [Last Accessed 14/12/16].

Purohit, H., Hampton, A., Shalin, V. L., Sheth, A. P., Flach, J., & Bhatt, S. (2013). What kind of #conversation is Twitter? Mining #psycholinguistic cues for emergency coordination. *Computers in Human Behavior*, *29*(6), 2438–2447. doi:10.1016/j.chb.2013.05.007.

Ranco, G, Aleksovski D, Caldarelli G, Grčar M, Mozetič I. (2015). The Effects of Twitter Sentiment on Stock Price Returns. PLoS ONE 10(9): e0138441.

Shinen Law Corporation. (n.d.). Twitterlogical: The Misunderstandings of Ownership. [ONLINE].Retrieved from: http://www.canyoucopyrightatweet.com/ [Last Accessed 08/10/16].

Simon, T., Goldberg, A., Aharonson-Daniel, L., Leykin, D., & Adini, B. (2014). Twitter in the cross fire--the use of social media in the Westgate Mall terror attack in Kenya. *PloS One*, *9*(8), e104136. doi:10.1371/journal.pone.0104136

S.A. Golder, M.W. Macy. (2014). Digital footprints: opportunities and challenges for online social research. Annual Review of Sociology, 40 (1) (2014), pp. 129–152.

Samaritans Radar (2016). [Online]. Samaritans.org. Retrieved from: http://www.samaritans.org/how-we-can-help-you/supporting-someone-online/samaritans-radar N.p., [Last Accessed 02/10/16].

The University of Sheffield (n.d.). Ethics Policy Governing Research Involving Human Participants, Personal Data and Human Tissue. [Last Accessed 01/05/2015].

Tomer, S, Avishay, G, Bruria, A. (2015). Socializing in emergencies—A review of the use of social media in emergency situations, International Journal of Information Management, 35 (5), pp. 609-619. Retrieved from: http://www.sciencedirect.com/science/article/pii/S0268401215000638 [Last Accessed 07/10/16].

Townsend, L & Wallace, C. (2016). [Online] Social Media Research: A Guide to Ethics. The University of Aberdeen. Retrieved from: http://www.dotrural.ac.uk/socialmediaresearchethics.pdf. [Last Accessed 15/08/16].

Twitter. (2016A,). [Online]. Twitter Privacy Policy. Twitter.com. Retrieved from: https:// twitter.com/privacy [Last Accessed 02/10/16].

Twitter. (2012b). Twitter / Twitter Terms of Service. Twitter.com. Retrieved from: https:// twitter.com/tos.[Last Accessed 02/10/16].

Twitter. (2016B). Twitter. Twitter Terms of Service. Twitter.com. Retrieved from: https:// twitter.com/tos. [Last Accessed 02/10/16].

UK Parliament, (1998). *Data Protection Act*. London: Her Majesty' Stationery Office. [Last Accessed 02/10/16].

What APIs Are And Why They're Important. (2016). [Online] *Readwrite*. Retrieved from: http://readwrite.com/2013/09/19/api-defined/. [Last Accessed 02/10/16].

Weller, K., Bruns, A., Burgess, J., Mahrt, M., & Puschmann, C. (Eds.) (2014). Twitter and Society. New York et al.: Peter Lang.

Zimmer, M. & Proferes, N. (2014). Privacy on Twitter, Twitter on Privacy. In Jones, S (Eds.) *Twitter and Society* (pp.55-67). New York, NY: Peter Lang.