Wisdom of the crowd

US University of Sussex DEMOS CASM Consulting Ipsos MORI

08/06/15

Unlocking the Value of Social Media Work package 3: Ethics

Stage 1: Scoping

Steven Ginnis, Harry Evans, Jamie Bartlett, Ian Barker

Contents

Summary

Summary

This literature review is part of the **Wisdom of the Crowd** project, sponsored by Innovate UK, the UK's innovation agency, with funding contributions from the TSB, the EPSRC and the ESRC. Ipsos MORI, Demos and the University of Sussex have come together, with the Centre for Social Media Analysis to critically examine the commercial possibilities for social media research.

As part of the Wisdom of the Crowd project, Ipsos MORI undertook a review of the current literature around social media research ethics. This was in order to identify a gap for future research and to inform what that research might look like. Below is a brief summary of the issues that the review unearthed.

Introduction to social media research

- There are a number of different types of social media research methodologies, including social media monitoring, ethnographic analysis, online communities and co-creational research techniques. These methods give rise to some shared and some diverse ethical concerns; however, they all need to consider how they are using people's data and engaging with participants during the research process.

- There is a distinction made in the market research literature between 'public' and 'private' social media; however this requires further clarification. One possible solution is to make a further distinction between truly open content (public), and content that can be accessed by anyone who creates an account to that social network (open-private). How does this relate to the processing of sensitive personal information? Is this, and should it be restricted by regulation?

- The Wisdom of the Crowd project's primary focus is on social media monitoring and the ethical concerns surrounding it. However, further research will also explore the public's understanding of different forms of social media research. Do people make a distinction between aggregation and individual study? And are people aware that this kind of research is taking place at the moment?

Legal considerations

- Though there is a clear legal definition of personal data, there is currently a lack of clarity as to how this applies to social media data. Removing the unique username of the author from the rest of the data may not be enough to guarantee anonymity if the content of the social media text contains personal information about the author or another person. To what extent is it possible to anonymise social media data? Should all social media content therefore be treated as potentially personal data and thus subject to the DPA?

- There are special exemptions for the processing of personal data under the DPA for research. However, it is important to acknowledge that these are different to exemptions for other disciplines such as journalism.

- Where exemptions apply for research, personal data still needs to be processed in a fair and lawful manner. Terms of use and privacy notices provide a key part of the structure under which data can be processing 'fairly'; however this may not apply to data which has been collected in private 'walled gardens'.

- Special caution ought to be taken where sensitive personal data is being provided, but which is allowed to be processed if declared public by the author. The status of this in somewhere like Facebook (which is open-private, rather than public) is important to establish in future research.

- Specifically, the possibility for using technology for deriving characteristics about an individual based on other content appear to have potential legal barriers that should be explored further. It will be a challenge to make sure that this data is 'accurate' and that it is does not process or identify sensitive personal information which has not been made public by the author.

- There is great uncertainty around how legislation may change with new European rules coming in outlining the rights of the digital citizen. Social media research needs to future-proof to ensure that a changing legal framework does not affect ongoing research.

Industry regulations

- Revised guidelines for market research have begun to make a distinction in the consent process required for different types of social media research. Though it is clear that informed consent must be obtained where personal data is collected directly from individuals, there is greater flexibility in the method of consent required to conduct social media monitoring research. Nonetheless, this type of research is still required to have a fair and lawful basis for collection.

- However there is a lack of clarity from guidelines as to what should be considered as public or private data, what constitutes personal data, how to treat sensitive personal data, and whether it is possible to fully-anonymise social media data.

- There is a direct contradiction between some guidelines and the practice of conducting social media research. Most notably:

    - Does the industry requirement to limit the processing of personal data and publish anonymous findings prevent research identifying key authors or networks?

    - How should researchers mask social media contributions and still adhere to copyright legislation; and is there a need to differentiate between individuals and organisations or companies?

    - To what extent should personal data be processed to enrich the data with key demographics to help identify the profile of the data and differences between users.

  - How should the industry treat content from social media users under 16 (which currently requires parental consent)? Should attempts be made to identify these groups and remove them from the dataset, if so this would require the processing of personal information?

- Are these regulations sufficient for public trust to be maintained? Further primary research may be able to suggest how people might react if consent could or should be sought.

Responsibility

- It is important to consider whether the legal and regulatory framework provides the safeguards that the public hope to see. Though there is an appropriate flow of data (and to contravene this appropriateness would mean to contravene a public ethics), this has the potential to damage the reputation of research.

- It is also clear that public ethics does not necessarily align with the regulatory and legal frameworks. Though there is some consensus that public opinion is not binding, it should have a central position in any research ethics.

- The concern about doing research with under-16s should also be examined further. This exists in research regulations for good reason and exploring the potential of suspending these regulations for the case of social listening should be looked at with the public and experts.

- Getting some more complete answers on where exactly the boundaries of the appropriate flow lie is vital to this project. Research should look at the ethical issues outlined in this chapter. Themes for exploration would be:

  - informed consent and whether the public thinks it can or should be sought;

  - the role of anonymity at different stages of the research process, including how best to report social media data;

  - acceptable levels of data processing;

  - the potential for capturing the views of under-16s, and the role of parental oversight; and

  - public awareness of the terms of use and understanding of what their social media data may be used for.

- Whilst some kind of exploration of the public attitudes towards social media monitoring is key, identifying all of the boundaries in the appropriate flow of information is, inevitably, an extensive task that will not be definitively completed within this project.

Introduction

1 Introduction

This literature review is part of the Wisdom of the Crowd project, sponsored by Innovate UK, the UK's innovation agency, with funding contributions from the TSB, the EPSRC and the ESRC. Ipsos MORI, Demos and the University of Sussex have come together, with the Centre for Social Media Analysis to critically examine the commercial possibilities for social media research.

The year-long project brings together some of the leading academics, technologists, thought leaders and insight specialists to address the technical challenges caused by the huge growth in social media data. This project aims to give credibility to the research of online conversations by placing non-technical analysts (researchers) at the centre of the data exploration process, with the necessary tools and expertise to satisfy client expectations.

This research project is intended to explore the possibility of applying new social listening technology to social and market research purposes. An important part of this process is ensuring that the tool and the processes behind this tool meet the high standards demanded by industry regulators as well as the required legal obligations.

While research ethics are often perceived to be a way of ensuring that researchers do not harm research subjects in any way as their primary focus; research ethics also considers the honesty, integrity, objectivity and openness of research practices. Ethics is a way of satisfactorily balancing the expectations of research subjects with the social benefits of conducting research. Respect for the law, intellectual property rights and social responsibility are also key ethical concepts that must be considered. By ensuring social media research is carried out ethically, researchers safeguard the reputation of the industry as well as safeguarding the public. The technology in social media research is fast moving, but public attitudes towards data do not always move in step with new technology. It is vital for researchers to step back and consider the people behind the research.

1.1 The ethics work package

The ethical evaluation in this project consists of three stages.

1. Secondary research: an exploration of the current literature of the ethics surrounding social media research. This will look at the legislation and industry regulation that governs analysis of social media. The purpose of this scoping is to identify areas for further research in stage 2.

2. Primary research: interviews and discussions with experts and users to build a definitive picture of the ethics and attitudes behind social media research and drive the recommendations in stage 3.

3. Recommendations: collate information together and assess best practice for researchers when conducting social media research.

1.2 Structure

This piece of work aims to evaluate the literature systematically, but also ensuring relevance to social media research. The below structure has been designed to separate the relevant literature into different strands.

1. In chapter two, we will outline the current methodologies within the field of social media research. This is not intended to be complete, but to offer context for later chapters of the report.

2. Chapter three looks at the bare minimum requirements as set out by the law in the UK and various pieces of official guidance that have been laid out on the topic of data protection and how it applies to social media research.

3. Chapter four considers how market and social research organisations have interpreted the legal framework, and explores how social media research fits in with the principles that guide research.

4. The last chapter looks beyond legislation and regulation to examine what the concerns of the public might be. This aims to document any research that has been done with users already and outlines the gaps where more primary research can be done to identify and pre-empt future ethical considerations.

1.3 Frameworks for discussion

The structure of this report makes an obvious distinction between **legal considerations**, **regulatory considerations** and **further responsibilities to research subjects;** each of these themes have separate stakeholders, constituents and objectives. However, it is worth stressing that many of the underlying principles behind each – such as respect for privacy and informed consent – are found in all three sets of considerations. It is the aim of the Wisdom of the Crowd project is to identify an ethical approach to social media methodology that finds the convergence of these dimensions and meets the requirements of all three.

In addition, it is also helpful to consider two further frameworks:

The first of these can broadly be considered as a spectrum between different types of social media analysis. In its most simple form, at one end of the spectrum 'qualitative' social media research aims to have direct, in depth communication and monitoring with respondents on a small scale; on the other side of the spectrum, 'quantitative' research aims to use anonymised aggregated datasets, most commonly without the direct involvement of social media users. Whilst the legal and regulatory frameworks make little distinction between the different types of social media analysis, these approaches will obviously raise very different legal and ethical considerations, and possibly different solutions.

It is also important to consider the different ethical considerations of working with personal data at each stage of the research process. These stages can be categorised as **data collection**, **analysis** and **reporting**. An example of this distinction might be provided by the question of anonymity.

For instance, at the collection stage it may be impossible to guarantee anonymity (regardless of your online methodology), but this anonymity can be introduced at the analysis stage or the reporting stage.

Social media research

2 Social media research

In 2011, the Market Research Standards Board of the Market Research Society published a discussion paper outlining their initial thoughts about ethics and best practice for social media research[1]. This discussion paper, and the subsequent response to submissions on the paper[2], purposely did not make concrete recommendations, citing the rapidly changing environment of social media research.

Social media research is still changing and it is not our intention here to present a complete taxonomy of the research methodologies. The purpose of this section is to provide a distinction of the different kinds of research that are currently being used, and to demonstrate how difficult it is to outline one single ethical treatment for the diverse umbrella of social media research.

2.1 An introduction to social media research

Social media is a relatively new source of information for researchers, and given its scale, complexity and variety, covers a wide range of different research methods and approaches. Moreover, it continues to change as new social media platforms – and new types of software and methodologies to understand it – are developed.

The Association for Internet Researchers (AoIR) outlines seven different inquiries that constitute internet research, some of which are applicable to social media research. These include utilising the internet to collect data or information through data scraping; study of how people use and access the internet including participating on social networks; and employing visual and textual analytics to study images, writing and media forms.[3]

The definition of internet research is too broad to deal specifically with social media research but the AoIR guideline does provide an initial distinction within social media research. This is to do with how researchers collect their data. Some types of social media research rely on automation due to dealing with vast quantities of data, whilst others are mostly observing a small number of users. A rough analogy might be drawn here with quantitative and qualitative research – how one gleans information from large datasets, whilst the other is inquiring into individual experiences.

[1] MRSB, 'Online data collection and privacy: discussion paper' (August 2011) https://www.mrs.org.uk/pdf/2011-07-19%20Online%20data%20collection%20and%20privacy%202.pdf (accessed 26/11/14)
[2] MRSB, 'Online data collection and privacy: response to submissions' (April 2012) https://www.mrs.org.uk/pdf/2012-04-04%20Online%20data%20collection%20and%20privacy.pdf (accessed 26/11/14)
[3] AoIR, 'Ethical Decision-Making and Internet Research: Version 2.0' (September 2012) https://cms.bsu.edu/sitecore/shell//-/media/WWW/DepartmentalContent/ResearchIntegrity/Files/Education/Active/AoIR%20Social%20Media%20Working%20Committee.pdf (accessed 26/11/14)

ESOMAR – an international market research regulator – published guidelines for social media research in 2011[4]. In this they outline several different kinds of social media research that their guidelines apply to.

- Social media monitoring (which encompasses everything from basic desk research to automated sentiment analysis).
- Ethnographic analysis (this includes observation of behaviour online and may include direct contact such as 'friending' subjects).
- Online communities (these may be created by the researcher, or pre-existing).
- Co-creational research techniques (feeding the ideas of users directly into new products).

In another similar publication, CASRO – the largest American market research membership body – listed similar methodologies within social media research[5]. Whilst there is significant variation within all of these, we attempt to outline a working explanation of each of these later in the chapter.

2.2 'Public' vs 'private' content

Guidance for researchers makes a distinction between private and public social media content. This is because some areas (sometimes known as 'walled gardens') may give the user an expectation of some kind of privacy. Private social media may include areas where privacy settings are set up to prevent individuals seeing your profile or posts. The definition of public and private content is important in relation to both the legal frameworks (such as the Data Processing Act), and ethical guidelines for market research.

While there is some social media data that is clearly public, such as content on Twitter (which is public unless made explicitly private), cases like Facebook are more complicated. The current market research guidelines do not make a clear distinction between content that is public and available to anyone, and content that is available to anyone who has a Facebook account. For the majority of Facebook content, users have a reasonable expectation that their 'public' information can only be seen by members of the site. While *anyone* could log into Facebook, there is some confusion whether current guidelines should class this type of content as a 'members-only forum' and therefore – according to the guidelines – inaccessible to researchers without express consent.

There is, therefore, the additional question of whether the current guidelines are fit for purpose or whether there ought to be a tripartite distinction between public, open-private and closed-private. Here, open-private would include any information that could be acquired by any individual signing up without further permissions to areas of the website. Closed-private would include areas of Facebook where the researcher would need to be friends with the data subject in order to get their data or where there is some further

[4] ESOMAR, 'Guideline on social media research' (July 2011) http://www.esomar.org/uploads/public/knowledge-and-standards/codes-and-guidelines/ESOMAR-Guideline-on-Social-Media-Research.pdf (accessed 26/11/14)

[5] CASRO, 'Social Media Research Guidelines' (October 2011) http://c.ymcdn.com/sites/www.casro.org/resource/resmgr/docs/social_media_research_guidel.pdf (accessed 26/11/14)

requirement to being allowed access. This distinction is also tied up with user-expectation. For open-private, an individual would reasonably expect that anyone could view their information so long as they had an account. They are also given the option of making this a closed-private account in accordance with Facebook's terms of use.

2.3 Types of social media analysis

2.3.1 Social media monitoring

This is a form of passive data collection, but often the collection is automated. Social media monitoring is typically a fully or semi-automated effort to collect and understand vast volumes of social media data that human analysts would not be capable of processing.

Many social media websites allow for the data of their users to be accessed through an Application Programming Interface (API) – in line with their Terms of Use[6]. Where the API gives a limited sample (content available through the API is sometimes not available past a certain date), it is also possible for researchers to buy in sample from third-party suppliers. However it is essential to establish that third-party suppliers themselves are following laws and regulations that govern this data.

In most social media monitoring, this sample then gets fed into a separate tool. There are several of these available doing similar things. The basic task of all the tools, however, is to generate an aggregated dataset that indicates the volume of social media content mentioning a certain topic and then gives some basic details about this aggregated dataset. The more complex tools will then do some analysis about the language being used in the social media content. Was the content positive or negative? Were people talking about the brand or the product? This kind of natural language processing analysis is still experimental[7].

Natural Language Processing (NLP)
NLP combines computer science, mathematics and linguistics to attempt to train a system to categorise content into a pre-defined classifier of meaning. When applied to social media, this enables the informed researcher to analyse complex categorisations of social media content.

Some tools produce only aggregated and anonymous data results to the researcher, others provide access to raw data. Some also require the analyst the code raw data to help classify the data into relevant groups or themes.

A further variation of social media monitoring is network analysis, which looks at how key social media accounts interact with the public and with other key accounts. This then gives you information about a key account's connections rather than the content of what is being received and sent.

It is therefore common that social media monitoring will allow the analyst to see the author of the comment – indeed identifying the 'most influential authors' is often a common objective of this type of research.

Some advanced tools also attempt to infer demographics based on either metadata or an accumulation of contributions from an individual. It may also be possible for the analyst to be aware of the assumed age, gender

[6] For more information about APIs in social media monitoring see Demos' report 'Vox Digitas' (2014), p. 85

[7] Timothy Baldwin, 'Social Media: Friend or Foe of Natural Language Processing?', (2012) *26th Pacific Asia Conference on Language, Information and Computation*, p.58-59

and location of an individual author. In some tools, it will also be possible for the analyst to see the contents of the metadata.

There is an additional element to consider in social media monitoring in the future. Facebook have announced the release of 'topic data'[8]. This utilises both private and public content, but is data which is aggregated and anonymised before leaving Facebook servers. This protects the identity of the user as well as providing organisations with unrestricted access to data.

Finally, some market research agencies have set up panels of social media users where they have asked for their expressed consent for their social media data to be used for the purpose of research. However, given the incidence of any individual taking part in a specific discussion of interest in social media, this is likely to limit the ability to collect all relevant data to a research project.

2.3.2 Ethnographic analysis / online communities

There is considerable overlap with communities that have been set up specifically for research purposes ('online communities') and online ethnography and there are different definitions of both. Both have the same processes involved, whereby there is no aggregation of data and individuals are looked at as individuals. Online communities may be undertaken without ethnographic techniques involved (like discussion forums) or they may include them if the area of discussion is 'naturally' occurring (such as monitoring Facebook activity between friends).

The key distinction is that online ethnography is a research technique, whilst online communities are a facility for carrying out research. Online ethnography seeks to apply principles of ethnography to the online sphere. Online ethnography may focus on individuals or examine communities, the only pre-requisite is that the social media content occurs naturally in some way. Online communities, on the other hand, are forums in which people are observed interacting – these may be natural or artificial depending on the requirements of the researcher.

In both circumstances the researcher is viewing content that user expects to remain private. As such, as discussed later in the report, current market research guidelines require the research to identify themselves and seek consent directly from the user.

2.3.3 Co-creational research techniques

This is a new area of research primarily used in product development to bring the consumer closer to the product development. This is a wide area and examples are diverse, and often exist outside of social media. An example of how social media is used to achieve co-creation is provided by Burberry who have developed an online discussion forum where users and product developers have conversations about what kind of products they expect next season[9]. In a further example of organisations engaging directly with social media users, the Department of Energy and Climate

[8] 'Privacy-first approach to Facebook Topic Data', http://datasift.com/products/pylon-for-facebook-topic-data/ (accessed 08/06/15)

[9] Merve Nazliogli, '5 examples of how brands are using co-creation' (October 2013) http://www.visioncritical.com/blog/5-examples-how-brands-are-using-co-creation (accessed 26/11/14)

Change conducted a 'Tweetathon' in November 2014 to help promote action against climate change. During the Tweetathon, people were encouraged to Tweet their questions to partner organisations to help inform the case for change.[10]

Below is a table summarising the different varieties of social media research in relation to relevant privacy concerns.

Table 2.1 – Privacy characteristics of different kinds of social media research

	Social media monitoring	Ethnographic analysis	Online communities	Co-creational research
Is personal data[11] processed by the researcher?	Dependent on method(s) used	Yes	Yes	Yes
Can informed consent be acquired for the specific project?	Unlikely	Yes	Yes	Yes
Could research theoretically go on without subject's knowledge?	Yes	Yes	In some cases – not in researcher-created communities	No
Aggregated data output?	Yes	No	Unlikely	Unlikely
Access to public or private social media content?	Public and open-private[12]	Public, open-private and closed-private	Public, open-private and closed-private	Public, open-private and closed-private

[10] #BackClimateAction tweetathon https://www.gov.uk/government/news/backclimateaction-tweetathon

[11] A definition of what is personal data warrants further discussion in chapter 3 of this report.

[12] Though, as discussed above, Facebook Topics data does allow researchers to request aggregated anonymous data for private content.

Broadly speaking, whilst there are basic similarities with all these forms of social media research, there is are clear differences in the kind of data interrogated by the researcher and the challenges this presents for collecting data, dealing with personal data, gaining informed consent and reporting findings.

The Wisdom of the Crowd project is developing a social media monitoring tool for aggregated social media analysis. The remainder of this report therefore focuses on the key legal, regulatory and societal considerations in conducting ethical social media monitoring.

Chapter Conclusions

- There are a number of different types of social media research methodologies which give rise to some shared and some diverse ethical concerns. However, all methodologies need to consider how they are using people's data and engaging with participants during the research process.

- There is a distinction made in the market research literature between 'public' and 'private' social media; however this requires further clarification. One possible solution is to make a further distinction between truly open content (public), and content that can be accessed by anyone who creates an account to that social network (open-private). How does this relate to the processing of sensitive personal information? Is this, and should it be restricted by regulation?

- The Wisdom of the Crowd project is primarily focused on social media monitoring and the ethical concerns surrounding it. However, further research will also explore the public's understanding of different forms of social media research. Do people make a distinction between aggregation and individual study? And are people aware that this kind of research is taking place at the moment?

Legal considerations

3 Legal considerations

This section takes a brief look at the initial, bare-minimum requirements that must be fulfilled under the law in the UK by researchers wanting to undertake social media research.

3.1 Data protection act

Within the United kingdom, the Data Protection Act 1998 (the DPA) is the primary piece of legislation pertaining to the processing of personal data. However it is important to note that the applicable data protection/privacy law relevant to a specific social media site will be dependent on the country in which the social media service provider is established, which can normally be determined by checking the social media site's terms of use and/or privacy notice.

Personal data
Personal data is any data where a living individual can be identified from that data or from that data combined with other data. It is also considered personal data in any cases where the data might have the potential to impact upon individuals, families or businesses.

The DPA defines "personal data" as meaning any data relating to a living individual who can be identified using that data, or from that data and other information which the researcher is in possession of, or is likely to obtain'[13]; however the literature does not make explicit what this means in regards to social media. There is currently a lack of guidance from the market research industry on what should be defined as personal data within social media research. From a legal standpoint, our review suggests that it is clear that any social media data fields could potentially be personally identifiable. For example, the contents of a tweet text could identify the name of the author or disclose personal information about themselves or another person; pictures could also reveal personal information about the author. It is therefore extremely difficult to anonymise social media content, even the removal of key fields such as the author's account name may not be enough to guarantee anonymity during data processing. We would therefore suggest that even de-authored social media data could be subject to the DPA.

The DPA is structured around eight 'data protection principles'[14] that all organisations and their staff, agents and sub-contractors must comply with when processing personal data. This includes researchers carrying out social media research, as the DPA principles still apply, which may provide significant barriers to some types of social media research. Before any consideration of the implications of the DPA on social research, it is important to note that the DPA does include provisions providing exemptions from some or all of the principles and other provisions.

One such exemption covers processing of personal data for research purposes. Subject to certain conditions, the research exemption permits the further processing of personal data obtained for purposes other than research purposes (exemption to 2nd principle). It also permits personal data being processed for research purposes only to be retained indefinitely (exemption to 5th principle), and exempts personal data being processed

[13] Section 1(1), Data Protection Act 1998 (as amended)
[14] See Part 1, Schedule 1, Data Protection Act 1998 (as amended) or ICO page 'Data protection principles' https://ico.org.uk/for_organisations/data_protection/the_guide/the_principles (accessed 26/11/14)

solely for research purposes from the right of subject access providing the research results are anonymous. However the research exemption does *not* include any exemption from the requirement for the personal data to be **fairly and lawfully** processed, which means researchers carrying out social media research must still determine the fair and lawful basis for the processing of any personal data collected and processed during the study[15].

3.2 Social media research - DPA implications

Rearchers must ensure any personal data processed for a social media research study: is held securely; is only processed to the extent necessary for the research; that any transfers outside the EEA comply with the 8th principle and the personal data is accurate; and that where relevant, up to date. In addition, the most significant implication on social media research is the need to ensure any personal data collected for the research is processed fairly and lawfully. For more traditional research, this is achieved by assuring personal data is collected with the voluntary, fully informed consent of the data subject – the research participant.

For many social media research projects, especially social media monitoring projects, obtaining prior informed consent is not feasible. As a first consideration, researchers must first decide if the project can be carried out by collecting and processing anonymous data. If it can, the DPA does not apply and the researcher can proceed; yet we have already identified above that in practice, it is likely to be possible to guarantee anonymity even if key fields are removed. Researchers must use the definition within the DPA, together with the guidance issued by the Information Commissioner's Office[16] and European Article 29 Working Party[17], to determine if the data of interest to their study may constitute personal data. However, in many forms of social media research data collected will include personal data (perhaps unavoidably).

In theory, the easiest way to ensure compliance is, as far as possible, to avoid collecting or otherwise processing personal data for the study. Where this is not possible, the most common lawful basis for the processing of personal data held by social media service providers, including disclosure to researchers, is if the "*processing is necessary for the purposes of legitimate interests pursued by the data controller or by the third party or parties to whom the data are disclosed, except where the processing is unwarranted in any particular case by reason of prejudice to the rights and freedoms or legitimate interests of the data subject*[18]*.* "Fair processing" is addressed within the social media service providers' terms of use and/or privacy notice.

It is important to note that if the personal data includes information relating to racial or ethnic origin, political opinions, religious beliefs, trade union membership, physical or mental health, sexual orientation or sexual

[15] Section 33, Data Protection Act 1998 (as amended)
[16] ICO detailed guidance on determining what is personal data v1,1 https://ico.org.uk/media/for-organisations/documents/1554/determining-what-is-personal-data.pdf (accessed 03/06/2015)
[17] Opinion 4/2007 on the concept of personal data http://ec.europa.eu/justice/policies/privacy/docs/wpdocs/2007/wp136_en.pdf (accessed 03/06/2015)
[18] Condition 6, Schedule 2 Data Protection Act 1998 (as amended)

behaviour or criminal behaviour, this constitutes sensitive personal data as defined by the DPA, which cannot be processed on the basis of legitimate interests. Where social media research may involve processing of sensitive personal data, researchers must identify the relevant lawful basis for the processing from the conditions set out in Schedule 3 of the DPA.

The most likely condition that would allow use of sensitive personal data would be if the information had been made public by the individual in question. For example, in conducting analysis of Twitter, Demos used information in activists' profiles to identify them as active in a given political party[19]. This would be classed as sensitive personal information but given that it has been made public by individuals in their own profiles, this gives a lawful basis for its use.

This last point outlines a certain problem for certain kinds of advanced analysis. While the DPA allows use of sensitive personal information where this has been made public, there are legal barriers to information being used where it has not been made public. Technological means of deriving certain types of sensitive personal data – such as making estimates of ethnic group based on name – could rub against the DPA in two ways.

- Firstly, all data held about an individual must be accurate. Such advanced analyses are often only correct to a certain degree of accuracy, but holding incorrect information about an individual opens an organisation up to legal challenge.

- Secondly, the definition of what is classed as 'public' sensitive personal information should be considered. For example deriving sensitive personal information from a name may not be compliant; furthermore the distinction between public, open-private and closed-private is also important.

[19] Demos, 'CASM and Sunday Times: The Twitter election', http://www.demos.co.uk/projects/casmandsundaytimesthetwitterelection (accessed 08/06/15)

Journalists and the DPA

These restrictions can often look strict given how liberally media outlets use social media information. Journalists are subject to the DPA when using personal social media data, but they also have several exemptions, which allow them to process the data more freely. The journalists' exemption falls 'into four elements:

- The data is processed only for journalism, art of literature,
- With a view to publication of some material,
- With a reasonable belief that publication is in the public interest, and
- With a reasonable belief that compliance is incompatible with journalism.'

This exemption gives journalists more freedom to process personal data than is allowed to researchers.

3.3 Fair processing - terms of use and privacy notices

Upon signing up to a social media service, users subscribe to the service accepting the services terms of use and processing of their personal data based on the service's privacy notice, more commonly referred to as the service's privacy policy" In the past, these documents have been accused of being overly long and complex, with research showing that the majority of users do not read these documents[20].

In September 2014 Facebook simplified and shortened its privacy policy from 9,000 words to 2,700 in order to make the policy more likely to be read by users[21]. With this new policy came a 'Privacy Basics' tutorial to condense this 2,700 words down even further. The current trend in terms of use and privacy policies is towards something shorter and more comprehensible. Twitter's terms of use and privacy policy now contains 'tips' embedded throughout the document to give the key points.

Despite the issue concerning whether users are in fact *aware* of what happens to their data, part of their using social media services entails that they *accept* this. That is to say, their use of the service is taken as agreement to the privacy policy and terms of use.

[20] Ipsos MORI, 'Understanding Society: the power and perils of data' (July 2014), https://www.ipsos-mori.com/DownloadPublication/1687_sri-understanding-society-july-2014.pdf (accessed 26/11/14)

[21] Reed Albergotti, 'Facebook gives its privacy policy a makeover' (November 2014), *Wall Street Journal*, http://blogs.wsj.com/digits/2014/11/13/facebook-gives-its-privacy-policy-a-makeover/ (accessed 26/11/14)

Each social media website has its own terms of use or privacy policy (and often both). To illustrate how they pertain to social media research, section 3.3.1 outlines the relevant clauses in Facebook and Twitter as case examples.

3.3.1 Facebook data use policy

The Facebook data use policy outlines what data is collected about users and how it is used[22]. In short, this outlines the types of personal data that Facebook processes about its members, which is basically all activities that a user can engage in on the website.

It makes clear from the beginning that information the user chooses to make public can be seen by anyone. This information can then be associated with that user and is accessible to any application or website that a user or that user's friends use. In addition, public information is accessible to anyone using the Facebook API.

There is certain information that is always publicly available and can only become entirely private through account deletion. This includes your name, profile picture and gender. Most other information can be made private, and some information (such as Facebook messages) is always private. In these cases the data is not available to other applications or through the API. In addition, it is possible for users to explicitly opt out of all sharing of data with applications (including the API) by adjusting their settings.

The data use policy expressly states that data may be made available through the API for public information and does not limit the purposes of this. There is nothing in the data use policy that dictates what the public data may or may not be used for and so it is legitimate to assume that research (of all shades) forms part of that. At various points, the data use policy highlights that the data can be seen by all.

3.3.2 Twitter terms of service

The Twitter terms of service is much more straightforward with regard to what can be done with the data[23]. Use of Twitter is consenting to the collection and processing of the data by Twitter. Content published to Twitter is available to Twitter is a worldwide license to 'use, copy, reproduce, process, adapt, modify, publish, transmit, display and distribute such Content in any and all media or distribution methods'.

In a 'tip', Twitter 'encourage and permit broad re-use of the Content. The Twitter API exists to enable this.' Twitter also make clear that their license to

[22] Facebook, 'Data Use Policy' (November 2013), https://www.facebook.com/full_data_use_policy (accessed 26/11/14)
[23] Twitter, 'Terms of service' (September 2014) https://twitter.com/tos (accessed 26/11/14)

the content enables them to share all data posted with companies and other organisations.

As with Facebook, users can opt out of their data being available through the API by protecting their account in the privacy settings.

3.3.3 Implications for research

Both Facebook and Twitter's terms of use are very far reaching in what they can do with public data. The data can be shared and processed in ways that are close to being unlimited, therefore disclosure and processing of personal data for social media research purpose compliant with those terms will be fair and lawful. Together with the exemption accorded to research in the DPA, the terms of use and privacy notices seem to allow for fairly wide ranging processing of the data that can be disclosed to third parties for social media research. This is most obviously of benefit to social media monitoring methods, but also seems to allow access for all forms of social media research that are conducted on public social media.

In private social media, such as 'walled gardens', and data that the user restricts access to via the social media site's privacy settings, the direct involvement of the social media service provider in the study, or the consent of the user to disclosure may be required. Whilst the research exemption may still apply, the terms of use and privacy notices do not provide a fair and lawful basis for the disclosure to third party researchers.

3.4 Intellectual property rights including copyright

Any review, however brief of the legal considerations for social media research must at least reference intellectual property rights, including copyright. Virtually all social media sites will include in their terms of use provisions asserting the service provider's rights over the content. In many cases those terms of use will include restrictions on, or even prohibit automated harvesting of data from the service, except for any data collected under licence using the service provider's API(s). This clearly has implications for using tools for social media monitoring and the purchase of data from third parties. Researchers must check the target social media service terms and conditions carefully before their study commences. Researchers are also advised to seek contractual assurances from suppliers prior to purchasing social media data.

3.5 Future legislation

The DPA was first enacted in 1998 and whilst it has undergone changes since, it is still a document formed before the recent rise of social media. The technological landscape has been fast moving, and there will soon be new legislation for research companies to comply with in the future.

The EU data protection reforms are due to be completed later this year (2015)[24] with the new regulation potentially coming into force as early as 2018, but some reports suggest it may take longer to come into force[25]. The implications of this could be wide-ranging, and have the capacity to fundamentally change the individual's rights with regards to personal data. Further research needs to be conducted to establish what this could mean for social media research.

Conclusions for further research

- Though there is a clear legal definition of personal data, there is currently a lack of clarity as to how this applies to social media data. Removing the unique username of the author from the rest of the data may not be enough to guarantee anonymity if the content of the social media text contains personal information about the author or another person. To what extent is it possible to anonymise social media data? Should all social media content therefore be treated as potentially personal data and thus subject to the DPA?

- There are special exemptions for the processing of personal data under the DPA for research. However, it is important to acknowledge that these are different to exemptions for other disciplines such as journalism.

- Where exemptions apply for research, personal data still needs to be processed in a fair and lawful manner. Terms of use and privacy notices provide a key part of the structure under which data can be processing 'fairly'; however this may not apply to data which has been collected in private 'walled gardens'.

- Special caution ought to be taken where sensitive personal data is being provided, but which is allowed to be processed if declared public by the author. The status of this in somewhere like Facebook (which is open-private, rather than public) is important to establish in future research.

- Specifically, the possibility for using technology for deriving characteristics about an individual based on other content appear to have potential legal barriers that should be explored further. It will be a challenge to make sure that this data is 'accurate' and that it is does not process or identify sensitive personal information which has not been made public by the author.

- With new legislation around the corner, it is important to ascertain what the future has in store for research using social media.

[24] EppGroup, 'Data Protection Reform Timetable', http://www.eppgroup.eu/fr/news/Data-protection-reform-timetable (accessed 08/06/15)
[25] Steve Henderson, 'EU Data Protection Reform – What you need to know' (June 2014), http://www.dma.org.uk/article/eu-data-protection-reform-what-you-need-to-know (accessed 30th December 2014)

Industry regulations

4 Industry regulations

Market and social research is a self-regulating profession. This chapter outlines what initial standards are placed on research companies over and above those demanded by the law.

The picture of research regulators in the UK is complex, but this paper will simplify it greatly by addressing the regulations put in place by the Market Research Society (MRS) and ESOMAR. The MRS is the predominant market and social research membership body in the UK, whilst ESOMAR are a key international market research membership body. Both are considered to be the industry standard in their respective domains. Both organisations issue a code of conduct to their members, and also provide specific guidance on ethically complex areas such as social media research. Both organisations require members to be aware of, and comply with each other's and other national market research codes applicable to a particular project.

The chapter also considers guidance from toe Social Research Association and the Economic and Social Research Council.

4.1 Market Research Society (MRS) guidance

4.1.1 MRS guideline history

In August 2011, the MRS published a discussion paper on social media research[26], and its current thinking on the ethics of current practice. The paper took a negative view of some research that was being undertaken in the field of social media research, stating that: 'because information can be *found* it does not mean it should be *used* for research.'

The MRSB outlined five areas that should be taken into account by anyone engaging with social media research.

1. Expectations of users and the terms of use – the fact that social media websites outline how this data might be used and under what circumstances privacy may be guaranteed by the terms of use.
2. Laws relating to privacy – there are numerous pieces of legislation that provide certain rights of privacy to an individual's personal communications and data, the Human Rights Act[27] and the DPA being the most relevant of these. It also includes various pieces of copyright and intellectual property law, as well as database rights law. Other pieces of legislation taken into account include 'the Privacy and Electronic Communications Regulations 2003'[28] and the

[26] MRSB, 'Online data collection and privacy: discussion paper' (August 2011) https://www.mrs.org.uk/pdf/2011-07-19%20Online%20data%20collection%20and%20privacy%202.pdf (accessed 26/11/14)
[27] HM Government 'Human Rights Act 1998' http://www.legislation.gov.uk/ukpga/1998/42/contents
[28] HM Government, 'The Privacy and Electronic Communications (EC Directive) Regulations (as amended) 2003', http://www.legislation.gov.uk/uksi/2003/2426/contents/made, (accessed 08/06/15)

'Regulation of Investigatory Powers Act 2000'[29].

3. The law of copyright – which applies in general to the following online content: blogs/profiles/updates, photographs, video clips and sound recordings. All of these may be analysed in some form of social media research.
4. The law of data protection – that personal data must be lawfully collected and processed. This bans the use of data collected for research purposes to be used for marketing purposes, for instance.
5. The principle of voluntary participation – and that all research should be conducted with the consent of research subjects.

This last principle was taken by MRSB to be the overriding one as it is a core principle of any ethical research. Voluntary participation was taken to be an inviolable right of research subjects. In doing so they took the opinion that 'virtual life is real life' and stated that the same principles should apply to online research as offline. They insisted that researchers ought to obtain 'the informed consent of all persons from or about whom data is collected'.

There were 'vigorous' responses to the MRSB's discussion paper, which were addressed in a response by the MRSB in April 2012[30]. The majority of comments were about the fact that the data is publically available. In response to this, the MRSB said that the Data Protection Act does not draw a distinction between personal data in public and personal spheres. Personal data is personal data wherever it appears. This would cause significant trouble to researchers who would be processing public personal data in a way that treated it as public, published data.

A typical comment was: '"The suggestion that explicit research consent should be collected in order to track online buzz is out of touch with reality."' It was perceived that the MRSB's discussion paper would be preventing researchers using even the most basic social media monitoring. In their response paper, they stated this was not the case, but that they advocated the pursuit of data minimisation. This would ensure that researchers do all in their power to avoid collecting personal data where possible. This would allow for some kinds of social media monitoring, but it still remained an open question how much metadata could be collected in conjunction with social media content.

The response paper clarified many problems that had clearly been raised by the discussion paper. The requirement for informed consent remained when dealing with non-aggregated data such as online ethnography and natural online communities. However, social media monitoring was given more room as the absolute requirement for voluntary participation.

[29] HM Government, 'Regulation of Investigatory Powers Act 2000', http://www.legislation.gov.uk/ukpga/2000/23/contents (accessed 08/06/15)

[30] MRSB, 'Online data collection and privacy: response to submissions' (April 2012) https://www.mrs.org.uk/pdf/2012-04-04%20Online%20data%20collection%20and%20privacy.pdf (accessed 26/11/14)

4.1.2 The MRS 2014 Code of Conduct

These new guidelines took a couple of years to become part of the binding MRS code of conduct. In 2014 a new code of conduct was published, with two new clauses about informed consent[31].

'Informed Consent

> **16** Members must ensure that participants give their informed consent where personal data are collected directly from them.
>
> **17** Members must ensure that they have a fair and lawful basis for the collection and processing of personal data from sources other than the data subject themselves.'

The inclusion of consent in cases where personal data is not collected 'directly' from the research subject allows passive data collection to take place on social media without the consent of the individuals involved providing the data collection is fair and lawful. Clause 17 backs this up by saying expressly that processing personal data from third parties may be allowed so long as it is fair and lawful as required by the DPA. This gives social media monitoring the scope to research social media so long as the data collection is done in compliance with the DPA and the social media service's terms of use and privacy notices.

Alongside the 2014 Code of Conduct, the MRS updated their 'Online Research Guidelines'[32]. With reference to clauses 16 and 17, it outlines several implications. Firstly, personal data must not be collected directly from research subjects without their consent. This is of particular relevance to researchers who are conducting non-aggregated, non-anonymised research (such as online ethnography) and removes the potential for a researcher conducting systematic observations of individuals in private 'walled gardens' without asking for consent.

Following on from this, the guidelines insist that transparency must be maintained. This means where research is conducted within private social media, the researcher must make their presence known and request consent for this activity. This includes social media monitoring: passive personal data must not be collected from private social media without first asking for consent.

Social data platform
A social data platform can provide the 'sample' of social media content when the website's API is too limiting for the researcher. These suppliers download all the content being produced on social media websites and store it on their servers ready to be distributed.

[31] MRS, 'Code of Conduct 2014' (September 2014), p.13 https://www.mrs.org.uk/pdf/mrs%20code%20of%20conduct%202014.pdf (accessed 26/11/14)

[32] MRS, 'MRS Guidelines for Online Research' (September 2014), https://www.mrs.org.uk/pdf/2014-09-01%20Online%20Research%20Guidelines.pdf (accessed 26/11/14)

The Social Research Association – 'Modifications to informed consent'

The Social Research Association (SRA) is a membership body for those undertaking social research. In December 2003, the SRA published ethical guidelines. While these do not deal explicitly with social media research, they do make clear in these principles that there are some occasions where informed consent can be modified so long as the process are 'in the spirit of the principle' of informed consent.

The contexts where these modifications may be acceptable are where large datasets are combined in such a way that the combination may increase the chances of re-identification of individuals. In cases like this where there is a secondary use of records, the SRA guidelines say that 'where possible and appropriate' researchers should first of all try and seek consent from individuals for a new enquiry – but they concede that this is not required for the DPA as 'there are no additional consequences for the data subject'.

This has important similarities and differences with aggregated social media data. While there may be chance of re-identification, this is not due to the combination of datasets – and there is no research requirement that the data need be published. However, as with the secondary use of data, social media research often acquires data without it being possible to acquire informed consent to gather this data. As there are no consequences for the data subject (in the form of their information being made available in a form they did not intend such as for marketing), consent is not required under the DPA.

Finally, the MRS code of conduct stipulates that research with children (defined as those under the age of 16) must not be conducted without the prior consent of the parent. Although many social media sites prohibit individuals aged under 13 from joining or using the sites, it is still legitimate for those aged 13-15 to be users; moreover, it is likely that children lie about their age to gain access to social media sites.[33]

[33] As noted in the AAPOR report, in the US, the Children's Online Privavy Protection Act defines a child as under 13; ESOMAR considers those under 14 as children and those aged as 14-17 as young people, both which require a special degree of care. https://www.aapor.org/AAPORKentico/AAPOR_Main/media/MainSiteFiles/AAPOR_Social_Media_Report_FNL.pdf

ESRC – guidelines for academic researchers

The Economic and Social Research Council (ESRC) is a major public investor in academic social research. As an organisation, the ESRC has its own ethical framework with which all of its funded projects must comply.

These principles map closely to those asked of MRS members. Informed consent is normally required for ethical research, as is transparency of approach and a guaranteed confidentiality for participants. The ESRC guidelines introduce a voluntary principle for participants as well. At the level of these principles there is a broad alignment between ESRC and MRS guidelines.

4.2 ESOMAR guidelines

ESOMAR is an international regulator of market and social research. It produces a code that has been adopted by, or incorporated into the codes produced by national market research associations, including the MRS. As such, the code does not differ from the MRS', and so we will not go into that here. Instead, we will focus on the 'ESOMAR guideline on social media research' that has already been referenced in this paper[34].

This paper built on guidelines that were produced by ESOMAR as internet research became more relevant. The 'Guide for online research'[35] outlined the researcher's responsibility in regards to personal data and also what were acceptable and unacceptable practices involving research technology (unrelated to social media) taking into consideration data protection and privacy laws enacted globally at the time of writing. The 'Passive data collection, observation and recording' guidelines[36] discuss 'incidental data' as personal data that is produced as a by-product of internet activity. This may be extended to include metadata, which it allows to be processed so long as it is for the purposes under which it was originally collected.

The ESOMAR website also offers consent guidance based on the 'Passive data collection, observation and recording' guidelines. It says that things people do in a public place should not be inaccessible to the researchers, and that consent in public places is not always possible to achieve. In these

[34] ESOMAR, 'Guideline on social media research' (July 2011) http://www.esomar.org/uploads/public/knowledge-and-standards/codes-and-guidelines/ESOMAR-Guideline-on-Social-Media-Research.pdf (accessed 26/11/14)
[35] ESOMAR, 'Guideline for online research' (August 2011), http://www.esomar.org/knowledge-and-standards/codes-and-guidelines/guideline-for-online-research.php (accessed 26/11/14)
[36] ESOMAR, ' Passive data collection, observation and recording, (February 2009), http://www.esomar.org/uploads/public/knowledge-and-standards/codes-and-guidelines/ESOMAR_Codes-and-Guidelines_Passive_Data_Collection-Observation-and-Recording.pdf (accessed 26/11/14)

cases, a public notice should be made of the ongoing research – in the case of social media the terms of use may provide such a notice[37].

In July 2011 – the same week as the MRS published their online data collection and research discussion paper – ESOMAR published it's guideline on social media research. There was initially some concern that ESOMAR's guideline permitted greater flexibility than the MRS discussion paper.

In the MRSB's response to submissions[38], they respond to the comment that: '"…we oppose the MRS's standpoint as we understand it but would support guidelines in line with ESOMAR's."' This was a frequently stated concern. There was an assumption that the ESOMAR guidelines allowed researchers to collect personal data without consent being given. The MRSB pointed out that this was not the case given that ESOMAR state that in all cases, researchers must comply with the existing data protection laws in their country. In the MRSB discussion paper, at this time they saw no difference between ESOMAR's principles and their own.

The ESOMAR guidelines deal directly with identifiable data[39]. In line with the MRS pursuit of data minimisation, where automated monitoring tools are being used, efforts must be made to depersonalise what the researcher sees – such as usernames, photos and links to profiles. If the data is being collected manually – as in netnography – analysis must be done with depersonalised data.

There is a question raised by the ESOMAR guidelines about whether consent is required for public social media. Within the MRS guidelines this appears to be a requirement where possible, but ESOMAR allows for consent to be obtained either 'directly or under the terms of use'[40], subject of course to applicable law. The qualifier to this is that researchers must only report 'depersonalised data from social media sources'.

With public data, the ESOMAR guidelines accept that 'data cannot always be 100% anonymised'[41] by removing the username. If quoting particular pieces of content, it may be necessary to seek permission from the user to report on this. If this is not possible, they recommend a process of 'masking' comments to ensure their anonymity. Masking should occur wherever the research has not sought permission to republish the content and the content could be traced back to the author.

Masking

Masking is a process of altering raw data so that the meaning is maintained but it is not traceable back to the source. This may be changing just a couple of words to altering the language used in the content. ESOMAR guidelines suggest that the extent of the masking is at the researcher's discretion.

4.2.1 Implications for social media research

[37] ESOMAR, FAQs, http://www.esomar.org/utilities/help-support/faq.php?idfaq=40 (accessed 26/11/14)

[38] MRSB, 'Online data collection and privacy: response to submissions' (April 2012), p. 6 https://www.mrs.org.uk/pdf/2012-04-04%20Online%20data%20collection%20and%20privacy.pdf (accessed 26/11/14)

[39] ESOMAR, 'Guideline on social media research' (July 2011), p. 6, http://www.esomar.org/uploads/public/knowledge-and-standards/codes-and-guidelines/ESOMAR-Guideline-on-Social-Media-Research.pdf (accessed 26/11/14)

[40] ESOMAR, 'Guideline on social media research' (July 2011), p. 6, http://www.esomar.org/uploads/public/knowledge-and-standards/codes-and-guidelines/ESOMAR-Guideline-on-Social-Media-Research.pdf (accessed 26/11/14)

[41] ESOMAR, 'Guideline on social media research' (July 2011), p. 7, http://www.esomar.org/uploads/public/knowledge-and-standards/codes-and-guidelines/ESOMAR-Guideline-on-Social-Media-Research.pdf (accessed 26/11/14)

Both MRS and ESOMAR guidelines ensure that researchers using social media monitoring must limit the personal data they process to the bare minimum required for the research. In line with good research practice and ethical standards, this means using anonymised or pseudonymised data where at all possible; with personal data only being processed for research purposes where this is absolutely necessary. This means limiting the personal data that is collected in the process.

This not only presents a considerable challenge, given that personal information is disclosed in an unstructured manner on social media (i.e. in more places than just the author ID), but also goes against the grain of some of the aims of social media research. A key objective of social media research is often to identify 'key influencers', or nodes within a network. In reporting this to a client, individuals are identified as important. How does this fit with regulations that prevent identification of research individuals without explicit consent? Given that this type of analysis is being broadly used by social researchers, it requires further exploration.

Furthermore, it is at odds with the aim of researchers who are attempting to add rigour to social media analysis by enriching social media content with inferred demographics. This requires the direct processing of personal data.

The MRS allows for the collection of social media data in an aggregated form, but insists that this be anonymised insofar as it can be. The MRS also allows for the collection of non-aggregated social media personal data providing both the provider and the researcher have clearly identified the legal fair and lawful basis for the processing, and the results of the research are anonymised to protect social media users' privacy and confidentiality. There is a question here as to whether there is more that can be done within the law. If adequate notice is served through the terms of use for a public social media site, does that mean the information made publicly available on that site is published and so allow for further research processing of personal data? Alternatively, should all the data that comes in be anonymised before the researcher sees it, and if so, what implications does this have for the potential analysis of the personal data held within profiles made available to the general public?

Social media researchers require a further challenge in protecting both copyright and privacy. For example, there is currently little recognition that anonymising the publication of a Tweet may not only infringe the terms of use of Twitter's API guidelines but also the copyright of individuals who want their comments attributed to them. There is also no practical distinction in advice between comments made by individuals compared to companies or organisations. Should a piece of stakeholder work using social media analysis be bound by the same guidelines as a project which explores the public opinion of individuals?

Further guidance is also required as to how researchers should look to identify and analyse social media contributions from users who are likely to be under 16. Do they require separate parameters for informed consent? Can we identify these users, and if so should their data be excluded prior to analysis?

Conclusions for further research

- Revised guidelines for market research have begun to make a distinction in the consent process required for different types of social media research. Though it is clear that informed consent must be obtained where personal data is collected directly from individuals, there is greater flexibility in the method of consent required to conduct social media monitoring research. Nonetheless, this type of research is still required to have a fair and lawful basis for collection.

- However there is a lack of clarity from guidelines as to what should be considered as public or private data, what constitutes personal data, how to treat sensitive personal data, and whether it is possible to fully-anonymise social media data.

- There is a direct contradiction between some guidelines and the practice of conducting social media research. Most notably:

  - Does the industry requirement to limit the processing of personal data and publish anonymous findings prevent research identifying key authors or networks?

  - How should researchers mask social media contributions and still adhere to copyright legislation; and is there a need to differentiate between individuals and organisations or companies?

  - To what extent should personal data be processed to enrich the data with key demographics to help identify the profile of the data and differences between users.

  - How should the industry treat content from social media users under 16 (which currently requires parental consent)? Should attempts be made to identify these groups and remove them from the dataset, if so this would require the processing of personal information.

- Are these regulations sufficient for public trust to be maintained? Further primary research may be able to suggest how people might react if consent could or should be sought.

Responsibility

5 Responsibility

The previous two chapters dealt with the minimum for market and social researchers who hope to engage in social media research. In this chapter, we explore the consideration that it is not necessarily the case that the legal and regulatory framework provides the safeguards that the public hope to see.

It is the 10th principle of the MRS Code of Conduct[42] that says: '[r]esearchers shall protect the reputation and integrity of the profession'. This involves looking at the ethical side of the research and how procedures can be put in place to ensure that the flow of data is appropriate and does not violate the public's ethics[43].

Contextual ethics

Research into ethics around privacy has led to development of theories of privacy ethics. One of these, proposed by Helen Nissenbaum, is that data flow does not necessarily need to be *restricted*, but it must be *appropriate*. That is to say, it needs to flow in a way that the public finds acceptable.

Researchers following this model of public attitudes should focus on *how* and *why* data is being share and processed and not just *how much*.

Susan Brenner outlines an important conceptual distinction between 'individual' and 'societal' expectations[44]. On the one hand, an individual may signal their expectation that their communications may be private. This could be by joining a 'walled garden' for instance. There is separately, but related, to an individual's expectation, a societal expectation of whether communications may be private. Reasonable individual expectations may often be driven by the expectations that have been laid out by society.

This chapter will identify several areas of contention that legislation and regulation may or may not deal with. The key element is that they present areas for online researchers to move beyond what is required to establish a best practice that protects research subjects and the integrity of the discipline.

5.1 Anonymity

Anonymity has consistently been cited as an issue with social media research. There are two primary areas where anonymity is broken: during

[42] MRS, 'Code of Conduct 2014' (September 2014), p.13 https://www.mrs.org.uk/pdf/mrs%20code%20of%20conduct%202014.pdf (accessed 26/11/14)
[43] Helen Nissenbaum, 'Privacy in Context' (2010), p.2-3, Stanford University Press
[44] Brenner, *Cybercrime and the Law: Challenges, Issues, and Outcomes* (Northeastern University Press, 2012).

collection (by the researcher) and during reporting (by whoever views the report).

5.1.1 Anonymity during collection

This is primarily a problem with any method where consent has not been sought directly from the user. As seen in the industry regulation chapter, consent needs to be sought wherever possible[45]. Where consent cannot be sought the question remains over what level of anonymity should be required. Many social media monitoring tools contain means of anonymising social media content seen by the researcher[46] (by removing the author ID form the content), but this does not guarantee that no personal information will be contained in the text, and searching the content could still give you the identity of the user. What level of anonymity would the public look for, if any?

5.1.2 Anonymity during reporting

This is more contentious than whether or not the researcher can identify someone. Consider the situation where someone's tweet is attributed to them in a report about racism on Twitter – does this do harm to the respondent? Consider the likely scenario where after this report being published, the tweet is deleted by the user. Does it then need to be removed from the publication? The alternative would be to anonymise all tweets in a report. But this could potentially be an infringement of copyright – some people want their tweets to be attributed to them[47]. Again, where consent is sought this is not as big an issue, but some research needs to be undertaken into under what circumstances attribution can take place, especially where consent has not been sought.

5.2 Research with young people

A key issue mentioned in the literature[48], as well as in the Twitter terms of use[49], is that distinguishing between children and their parents is challenging online. Currently MRS regulations stipulate that no research should occur with participants under the age of 16 without first seeking consent from a parent or guardian. Email correspondence is not considered adequate and so the MRS guidance insists on more rigorous methods of communication with parents[50].

Despite this, the area that may cause more concern is that people of all ages can be using social media. Social media monitoring may be

[45] ESOMAR, 'Guideline on social media research' (July 2011) http://www.esomar.org/uploads/public/knowledge-and-standards/codes-and-guidelines/ESOMAR-Guideline-on-Social-Media-Research.pdf (accessed 26/11/14)
46For more information about APIs in social media monitoring see Demos' report 'Vox Digitas' (2014), p. 85
[47] NatCen, 'Research using Social Media; users' views' (February 2014), p. 7, http://www.natcen.ac.uk/media/282288/p0639-research-using-social-media-report-final-190214.pdf (accessed 26/11/14)
[48] MRS, 'MRS Guidelines for Online Research' (September 2014), https://www.mrs.org.uk/pdf/2014-09-01%20Online%20Research%20Guidelines.pdf (accessed 26/11/14)
[49] Twitter, 'Terms of service' (September 2014) https://twitter.com/tos (accessed 26/11/14)
[50] MRS, 'MRS Guidelines for Online Research' (September 2014), https://www.mrs.org.uk/pdf/2014-09-01%20Online%20Research%20Guidelines.pdf (accessed 26/11/14)

aggregating the data of children under the age of 16 – both Twitter and Facebook have a minimum age requirement of 13. Terms of use agreements will usually state that children must have permission of their parents to use the service. This could be taken as a form of 'parental consent' to the terms of use and thus to their data being processed for research purposes.

However, whilst this may be enough to satisfy the legal requirements, it may not be something parents are happy with. There is, indeed, no guarantee that a parent was present when the child was signing up for the social media website. Are the public happy with children and their personal data being included in these aggregated datasets? More research is required to establish whether more ought to be done to limit the presence of children in the monitoring process.

There is also a separate question here relating to whether a social media comment, however public, should be attributed in a report if the researcher cannot say for certain that the originator of the comment is over 15.

5.3 Informed consent

The industry regulations make clear that informed consent must be sought where possible[51]. This will mean that in cases such as netnography and where online communities are being monitored at the individual-level there should be a level of consent received from all parties. For social media monitoring and other types of data aggregation, this requirement is relaxed as long as the data is collected and processed in a fair and lawful manner.

Despite this, the views of the public on informed consent remains mixed. A NatCen report early in 2014 demonstrated that whilst many people consider public social media data to be accessible to anyone, there was still an obligation for the researcher to seek consent from subjects. Whilst many accepted that using automated monitoring software may make seeking consent impractical, NatCen found that there was a group that were concerned about their privacy in these circumstances. In qualitative discussion, they found that 'participants who wanted consent to be gained did not think the logistical burdens of doing so were a justification for not seeking permission.'[52]

It is important to establish where the public believe informed consent can be contravened, and simply because big data analytics is fast becoming the norm does not mean we should assume that attitudes are moving in step with this. Even where industry practice and the law both suggest that consent can be inferred from terms and conditions stipulations, the NatCen research shows that this is not necessarily in line with some users' expectations. Indeed, research into the subject of expectations of privacy on social media find that public attitudes are extremely fragmented on the subject.

[51] ESOMAR, 'Guideline on social media research' (July 2011) http://www.esomar.org/uploads/public/knowledge-and-standards/codes-and-guidelines/ESOMAR-Guideline-on-Social-Media-Research.pdf (accessed 26/11/14)
[52] NatCen, 'Research using Social Media; users' views' (February 2014), p. 27, http://www.natcen.ac.uk/media/282288/p0639-research-using-social-media-report-final-190214.pdf (accessed 26/11/14)

5.4 Processing of data

There is little discussion as to what is considered as acceptable processing of social media data, and what might go beyond users' initial expectations under the terms of service. A good example of this is the academic work being undertaken to infer demographics from metadata and from an individual's social media contributions over time. On the one hand, this work in crucial to add rigor and quality to the insight drawn from social media data as it helps identify the profile of any one conversation; on the other hand, users may be are unaware that this type of data processing is being undertaken and in relation to some demographics raises questions regarding the fair and lawful basis for processing what may be sensitive personal data. Various studies have found that it is possible to infer a lot of information about an individual based on meta-data analysis that they are unlikely to have reasonably expected to be taking place. There may therefore be new personal data *created* during he process of data analysis.

Similarly, a combination of human coding and natural language processing is used by researchers to help conduct thematic analysis and identify patterns within the data. Further research is required to understand users concerns around these processes – this is likely to be more sensitive for emotive topics, for example the publication of a report which uses natural language processing to identify discriminatory language.

5.5 Public awareness

There is a final question to be raised around whether users of social media are aware of what data is being collected about them. The NatCen research shows that awareness of the public nature of social media is high. Participants in the research were quite open to the idea that their activity was open for all to see. However, we saw earlier that many individuals do not read the terms of use of social media sites, and it may not be clear to people how much of their additional data is available through the API. For instance, it is possible that a user's location may be attached to social media content, despite them never having input this information themselves. The wide-ranging extent of metadata is not something that can be assumed to be widely known. This is a problem even where consent is attained – are the participants aware that the researcher can see both their Facebook posts and where they were posted from? There is also evidence of social media users becoming more aware of the potential negative aspects of sharing informaton publicly, with more users restricting social media content to authenticated users, and/or using privacy settings to further limit access to only those social media service members the user grants access.

Case study: Samaritans Radar

The Samaritans came under fire recently for producing an app that warned social media contacts if someone they already followed on Twitter was using suicidal language. This caused a widespread concern among the media and privacy campaigners for a number of reasons, but a key one was privacy. The primary concerns were:

- stigmatisation of individuals (and potential bullying);
- that there was no opt-out mechanism for people who did not want to have their data collected (there was an opt out but you had to know about it to opt out); and
- that it may breach the DPA.

Whilst the ICO displayed concern at the way the app was being used, it was not immediately obvious that it contravenes the DPA. It was, however, received with significant media coverage. Something in the processing of public social media data broke the perceived 'appropriate flow' of data, and it is important to explore with the public what this is and where the sensitivities lie.

Lastly, there is a lack of public awareness surrounding the kinds of data analytics that can be done with big datasets[53]. For instance, the tools that are being developed now for social media monitoring are capable of discerning some individuals' locations *even if* they do not have geo-tagging enabled.

5.6 Responsibility and future-proofing

This public awareness dovetails into future-proofing against legislation and regulations that are to come. In November 2014, the Science and technology Committee published a report exploring the responsible use of social media data[54]. In a section dedicated to informed consent, the Committee voiced their concern that the public do not necessarily see terms of use as a proxy for informed consent. The Committee calls on companies to be more transparent in how they intend to use data, and to communicate more effectively.

This may be taken as a sign that awareness of the Terms of Use may become more widespread as companies begin to take on board these recommendations. As researchers, it is important to explore ways that we can be transparent too in the way that we are using this data. This will also allow us to be ahead of the curve when updated legislation comes in – ensuring both a sense of responsibility towards the public and a methodology proofed against future changes in the legislative landscape.

[53] For details on the Samaritans case study, see Jamie Bartlett, 'Below the Samaritans Radar' (November 2014), http://www.demos.co.uk/blog/belowthesamaritansradar (accessed 26/11/14)
[54] House of Commons, Science and Technology Committee, 'Responsible use of data', Fourth Report of Session 2014-15 (November 2014)

Conclusions for further research

- There is an appropriate flow of data and to contravene this appropriateness would mean to contravene a public ethics. This has the potential to damage the reputation of research.

- It is also clear that public ethics does not necessarily align with the regulatory and legal frameworks. Though there is some consensus that public opinion is not binding, it should have a central position in any research ethics.

- The concern about doing research with under-16s should also be examined further. This exists in research regulations for good reason and exploring the potential of suspending these regulations for the case of social listening should be looked at with the public and experts.

- Getting some more complete answers on where exactly the boundaries of the appropriate flow lie is vital to this project. Research should look at the ethical issues outlined in this chapter. Themes for exploration would be:

    - informed consent and whether the public thinks it can or should be sought;

    - the role of anonymity at different stages of the research process, including how best to report social media data;

    - acceptable levels of data processing;

    - the potential for capturing the views of under-16s, and the role of parental oversight; and

    - public awareness of the terms of use and understanding of what their social media data may be used for.

- Whilst some kind of exploration of the public attitudes towards social media monitoring is key, identifying all of the boundaries in the appropriate flow of information is, inevitably, an extensive task that will not be definitively completed within this project.

Steven Ginnis
Ipsos MORI
Steven.ginnis@ipsos.com

Harry Evans
Ipsos MORI
Steven.ginnis@ipsos.com

Ian Barker
Ipsos MORI
Ian.barker@ipsos.com

Jamie Bartlett
Demos
jamie.bartlett@demos.co.uk

Ipsos MORI
79-81 Borough Road
London SE1 1FY

t: +44 (0)20 7347 3000
f: +44 (0)20 7347 3800

www.ipsos-mori.com
www.twitter.com/IpsosMORI

The Social Research Institute works closely with national governments, local public services and the not-for-profit sector. Its c.200 research staff focus on public service and policy issues. Each has expertise in a particular part of the public sector, ensuring we have a detailed understanding of specific sectors and policy challenges. This, combined with our methodological and communications expertise, helps ensure that our research makes a difference for decision makers and communities.