

Social Media Research: A Guide to Ethics

Dr. Leanne Townsend and Prof. Claire Wallace

The University of Aberdeen

With contributions and feedback from: Christian Fuchs, University of Westminster; Curtis Jessop, Natcen; Dave Harte, Birmingham City University; Elvira Perez Vallejos, University of Nottingham; Hayley Lepps, Natcen; Jennifer Holden, University of Aberdeen; Kate Orton-Johnston, University of Edinburgh; Luke Sloan, Cardiff University; Michael Smith, University of Aberdeen; Samantha McGregor, Economic and Social Research Council; Suzanna Hall, University of Roehampton.



This work was supported by the Economic and Social Research Council [grant number ES/M001628/1] and was carried out at The University of Aberdeen.



Contents

1. Introduction -----	p3
1.3 Background -----	p3
1.2 Aims of this document -----	p3
1.3 Social media, research and ethics -----	p4
2. Key areas of concern within social media research -----	p5
2.1 Private vs. public? -----	p5
2.2 Informed consent -----	p6
2.3 Anonymity -----	p6
2.4 Risk of harm -----	p7
3. Framework for ethical research with social media -----	p8
3.1 Terms, conditions and legalities -----	p9
3.2 Privacy and risk -----	p10
3.3 Re-use and re-publication -----	p12
4. Concluding comments -----	p15
5. Further reading and resources -----	p15

1. Introduction

1.1 Background

In recent years the number of people engaging with social media has exploded. Social media platforms are now utilised as key locations for networking, socialising and importantly, for reflecting on all aspects of everyday life. Such online spaces therefore hold vast quantities of naturally-occurring data on any number of topics, from consumer behaviours, to attitudes towards pro-environmental policies, to political views and preferences. This provides researchers with a huge opportunity to gather data that would otherwise have taken much time and resource to obtain. Yet this opportunity is accompanied by responsibility to ensure that how we obtain and reuse such data is done to the highest possible ethical standards. Traditional ethics frameworks can inform researchers to some extent in this, but social media data brings new contextual challenges which the more traditional approaches are not equipped to deal with. This calls for a new consideration of best practice in this domain.

This document has resulted from research and workshop activities carried out by researchers at the University of Aberdeen, on the ESRC-funded project “Social Media, Privacy and Risk: Towards More Ethical Research Methodologies”. The purpose of this project was threefold – 1) to carry out research into approaches taken by researchers and ethics committees to ensuring ethical approaches; 2) to understand user perceptions and expectations in regards to the use of their social media data for research; and 3) to produce a set of ethics guidelines for use by researchers, students, ethics committees, and anyone else with an interest in the ethics of online research methodologies. It is the third aim of the project – the production of ethics guidelines for social media research – which informs the contents of this document. These guidelines were co-produced by participants at a two-day workshop that was held in February of 2016 in Aberdeen. The participants were some of the key thinkers in this area, and they are acknowledged on the title page of the document for their contributions and feedback.

1.2 Aims of this document

This document is intended for use by researchers, students, members of ethics committees, employees of funding bodies, and anyone else with an interest in the ethics of working with



Copyright Sebastiaan Ter Burg via Flickr

social media data. The main purpose of the document is to provide clear guidelines on the ethical use of social media data in research. In this document, the term “social media” can refer to any social online data with the exception of email – social media platforms such as Twitter, Facebook

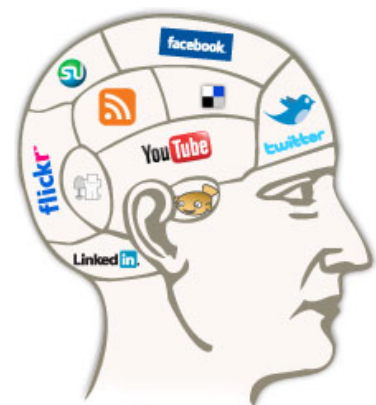
and Instagram are particularly popular with researchers due to their scale and

popularity, but researchers also work with data gathered in discussion forums, chat rooms, and via blogging sites such as Wordpress. The work reflects on guidance provided in

previous work (outlined in the Further Reading section of this document), whilst updating this guidance based on more contemporary understandings of social media – for example, some previous work refers more to email, discussion forums and personal messaging services, being written before the explosion of social media use. The intention is to provide an accessible and concise guide, rather than a lengthy inaccessible document. To this end, a framework has been produced which enables readers to work through the relevant areas and issues that are likely to arise. This framework is intended to assist individuals in making informed decisions about the most ethical approach for their research. As well as a diagrammatic framework, we have provided readers with additional guidance on each area covered in the framework, in the form of sub-sections within the text. We have also provided a number of case studies which can further help to clarify the best approach in the different social media contexts that a researcher may find him- or herself working within. It should be made clear that researchers will need to seek guidance beyond what this document provides; in particular individuals will need to read thoroughly the terms and conditions of the specific social media platform they are accessing data from, as well as consulting with their own institutional guidelines and any stipulations set out by the funding body supporting their research. Social media is a rapidly evolving landscape, and the authors are unable to predict what new ethical dilemmas may arise in future social media platforms (or future versions of those we are already familiar with).

1.3 Social media, research and ethics

Social media use is a growing phenomenon in contemporary society. Social media platforms offer their users an easy way to access and develop networks of friends, family and relevant professionals. Online communities of interest can be found to suit the interests of almost anyone. Social media platforms are increasingly used by many as a means of communication, sharing information and - importantly for this document – the sharing of attitudes and behaviours on a huge breadth of topics. It is this user-generated content that presents such a valuable opportunity to researchers. Whereas before, researchers gathered information on attitudes and behaviours through a variety of methods such as questionnaires, in-depth interviews and observation, such data is often now accessible at the mere ‘touch of a button’ (or more accurately, typing a few search terms into a platform’s search bar). Such data, found on social media platforms, online discussion forums and blogs (to name a few) is typically rich, numerous and naturally occurring (NatCen 2014). Not surprisingly then, social media platforms such as Twitter are becoming popular field sites for data collection by researchers across diverse disciplines.



Copyright Kjyrstenolson

As with other forms of data collection, the use of social media data in research poses important ethical concerns – the key concerns are explored in detail later in this document. Indeed, given the relatively new and emerging context of social media platforms as research sites, there is as yet no clear ethical framework for researchers entering this field. There have been some notable contributions in the form of guidance in recent year (see the Further Reading section of this document), yet these are often conflicting (Evans et al.

2015). This document therefore aims to support researchers with a framework which will help them to navigate the complex ethical concerns of working with social media data.

2 Key areas of concern within social media research

The following sections, outlining the key areas of ethical concern in terms of social media data, have been informed by a literature review of the available materials, interviews with researchers, ethics committee members and social media users, and the results of a two-day workshop bringing together scholars with a strong interest in the ethics of social media research.

2.1 Private vs. public?

One of the biggest areas of concern with social media data is the extent to whether such data should be considered public or private data. Key to this argument is the standpoint that social media users have all agreed to a set of terms and conditions for each social media platform that they use, and within these terms and conditions there are often contained



Credit: <http://www.intoconnection.com>

clauses on how one's data may be accessed by third parties, including researchers. Surely, if users have agreed to these terms, the data can be considered in the public domain? In our interviews with researchers, a number of responses indicated such a view e.g. *"it's public data, people know that when they sign up. So I can use that data however I like"*. But according to boyd and Crawford, *"... it is problematic for researchers to justify their actions as ethical simply because the data are accessible... The process of evaluating the research ethics cannot be ignored*

simply because the data are seemingly public" (boyd and Crawford 2012, p672). Questions of whether online postings are public or private are determined to some extent by the online setting itself, and whether there is a reasonable expectation of privacy on behalf of the social media user (British Psychological Society 2013) – for example a password protected 'private' Facebook group can be considered private, whereas an open discussion on Twitter in which people broadcast their opinions using a hashtag (in order to associate their thoughts on a subject with others' thoughts on the same subject) can be considered public. Questions of whether the data is public or private relate to the extent to which we are ethically bound to seek informed consent from social media users (see next section)¹.

¹ This could have implications for how and if the Data Protection Act is applied, see:

<https://ico.org.uk/media/for-organisations/documents/1541/big-data-and-data-protection.pdf>

There is also the issue of social media data containing data from people from broader networks, as in the case of people commenting on a social media user's post.

2.2 Informed consent

Informed consent is a critical component of the ethics of all types of research. In more traditional research approaches, informed consent is usually built in to the research design, for example in the form of consent forms or boxes to be ticked and signed on questionnaires. Social media-based research on the other hand presents problems concerning the informed consent of participants. In many cases, a social media user's data is accessed and analysed without informed consent having first been sought. 'Participants' in such research are rarely aware of their participation. Acquiring informed consent becomes more problematic the larger the data set, and can seem virtually impossible in aggregate data containing thousands or even hundreds of thousands of data units. Further, it is tempting to conflate a social media user having agreed to the terms



Credit: <http://www.pixabay.com>

of conditions of the platform (many of which include clauses on the accessing and re-use of data by third parties) with informed consent in research (Salmons, 2014) – problematic especially given that many social media users report not having read the terms and conditions properly. Important aspects of informed consent, such as the right to withdraw, are made more complicated in social media research (British Psychological Association, 2013) – for example, does deleting a post or account equate with a withdrawal from research, and is a researcher aware when this happens? When working with social media data, there are some conditions in which researchers will be more ethically bound to seek informed consent, such as when accessing data which social media users expect to be private (see above section).

2.3 Anonymity

Anonymity is a key consideration in research ethics, particularly in qualitative research practices or when data sets are shared outside of the original research team. Concerns over anonymity and online data are not new – Kleinberg highlighted the potential for anonymity breaches with social network data in 2007. With traditional forms of research, it is generally straightforward to anonymise data so that research participants cannot be identified. When working with social media data, however, anonymising data is more complex – anonymisation procedures are still evolving for aggregated or big data, and it is difficult to anonymise individual data extracts (such as Tweets) when these are reproduced in publications and during presentations (Narayanan & Shmatikov 2008, 2009). This is further complicated when some platforms insist on units of data being republished only in their original form and attributed to the original poster. Different issues arise for different types of data too – the information contained within a text-based unit of data is different to what can be gleaned from images, audio- and video-format social media data. Given that social

media companies tend to store data and meta-data for long periods, and that much of this data is searchable, anonymisation in secondary use of data in some cases becomes challenging. Further problems arise when data sets are exported to external coders and research partners. Issues of anonymisation become more critical in cases where data sets or individual units of data are published – for example online, in journal papers and at academic conferences. Protecting the identity



of unwitting participants becomes even more crucial when the data accessed refers to sensitive subject matter, particularly when exposing such data in new contexts and to new audiences may place the social media users at potential risk.

2.4 Risk of harm

Related to concerns over identity breaches is the risk of harm that researchers potentially place on their research subjects. The Association of Internet Researchers (2012) suggest that a researcher's responsibility towards his or her participants increases with the increased risk



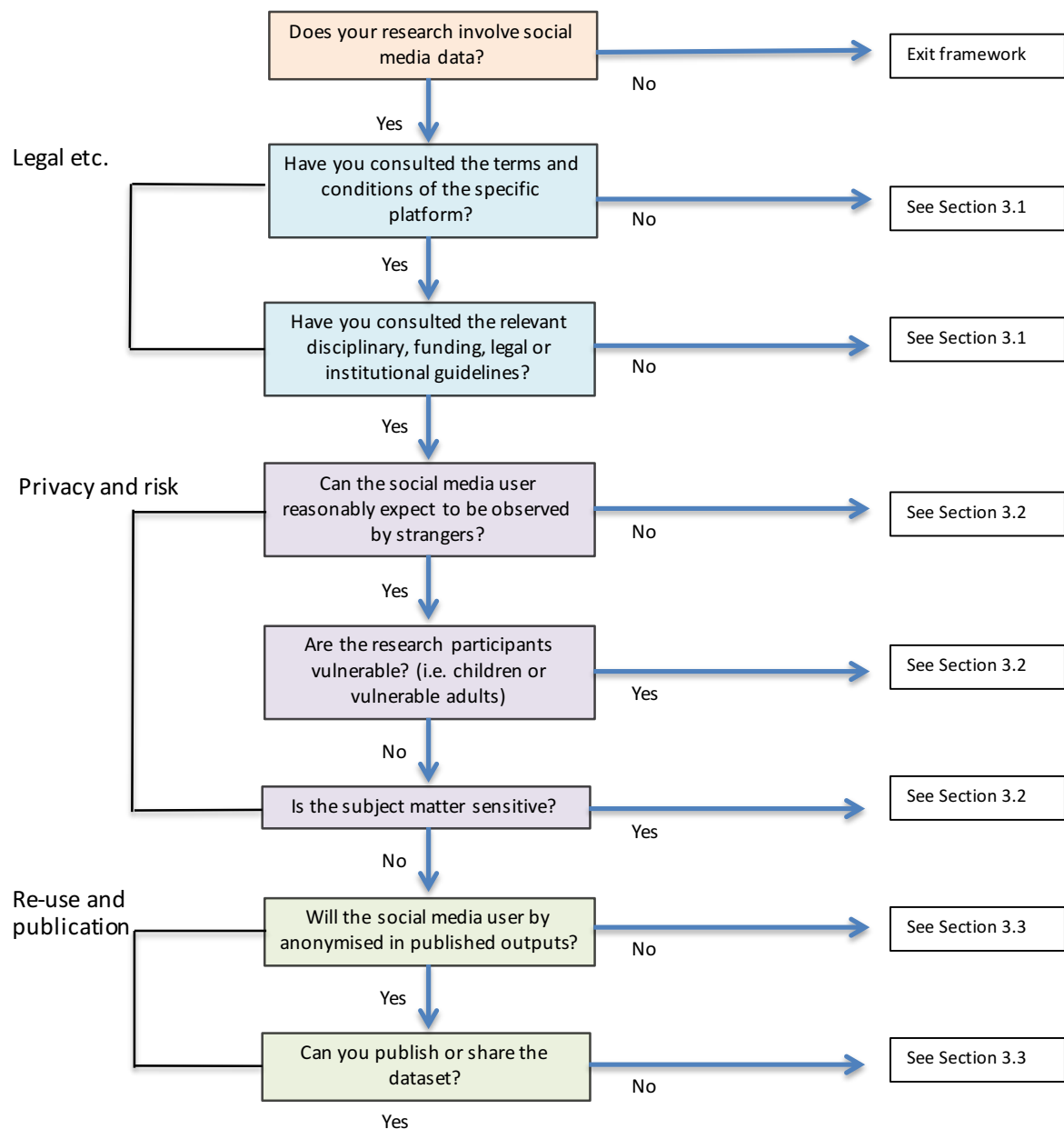
of harm to those participants, or increased vulnerability of individuals or groups online. This risk of harm is most likely where a social media user's privacy and anonymity have been breached, and is also greater when dealing with more sensitive data which when revealed to new audiences might expose a social media user to the risk of embarrassment, reputational damage, or prosecution (to name a few examples). This, of course, must be balanced with a duty of

care on the part of the researcher to report concerns such as abusive or threatening behaviour online to the appropriate channels. It is not always clear to the researcher whether or not the data they have accessed, collected, analysed or reused can be retraced in its original online context, or what the repercussions of such retracing might be. Of particular concern is the republishing of quotes that have been taken from social media platforms and republished verbatim, as these can lead us, via search engines, straight back to their original location, often then exposing the identity and profile of the social media user they originate from (British Psychological Association, 2013). There may be issues in verifying information such as whether a participant is a child, or of sound enough mind to understand the easily accessible nature of their data. This becomes of increased importance when dealing with sensitive or potentially embarrassing data. Therefore, where data deals with very sensitive topics, it becomes important to revisit the other concerns, ensuring that confidentiality and anonymity has been fully protected, and to consider whether or not to seek informed consent. Risk of harm might not be present in all instances in which a researcher wishes to cite social media data, for example when such data is shared by public

bodies or organisations, or when the social media user is clearly aiming for broad readership (e.g. by using hashtags in Twitter).

3 Framework for ethical research with social media data

Social Media Ethics Framework:



The framework presented here is the result of a workshop held in February 2016 with key scholars working in the field of social media data. Key issues were raised and discussed, and the ethics framework was co-produced between participants. The framework has been further informed by own extensive research to understand the issues and dilemmas facing researchers working with social media data today. We have also obtained extensive feedback on the framework through delegates at relevant conferences and workshops and through members of the #NSMNSS (New Social Media, New Social Science) network.

The framework represents guidelines as opposed to rules, recognising that principles need to remain flexible in order to respond to the varied contexts in which social media data is found (i.e. platform used, the target population, the topic of focus, the methodology employed and the type of data collected - text, images, video etc.). It is designed in such a way that a researcher may use it to guide and support their own decisions, rather than providing definitive answers or a 'one size fits all' approach (Association of Internet Researchers, 2012). This ultimately does leave the responsibility with the researcher, along with his or her corresponding ethics committee. We propose that researchers make reference to this document when submitting applications for ethical approval for new projects in this domain. For further support with social media research ethics, we suggest researchers consult the 'New Social Media, New Social Science?' (NSMNSS) online network, as a place to pose questions to a community of social media researchers².

3.1 Terms, conditions and legalities

Before considering other aspects of the ethics of your social media project, it is important to consult with all other relevant terms, conditions and guidelines. Firstly, you should carefully read through all of the relevant terms and conditions of the platform(s) that you will be using to obtain your data. These terms and conditions will include those aimed at users, and might also include those aimed at third parties wishing to access data from the platform. Even if you have read the terms and conditions of a specific platform at some point, it is worth bearing in mind that the terms and conditions of social media platforms change regularly in accordance with changes made to the platform, or changes in how the platform owners wish to make profit from the platform. Reading these documents may seem tedious – indeed, many platform users do not read them, and inadvertently agree to things they do not realise they have agreed to (such as use of their data by third parties!). But being familiar with the most current terms and conditions will protect you from potential legal action should you violate them.

You must also ensure that you are compliant with all terms and conditions relevant to your university or research organisation (for example on your university webpage or guidance provided to you by your ethics committee); external funding bodies that are funding your research (either in your documentation from the funding body, or on their website), and any disciplinary guidance provided either through your own university or through major disciplinary bodies such as the British Psychological Association, the British Sociological

² <http://nsmnss.blogspot.co.uk>

Association or the Academy of Social Sciences. Many such associations provide ethics guidance on their websites.

3.2 Privacy and risk

The next task in working through the framework is to determine whether the data you wish to access is really public, and if it is not to decide how - or indeed if - to proceed. The question as to whether to consider social media data as private or public comes down, to some extent, to whether or not the social media user can reasonably expect to be observed by strangers (British Psychological Society, 2013; Fuchs, forthcoming). Things to consider here are: is the data you wish to access on an open forum or platform (such as on Twitter), or is it located within a closed or private group (e.g. within Facebook) or a closed discussion forum? Is the group or forum password protected? Would platform users expect other visitors to have similar interests or issues to themselves? Does the group have a gatekeeper (or admin) that you could turn to for approval and advice? How have users set up their security settings? Data accessed from open and public online locations such as Twitter present less ethical issues than data which are found in closed or private online spaces. Similarly, data posted by public figures such as politicians, musicians and sportspeople on their public social media pages is less likely to be problematic because this data is intended to reach as wide an audience as possible. If the data you wish to access is held within a group for which you would need to gain membership approval, or if the group is password protected, there are more ethical issues to take into consideration.

Case Study 1:

Context: The researcher wishes to study support mechanisms between members of a discussion forum which deals with mental health issues such as depression and feelings of suicide. The forum is a closed forum which is password protected and registration must be approved by a gatekeeper (a site admin).

Concerns: The researcher is aware that this data is private – there is a high expectation of privacy on behalf of the users who feel it is a safe space where they will only be conversing with other people in the same situation. This raises questions about the ethics of accessing the data, and how to report the findings of the data if it is accessed.

Solution: *The researcher needs to treat this data as private and sensitive. In order to access the data, the researcher should consider seeking consent from the gatekeeper of the community (site admin), who might seek the approval of the group more widely before deciding. Once consent has been granted the researcher might wish to make themselves known to the community, and give participants the right to opt out (so that their data is not republished or analysed). The gatekeeper might grant the researcher access to a certain area of the site, and retain a 'safe' space to accommodate community members who are not comfortable with the researcher's presence. If the researcher wishes to republish certain units of data in order to illustrate their research findings, it is ethical to seek informed consent from each forum user whose data will be republished. Community members should be fully anonymised in any research outputs.*

Your first port of call should be to make contact with the site or group admin. They will have an understanding of the social dynamics of the group and will decide how to proceed. They may wish you to seek consent from individual group members for you to access their data, or offer group members the option to 'opt out' of the research (therefore you could use peoples' data unless they specify otherwise). Will you be asking questions of social media users in order to produce new data on a given subject? If so, it is vital that you are transparent about your own identity (a researcher in a university) and that responses will be used as data in your research.

The 'blurring of boundaries' between researcher & participant is a further consideration - your own social media activity (or that of people you know) may be part of the dataset you are researching, which is potentially problematic. In this case care needs to be given to how such research methodologies and findings are reported. Also, the researcher themselves might become searchable by participants, meaning that you should pay attention to your own online identity and privacy.

Another consideration here is whether or not you might be dealing with young or vulnerable participants. You must ensure that you have taken all possible precautions to rule out the use of data by vulnerable adults (i.e. those with additional educational needs) or children (or in the case of children, seeking parental consent). Social media can often make it difficult to identify such individuals, not least because people often shield their true identities on social media platforms and discussion forums. Importantly, if data is suspected to originate from young or vulnerable individuals, informed consent cannot reliably be given so this data should be eliminated from the research.

A final consideration is whether the data is potentially sensitive. Is the data about fairly mundane daily activities or opinion, or is there the potential to cause harm to social media users should their data be exposed to new audiences? Less sensitive data might include postings about, for example: the weather, recipes or consumer preferences. More sensitive data includes postings about, for example: criminal activity such as driving offences or the use of illegal drugs; financial problems; mental health issues and feelings of suicide; extramarital sexual activity; controversial political opinions and activism. It is your responsibility to decide whether the content is sensitive and if so to determine an ethical way of working with the data³. If there is risk of harm to individuals whose data you are using, you must either a) paraphrase all data which is republished in research outputs, having taken steps to ensure that the paraphrased data does not lead interested parties to the individual's online profile; b) seek informed consent from each person, should you wish to (or need to) use their data in its original form in research outputs or c) consider using a more traditional research approach where consent and confidentiality can be more safely ensured. It is also important to take these things into consideration in terms of whether you can share data sets (covered in more detail in Section 3.3). There might be cases where it is not straightforward to seek consent. Conducting critical discourse analysis of harmful or ideological social media content (such as found in Neo-Nazi online groups) is one such

³ It may be useful to refer to the EU Data Protection Reform, which legally sets out what can be considered sensitive data: http://europa.eu/rapid/press-release_MEMO-15-6385_en.htm

example (for one thing, it might be dangerous for a researcher to get in touch with these social media users). We could argue for such material as being exempt from the seeking of informed consent, in order to both protect the safety of the researcher, and to ensure that social media research ethics does not result in an indirect censorship of critical research. Here though we would advise paraphrasing quotes, not least to protect the researcher from being targeted, and to ensure an ethical approach.

Case Study 2:

Context: A researcher wishes to study pro-legalisation narratives on marijuana use. The data will be collected from Twitter, so it is open public data. The researcher will gather data over the last 7 days posted with the hashtags *#cannabis*, *#legalize* and *#ismokeit*.

Concerns: Firstly, the subject matter is sensitive because it refers to an activity that is still illegal in the UK. Secondly, there may be users under the age of 18 contributing to the debate. Therefore, the researcher must work out how to handle the data in terms of protecting anonymity.

Solution: *the researcher decides that the data is public, because it is posted on Twitter, a platform on which the default setting for posts is public; most profiles are set to public and can be viewed and followed by anyone. Furthermore the use of hashtags implies that platform users are keen to contribute to a community or debate and therefore expect an even greater number of people to see their data. The subject matter is sensitive though, and there could be children contributing data, so there is considerable risk of harm. The researcher decides it is ok to access the data and present results from aggregate data, but it is not ok to publish a data set (prohibited by Twitter anyhow) or republish direct quotes which will lead interested parties to the user's profile, hence compromising anonymity. The researcher will therefore present paraphrased quotes (removing ID handles) to reflect the themes that emerge, and provide details on how interested parties might recreate the data search for themselves. Some direct quotes may be used with informed consent from the platform user, but the researcher knows he must take steps to ensure that the user is over the age of 18.*

3.3 Re-use and re-publication

There are different types of re-use or re-publication to be taken into consideration when you are working with social media data. You may wish to publish your research results in a number of different formats, for example: online blog posts; journal papers; conference presentations (including the submission of papers to conference proceedings); book chapters; articles online such as in *The Conversation*. When reporting your findings, do you want to use units of data (such as individual tweets or Facebook postings, or Instagram images with corresponding text) to illustrate the themes that have arisen in the data? If so, you need to refer to Sections 3.1 and 3.2 in deciding whether it is ethically sound to do so. You may feel that you need to either paraphrase the data or seek informed consent from individual platform users before you can do this. You should also give consideration to the format of the data you are working with – if the data is in photographic, audio or video

formats, are there copyright issues to be considered when re-publishing? Can such data compromise the anonymity of individuals or groups?

Sharing data sets is another consideration and something that is increasingly expected, for example by external funding bodies wishing to encourage a transparent and replicable research process. You should consider whether it is ethically sound to share your data set, as well as checking the platform terms and conditions to determine whether they allow or prohibit it. If it contains data that could cause harm if re-published, then either the sensitive data should be removed or paraphrased, or the data set should not be shared at all. In cases of aggregate data where the individual units (or postings) are no longer discernible, it is generally safe to share the data set. If the data set does not contain sensitive data, or if it is not possible to identify individuals based on the data set, it is also safe to share. If you are not happy about sharing the data you will need to explain the reasons why when asked to upload or otherwise share the data set.

Special consideration needs to be given to anonymisation of social media users. In almost all cases it is important to ensure that users are anonymised in research outputs. There are some exceptions to the rule, for example public figures and organisations seeking to share their data as widely as possible. Arguably, data that is not in any way sensitive (such as postings about the weather or consumer preferences) are unlikely to cause harm to individuals, therefore you can argue that it is unnecessary to anonymise content of this nature.

Case Study 3:

Context: A researcher wishes to study public interactions on a dating platform such as Tinder. Although the posts under scrutiny are public, rather than through private messaging, she needs to sign up to Tinder to view them. By signing up, she has to fill in a registration form including questions such as “I am a woman looking for a man/woman” etc. It is therefore reasonable to think that users of the platform expect that other people viewing their profile might be doing so for similar (dating) reasons. The researcher is also aware that there may be people under the age of 18 using the platform. The users of the platform are aware that there is a very large number of people using the platform and potentially able to access their profile.

Concerns: Firstly, can the researcher ethically access and re-publish this data, given that the users of the platform have a reasonable expectation that people seeing their data are like-minded (i.e. using the platform for similar reasons)? Secondly, is there a chance that vulnerable people (such as children) could be using the platform? Thirdly, is the data likely to be sensitive?

Solution: *the researcher decides that, although the platform users may expect others viewing their profile to be like-minded, they will be expecting strangers to view their profile – so the data is not private. There is however a chance that children could be using the platform, and the data is potentially sensitive (e.g. underage children engaging in sexual talk or activity, people looking to engage in extra-marital relationships etc.). The researcher therefore can access and analyse the data, but needs to be careful with re-publishing. She does not publish the data set, and when writing up her results she only uses quotes that are paraphrased (and she is sure cannot be used to identify the platform user). Consent to use data is problematic here because the platform is popular with those under the age of 18, who may be dishonest about their age or use a misleading photograph.*

Case Study 5:

Context: A researcher wishes to explore the dominant themes in the social media postings by Olympic athletes on their social media profiles. The profiles are public and typically have hundreds of thousands of followers. The platforms under scrutiny include Twitter and Facebook.

Concerns: Can the researcher consider these postings to be public, and is it ethical to publish their data verbatim?

Solution: It is reasonable for the researcher to consider this data as public, because the sportsperson is posting on a public profile with the aim of raising their profile more generally and with the intention of reaching as many people as possible. In this case it is also reasonable for the researcher to republish their data – the sportsperson has a high expectation that (large numbers of) strangers will be viewing their data and indeed this is often desired. So the data is unlikely to be sensitive. Also given the size of their existing audience, the researcher is unlikely to pose potential harm to the sportsperson over and above any potential risk they place on themselves. The quotes can be republished in their original wording.

Case Study 4:

Context: A researcher is studying viewpoints on a public health campaign. The Twitter data is accessed via a paid-for search service. Hundreds of thousands of Tweets are collected, and a computer-programmed analysis method called 'sentiment analysis' is used to present the results of the data. This means that the data is aggregated and no individual Tweets are republished or even maintained in the data set. Instead the data set is composed of chunks of data which no longer resemble individual Tweets.

Concerns: Can the data be collected and analysed in this way? How is privacy and anonymity protected for each Twitter user?

Solution: The data is posted on a public site (Twitter) and although it could be sensitive in some cases, anonymity and privacy is maintained because the data is aggregated and individual Tweets are not republished or presented in the data set. Twitter handles (profile names) should be removed from the data set (but check the current API T&C of Twitter to ensure that it is ok to share a data set of aggregated data, with Twitter handles removed – T&Cs change regularly).

Case Study 6:

Context: A researcher conducts a critical discourse analysis of a dataset of tweets using the hashtags #DonaldTrump; #TrumpTrain; #VoteTrump2016; #AlwaysTrump; #MakeAmericaGreatAgain or #Trum2016. These are analysed in order to find out how Trump supporters argue for their candidate on Twitter.

Concerns: Can we consider this data public? Are there any issues of sensitivity or risk of harm? Do we need to seek informed consent before quoting these tweets directly?

Solution: Trump supporters use these hashtags in order to reach a broad public and convince other people to vote for Trump. It is therefore reasonable to assume that such tweets have public character: the authors expect and want to be observed by strangers in order to make a political point that they want others to read. The researcher can therefore directly quote such tweets without having to obtain informed consent. It is, however, good practice to delete the user IDs of everyday users, who are not themselves public figures.

4. Concluding comments

As argued by the Association of Internet Researchers (2012), no set of Internet research guidelines can be static, because technologies and the way that technologies are used are constantly changing. Consequently, conversations on ethical standards in social media research need to be dynamic too. Therefore, this set of guidelines should be considered flexible; we hope to see new frameworks emerge as changes to the nature and uses of social media make this necessary. Ultimately, the responsibility lies with the researcher and his or her corresponding ethics committee to ensure an ethical approach is taken to the collection, analysis and re-use of data collected from social media platforms. This framework, or any framework that supersedes it, cannot be prescriptive. Each social media research context is unique, having a unique set of characteristics and ethical challenges. We hope that this framework will provide researchers with the decision-making tools they require to take the most sensible and ethical approach to their social media research.

5. Further reading and resources

Association of Internet Research (2012). *Ethical Decision-Making and Internet Research*. Available at: <http://aoir.org/reports/ethics2.pdf>

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

British Psychological Association (2013). *Ethics Guidelines for Internet-Mediated Research*. Report available at: <http://www.bps.org.uk/system/files/Public%20files/inf206-guidelines-for-internet-mediated-research.pdf>

CASRO (2011) *Social Media Research Guidelines*. Available at: http://c.ymcdn.com/sites/www.casro.org/resource/resmgr/docs/social_media_research_guidel.pdf

Collaborative Online Social Media Observatory (COSMOS) *Ethics Resource Guide*. Available at: <https://www.cs.cf.ac.uk/cosmos/ethics-resource-guide/>

ESOMAR (2011) *Esomar Guideline on Social Media Research*. Available at: <https://www.esomar.org/uploads/public/knowledge-and-standards/codes-and-guidelines/ESOMAR-Guideline-on-Social-Media-Research.pdf>

Evans, H., Ginnis, S. & Bartlett, J. (2015) *#SocialEthics: a guide to embedding ethics in social media research*. Report available at: <https://www.ipsos-mori.com/Assets/Docs/Publications/im-demos-social-ethics-in-social-media-research-summary.pdf>

Fossheim, H. & Ingierd, H. (2015). *Internet Research Ethics*. Available at: <https://press.nordicopenaccess.no/index.php/noasp/catalog/view/3/1/9-1>

Fuchs, C. (2014) *Social Media: A Critical Introduction*. London: Sage.

Jones, C. (2011) *Ethical issues in online research*. British Educational Research Association on-line resource. Available at: <https://www.bera.ac.uk/wp-content/uploads/2014/03/Ethical-issues-in-online-research.pdf?noredirect=1>

Kleinberg, J. M. (2007) *Challenges in mining social network data: processes, privacy and paradoxes*. Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining, New York, 2007.

Narayanan, A., & Shmatikov, V. (2009) 'De-anonymizing social networks', IEEE Symposium on Security & Privacy, Oakland, CA. Available at: http://www.cs.utexas.edu/~shmat/shmat_oak09.pdf

Narayanan, A., & Shmatikov, V. (2008) 'Robust de-anonymization of large sparse datasets (How to break anonymity of the Netflix prize dataset.)', IEEE Symposium on Security & Privacy, Oakland, CA. Available at: <http://arxiv.org/pdf/cs/0610105v2.pdf>

Natcen (2014) *Research using Social Media: Users' Views*. Available at: <http://www.natcen.ac.uk/media/282288/p0639-research-using-social-media-report-final-190214.pdf>

Salmons, J. (2014) *New Social Media, New Social Science... And New Ethical Issues!* Report available at: <https://drive.google.com/file/d/0B1-gmLw9jo6fLTQ5X0oyeE1aRjQ/edit>