Wisdom
of the crowd

US
University of Sussex

DEM☉S

CASM
Consulting

Ipsos MORI

#SocialEthics
a guide to embedding ethics in social media research

Harry Evans; Steve Ginnis; Jamie Bartlett   12 November 2015

Summary

This report on social media research ethics is a 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, CASM Consulting LLP, Demos and University of Sussex have collaborated in this project to critically examine commercial possibilities for social media research.[1]

One of the focuses of the Wisdom of the Crowd project is to examine the ethical landscape surrounding aggregated social media research. In spring 2015, the first publication of this ethics strand contained a review of the legal and regulatory framework for using social media in market research[2]. This second and final report builds on these findings, presenting our conclusions from quantitative and qualitative primary research with stakeholders and social media users, and outlining our recommendations for how the research industry should look to proceed if it is to be at the forefront of using social media data in an ethical way.

The scope of the ethical review is focussed on large-scale, aggregated analysis of social media data – sometimes referred to as 'social listening'. We regard this kind of research as potentially fruitful in the social insight it can be provide, and we are generally excited about the possibilities for this a new social research methodology; however, we are concerned that the guidelines for ethical best practice are incoherent and inadequate. The volume of data collected through this method presents barriers to traditional ethical research frameworks; this new kind of research also fits into the wider ethical context of using algorithms to analyse people's personal data. Consideration therefore needs to be given to how this kind of research can be conducted ethically.

Social media research is a relatively new discipline and does not always fit easily within existing guidelines which cover market research. Current guidelines for researchers are not always comprehensive, and sometimes conflict with one another. Moreover, guidelines and legislation do not always work well together, and sometimes these regulations have not been designed with social media research in mind. The review presented within this document attempts to fill some of the gap about what best practice looks like for researchers conducting social media analysis.

The conclusions and recommendations here are aimed primarily at those working within regulatory and legal frameworks within UK market research, and thus under the regulation of the Market Research Society (MRS) and the Data Protection Act (DPA). However, it is hoped that the learnings and practices presented here can be used to inform this debate in other jurisdictions and disciplines.

---

[1] Further information about the Wisdom of the Crowd project can be found here: https://ipsos-mori.com/wisdomofthecrowd. The project is funded by the former Technology Strategy Board (TSB), Engineering and Physical Sciences Research Council (ESPRC) and Economic and Social Research Council (ESRC).

[2] All Wisdom of the Crowd publications can be found here: https://www.ipsos-mori.com/ourexpertise/digitalresearch/sociallistening/wisdomofthecrowd/publications.aspx

The ethical review of this project consisted of three stages:

1 Secondary research: a review of the current literature about social media research and its ethical, legal and regulatory implications.

2 Primary research: interviews and discussions with experts and users to understand more deeply some of the issues raised at stage 1. The primary research informed our understanding of the issues as well as providing a basis for the recommendations.

3 Conclusions and recommendations: combining the findings of stage 1 and 2 in order to make recommendations for best practice for researchers conducting social media research.

The context, and need for a new form of ethical social media research

- **Social media analysis is a valid and important research methodology.** Like any other methodology it is not well suited to all research questions; moreover, it should not be used as a cheap and accessible proxy for drawing conclusions on the general population. However, there is immense value in delivering insight from research into the content, engagement, and relationships generated by social media itself. This type of research can be conducted in an ethical way.

- **All social media analysis should be conducted to the highest quality and with the same ethical considerations regardless of who conducts the research.** Whilst social media analysis conducted across government, academia and journalism will be subject to different ethics structures, we note that a considerable amount of social media analysis in the UK is conducted outside of any formal ethical structures. This includes social media analysis conducted by insight and communications departments of private sector brands, and the services provided by social media analytics platforms. According to public record, none of the major social media analytics platforms operating in the UK have signed up to the ethical code of the MRS. Though recommendations of how best to guarantee this are out of scope for this project, it is crucial that the UK government and regulatory bodies consider how best to ensure that social media analysis is being conducted within the best interests of social media users.

- **Professional market research should play a prominent role in social media analysis.** The professional market research industry prides itself on quality and ethics, and should continue to set be at the forefront of developing best practice for all research methodologies. The commissioning of a professional research agency assumes that the work will be conducted to legal and regulatory principles that underpin the industry. Within social media analysis, this requires standards of data collection, analysis and reporting that are distinctly different from an individual accessing publically available social media data to draw their own conclusions. Currently, not all research agencies and social media analytics tools are adhering to these standards; research methodology is being led by what is technically possible, not always what is ethically appropriate.

- Researchers should apply the same ethical principles to all projects, regardless of whether they have access to raw social media data, or use a third party tool to collect aggregated and anonymous findings. It is the responsibility of the researcher to ensure that data has been collected and analysed with the same respect for legal and regulatory principles, and within expectations of the user.

- Whilst there is a fair and lawful process for analysing social media data on quantitative scale, this is not synonymous with user consent to be included in a research project. By signing up to terms and conditions of some social media sites, users are technically agreeing to social media research taking place; however this is not the same principle as informed consent applied to more traditional research methodologies. Each individual social media project therefore needs to take appropriate action to consider the rights and expectations of social media users during the collection, analysis and reporting of social media content.

- There can be no guarantee that personal data will not be collected during social media research. By its very nature, social media data is largely open text, unstructured data. Even if attempts are made to withhold meta data fields (such as author, age or gender) from analysis, it is highly likely that data personal to the author or another individual will be present in open text content. It should therefore be assumed that personal data will be processed during the project, and therefore subject to the Data Protection Act (DPA). This applies to the collection, processing and storage of social media data.

- There can be no guarantee of full anonymity within social media research. Even though it is possible to report at an aggregate and anonymous level, it is not possible to present raw anonymous content to the analyst, client or reader. If the author field is removed, it would still be possible to search for the content online; moreover, it is unlikely that 'masking' content can fully guarantee that the author cannot be identified. This makes it all the more important to take steps to minimise the chances of identification at each stage of the project, even if this risk can never be fully nullified.

- Social media analysis best practice should be continuously reviewed alongside changes in technology, legislation and use expectations. As new social media platforms emerge, and trends in social media change, so too should guidelines for how to conduct ethical social media analysis. For example, further work is needed to provide guidance on the analysis of images generated through social media, and on practices known as 'scraping' where data is collected without access to a formal application program interface (API).

API
An Application Program Interface is the most common way of downloading social media content for research purposes. Set up by social media organisations, APIs act as portals for accessing content in bulk.

Key challenges in the current legal, regulatory frameworks

- **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.

- **There is a lack of clarity in the guidelines for researchers** on: how researchers should consider research with privately shared data presented at an aggregated and anonymous level; what constitutes 'personal data'; how to treat sensitive personal data; and whether it is possible to fully-anonymise social media data.

- **There is also currently no guidance on how research should be conducted bearing in mind that under-16s may be included in the research.** While traditional forms of research have stricter safeguards on research with young people, this has largely been ignored within social media research.

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

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

  o How should researchers mask social media contributions and still adhere to the brand guidelines of social media organisations. Is there a need to differentiate publication of social media content between individuals and companies or organisations?

  o 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.

- **There has been little research with the public undertaken to date on this topic,** but that which has been done has mostly focussed on general concerns about data and usage of social media data rather than the specific usage of the data for research. Research with the public has shown up annoyances with data used for targeted marketing as well as a concern that the terms of use are insufficient for informed consent. However, this research has also demonstrated that awareness of the public nature of social media is high, even if the understanding of the wide-ranging terms of use is low.

- **It is clear that public perceptions of how their data should be used do 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.

- While public opinion is not necessarily binding when it comes to research, **researchers have a responsibility to identify the concerns and fears of the public in how their social media data is used**. Other guidelines are in place for academic researchers, and these provide a good basis for commercial research, but these do not necessarily fit to transpose in their entirety onto a commercial environment.
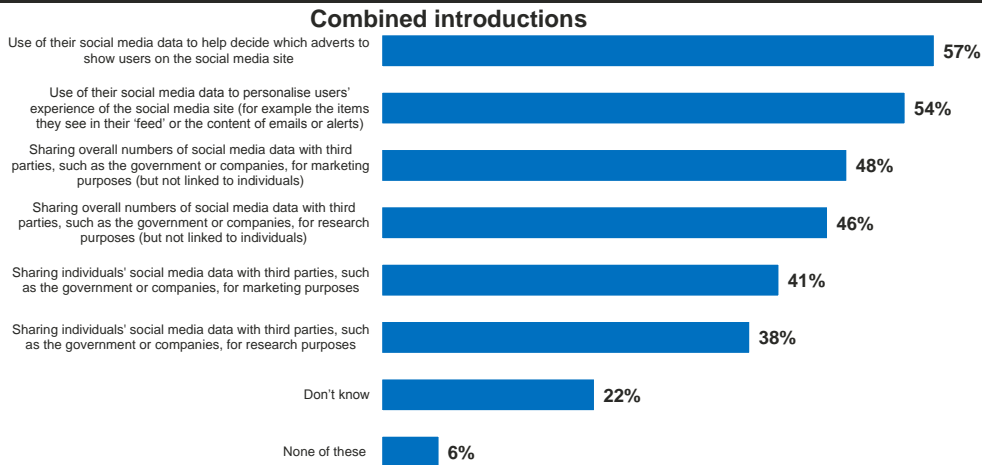
Findings from primary research with the public

It was felt that in order to fundamentally future proof and inform our understanding of the ethical issues at hand it was vital to conduct research with the public. Three kinds of research were carried out with the public in order to support these objectives:

1 An online quantitative survey of 1,250 adults aged 16-75 in the UK asking about people's attitudes towards possible uses of their social media data, and specifically, the value of social media research;

2 Three qualitative workshops in which participants discussed use of social media content, and the principles of ethical social media research. Two of these groups were carried out with adults, and one was carried out with 13-15 year-olds.

3 Statistical analysis (conjoint analysis) undertaken as part of the survey, during which respondents were asked to imagine they were on an ethics board and mark whether they would be likely to approve a series of social media research projects with different features.

All fieldwork took place in July and August 2015. More details about the conjoint and quantitative studies can be found in the appendices of this report. In summary, the survey found that:

- **Fewer than two in five people (38%) said that they thought sharing social media data with third parties for the purposes of research currently happens under the terms and conditions they sign up to on social media sites.** As well as a lack of awareness, research with social media data was also one of the least popular activities on the list. When asked to select activities that they thought *should not* happen, **a majority (60%) thought that data should not be shared with third parties for research purposes** under the format of terms and conditions. A third of people thought that sharing overall numbers (i.e not linked to individuals) for the purposes of research should not happen (32%).
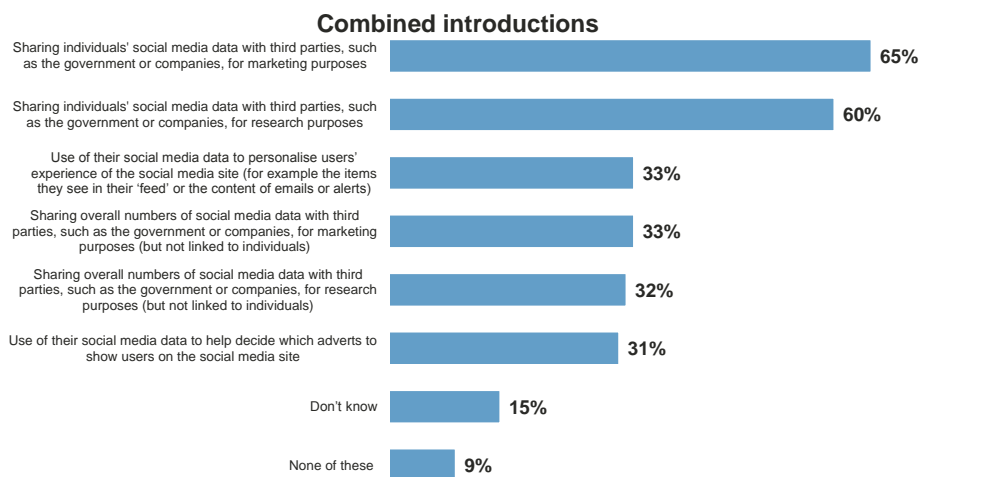
Which of the following examples of using people's social media data do you think currently happen under the terms and conditions of social media sites?

Combined introductions

Use of their social media data to help decide which adverts to show users on the social media site	57%
Use of their social media data to personalise users' experience of the social media site (for example the items they see in their 'feed' or the content of emails or alerts)	54%
Sharing overall numbers of social media data with third parties, such as the government or companies, for marketing purposes (but not linked to individuals)	48%
Sharing overall numbers of social media data with third parties, such as the government or companies, for research purposes (but not linked to individuals)	46%
Sharing individuals' social media data with third parties, such as the government or companies, for marketing purposes	41%
Sharing individuals' social media data with third parties, such as the government or companies, for research purposes	38%
Don't know	22%
None of these	6%

Base: 1,250 GB adults   Source: Ipsos MORI

And which, if any, do you think the following examples should not happen?

Combined introductions

Sharing individuals' social media data with third parties, such as the government or companies, for marketing purposes	65%
Sharing individuals' social media data with third parties, such as the government or companies, for research purposes	60%
Use of their social media data to personalise users' experience of the social media site (for example the items they see in their 'feed' or the content of emails or alerts)	33%
Sharing overall numbers of social media data with third parties, such as the government or companies, for marketing purposes (but not linked to individuals)	33%
Sharing overall numbers of social media data with third parties, such as the government or companies, for research purposes (but not linked to individuals)	32%
Use of their social media data to help decide which adverts to show users on the social media site	31%
Don't know	15%
None of these	9%

Base: 1,250 GB adults   Source: Ipsos MORI

- When asked to review how likely they would be to approve a social media research project,[3] the average approvability score for a project was 5.02 out of 10. Furthermore, 41% gave an average score of 4 or below, which is a broad a proxy for unlikely to approve social media research projects. This shows that there are still a lot of safeguards that need to be put in place to introduce a broad-based trust in social media research happening. This is backed by the reactions in our workshops, which demonstrated a real distrust of organisations using this data without seeking their direct consent and a feeling that they were losing control of data that is being shared on social media.

- Our study suggests that the biggest factor in the likelihood to approve a social media research project is whether the social media data is already publically available prior to inclusion in the research project. This factor

---

[3] On a scale of 1-10 where 1 is 'would definitely not approve' and 10 is 'would definitely approve'.

explains about a quarter of the variation in approval seen for different combinations. The level of anonymity was also a big part of these decisions, along with who the project is for and how personal the information that is being used is.

- **The statistical analysis shows that people opted for data minimising options over projects with more data seen by the researcher.** The acceptability was greatest for options where the names were not visible to the researcher, but there was no great difference between the researcher looking at de-authored data and overall aggregated numbers being provided to the researcher.

- **The analysis also reiterated the message that context is key when dealing with attitudes towards data**. However, while previous work has underlined the importance of *who* is doing the research, *what* they are doing it for and the *kind* of content being looked at, the conjoint suggested that the methodological context is key.

- **These fears were also voiced in our workshops with users.** The concerns of participants were often based in a fear that they had lost control over their data. They saw their data being used for research as *just another thing* that people wanted to do with their data, though they were often unable to articulate exactly why they did not want their social media information being analysed. Important factors that could make participants more comfortable included a transparency with what their data was being used for, a beneficial end purpose and an opt-in/opt-out mechanism.

- While the statistical analysis did present permission to use data as a significant factor, it was lower than many other factors. Within this, **relying on the terms and conditions as sufficient for consent appeared to have the least approvability, while providing an opt-out for the research (or, even better, an opt-in) was considered more preferable.** Participants in the workshops were also keen to reintroduce at least an opt-out system for their social media data.

How likely, if at all, would you be to approve the following research project on a scale of 1 to 10, where 1 is 'would definitely __not__ approve' and 10 is 'definitely __would__ approve'.
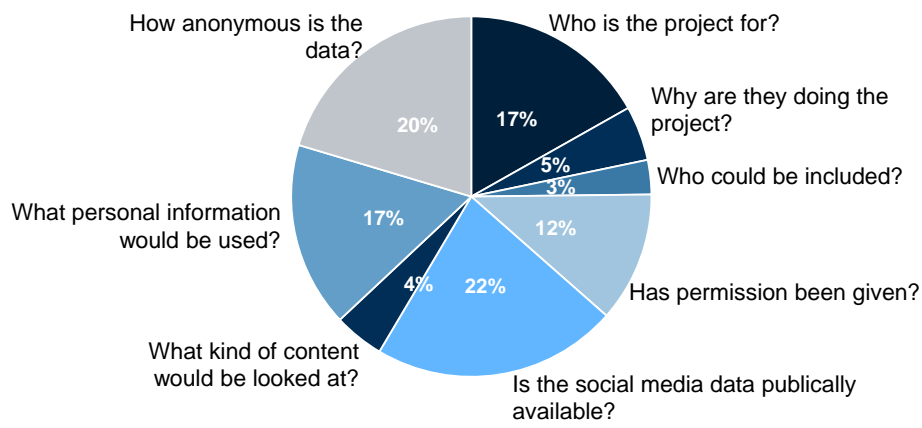
Base: 1,250 GB adults   Source: Ipsos MORI

Recommendations for embedding ethics in social media research

In light of the low level of awareness and trust of social media research among social media users, and the clear priorities for how best to improve the acceptability of social media research projects, we have developed 19 recommendations to regulators and social media organisations. These recommendations offer a series of sensible, positive, and practical steps for improving ethical standards in social media research. The full rationale for these recommendations is outlined in section four of the main report, describing the thought process and evidence base behind each of these suggestions.

The recommendations in this report are aimed primarily at researchers under the regulation of the Market Research Society (MRS) and the Data Protection Act (DPA). Ipsos MORI in the UK will move towards adopting these recommendations in our practices, recognising that this change cannot happen overnight. The changes that Ipsos MORI is setting out will require trialling in order to ensure they are practicable and useful in ensuring an ethical best practice for social media research. Ipsos MORI also recognises that implementing these recommendations will be iterative, and that there may need to be changes in practice, not least because digital communications develop extremely quickly. It is not assumed that all researchers outside of research organisations such as Ipsos MORI will be able to adopt these recommendations wholesale, but we hope this document contributes to debates on this topic in related sectors.

Demos recognises that social media research is developing rapidly, and this paper is an important step in understanding how it can be best conducted legally and ethically. They are therefore currently developing a code of conduct that draws on this report and makes it applicable for think-tank research.

Recommendations 1-2: boost awareness to build trust

Recommendation for researchers

1   Researchers should aim for transparency when conducting research projects that use social media analysis. Where possible, details of the project should be made available online, providing an explanation of whose data will be collected and for what purposes it is being used. With respect to commercially sensitive information and only with consent of the client, an abstract for each project should be made available online, including, where appropriate:

   o   The commissioner of the research;

   o   the purpose of the research;

   o   what kind of data will be collected and how; and

   o   whether the data will be published and in what format.

Recommendation for social media organisations

2   Social media organisations now have a raft of preconceptions to overcome about the difficulties of reading privacy documents. Ethical research relies in part on people understanding how their data might be shared. Work has been done to simplify these documents and shorten them, but less has been done to advertise how their data might be used. It is recommended that social media organisations continue to review terms and conditions so that the possible uses of the data easier are easier to understand, including use of social media content for research purposes.

Recommendations 3-5: the option to opt-out

Recommendation for researchers

3   Researchers should work to put in place systems to allow for users to opt-out from individual social media projects, or to opt out from all social media research conducted by that organisation. Compliance teams already have similar systems in place for blacklisting numbers and email addresses for traditional research. At the very least, an email address should be provided for people to easily put their profile or account on a list which will be excluded from future social media research by that research organisation.

It is acknowledged that researchers are only able to guarantee the execution of the opt-out when working with raw social media data. The full implementation of an opt-out mechanism requires the collaboration of social media analytics platforms, who often provide social media content to research organisations either as raw data or in an aggregated and anonymous format. One solution would be for research organisations to provide analytics platforms with a list of users to be removed prior to receiving either raw content or aggregated data.

Research organisations need to work with analytics platforms to ensure they understand the importance of giving the opportunity for privacy-conscious research subjects to remove themselves from analysis.

Recommendation for regulators

4   An opt-out for this kind of research being maintained by separate research organisations is over-complicated for members of the public. We call on regulators to look into creating an industry-wide opt-out mechanism that would work across the research organisations registered with them.

Recommendation for social media organisations

5   The final step would be to encourage social media organisations to build in an opt-out for research. We suggest that social media organisations explore ways of incorporating consent-for-research opt-out into their system. This might take place at sign-up, or be available in the account settings. While it is understood that such a system may not be technically desirable, it would likely be an

unobtrusive and an effective means of building a broader consent basis for research.

Recommendations 6-7: minimising unnecessary personal data collection

Recommendation for researchers

6 Researchers should look to put in place restrictions on what the researcher can see in a social media analysis tool depending on the scope of the project. This will involve working with the designers of the technology to ensure that data can be removed if not required by the project. It is important to try and move towards a culture of questioning whether the data that is being collected is really necessary for a research project. Each project should be engaging in its own ethical review (see recommendation 19) to establish potential harms to the data subjects, as well as what data is necessary to answer the research questions.

Examples of data minimisation for a project might include, but is not limited to:

o   removing the author's name and @tag from the researcher's sight;

o   stripping out other data that is downloaded in the content of a social media post, such as named persons or place names;

o   removing metadata that is not relevant for the purposes of a research project, such as GPS data that might be attached to the social media post;

o   creating generalised groupings of data instead of analysing specific data. For instance, generalising locations by cities instead of exact street locations; and

o   identifying where the need for creating derived characteristics is crucial to a project, and not running these algorithms as standard.

Recommendation for social media organisations

7 We want to work with social media organisations to make the process of data minimisation easier for researchers. Currently, downloading data from the API makes it impossible to avoid downloading data such as names, locations and other details, which are then immediately subject to conditions of 'further processing' under the DPA. For some of these fields, social media organisations should explore the possibility of limiting some of the fields of data that are downloaded.

Recommendations 8-10: removing under-16s from social media research

Recommendation for researchers

8   Under-16s remain a particularly vulnerable group social media. In order to maintain this principle, researchers should make efforts to remove under-16s from the data. At the moment there is no way of excluding this group from the data collected from social media through their APIs. Until such a function is in place, the second best option is deriving the age of this group from the content they post. This will give an imperfect, probabilistic estimate of whether a user is under 16 so that a great deal of young people can be excluded from the research.

While this method appears to go against recommendation 3 - which asks researchers to question the need to derived characteristics – we would suggest it is acceptable to infer age for the specific purpose of removing a vulnerable group from the dataset.

There may be some very legitimate reasons to purposefully include or even target children under 16 in social media research – for example engagement in a wellbeing campaign or review of support services used on social media. However, these exceptions to the principle should be considered carefully and only undertaken if approved by an internal ethics review.

Recommendation for regulators

9   There are currently no suggestions from regulators on how removing under-16s from the research can be done. Regulatory bodies need to provide details with the means to comply with the principle of avoiding conducting research with young people without parental consent.

Recommendation for social media organisations

10 Ethical researchers could be assisted by social media providers in this endeavour. Where aggregated and anonymous data is being provided (to either third party analytics platforms or directly to researchers), it should be possible to request that those known to be under 16 are removed from the dataset.

Recommendations 11-16: permission for publication

Recommendation for researchers

11 All social media research projects should question whether there is a need to publish verbatim content, and ask whether publication of aggregated and anonymous data would be sufficient to answer the research question. Any projects that wish to show verbatim text should first seek approval to do so through an internal ethical review. All steps should be taken to reduce the risk of harm to the participants.

12 If a project wishes to cite examples of content as illustration in a report or presentation, researchers should aim to contact social media users to ask them if they would be happy for their content to be cited. This would mirror existing good practices in other disciplines, as well as matching expectations of the public, who expect anonymity as standard. To adhere to brand guidelines, where consent has been given, researchers should keep the author next to any content, and avoid adjusting the text of the content.

13 Where a project wishes to show verbatim comment as part of an automated dashboard that provides a live stream of content on a certain topic, and often filtered cut by certain variables, the issue of anonymization is distinctly more difficult to implement. Where there is a clear need demonstrated from the client to include these which has been approved be an ethical review process, dashboards should be password protected and de-authored as minimum to maintain some level of anonymity to the wider public.

14 In addition, where it is agreed that raw content will be seen by the client, researchers should consider putting in place an agreement with their client that they will not try to re-identify de-authored individuals' from the dashboard. There are models for these kinds of agreements in other kinds of research.

Recommendation for regulators

15 Regulators should attempt to form a clear definition for what constitutes a 'brand' on social media. For example where social media accounts have a number of followers or friends that can be counted, is it reasonable to expect that those with a significantly large volume of followers expect less privacy than someone with fewer followers? Is it also reasonable to suggest that public accounts of prominent individuals, such as MPs or company CEOs, should be treated differently?

This definition of a brand can then be used to identify individuals in a network who are of special interest to a commissioner of research and that can be named in a report. We acknowledge that this might sometimes appear arbitrary but regulators should focus on balancing expectation of privacy with genuine research interest in high-profile users.

Recommendation for social media organisations

16 We would like to see social media organisations adopt developer guidelines that allows researchers the flexibility to make changes to the social media content they collect for the purpose of research. This would give researchers the ability to mask and anonymise content from individuals when republishing content in a research project. Care would be taken by researchers to uphold the meaning and content of the social media post, and attribution would be given to the platform brand. We also seek guidance from social media organisations as to what format they would ideally like these masked, modified social media posts in, and how they should look.

Recommendations 17-18: Defining "private"

Recommendation for researchers

17 It is the responsibility of researchers to have a sound understanding of whether the data they have collected, analysed or enriched consists of public or private data. Analysis of private content should only be conducted with approval from an internal ethics review.

Recommendation for regulators

18 It is not clear from the primary research whether users are happy with data they share privately being used for research, even if new functionality allows this to be conducted at the aggregated and anonymised level. We recommend that regulators provide further clarification on extent to which this type of approach would comply with current research guidelines.

Recommendation 19 – establishing ethics reviews for social media research

Recommendation for researchers

19 Researchers should undertake an internal ethics review for all social media research projects that do not seek consent directly from research subjects. Researchers should assess the context of the research to try and understand where potential harm to participants may arise and identify what steps can be put in place to meet user expectations and protect users from harm.

Steve Ginnis
Head of Digital Research
Social Research Institute
Ipsos MORI
steven.ginnis@ipsos.com

Harry Evans
Research Executive
Social Research Institute
Ipsos MORI
harry.evans@ipsos.com

Jamie Bartlett
Director
Centre for the Analysis of Social
Media at Demos
jamie.bartlett@demos.co.uk

For more information

About Wisdom of the Crowd

Four of the country's leading technology and social policy research organisations – Ipsos MORI, Demos, CASM Consulting LLP and the University of Sussex - have teamed up to launch a year-long project that aims to change the way social media information is used.

The **Wisdom of the Crowd** project is sponsored by Innovate UK, with funding contributions from the TSB, the EPSRC and the ESRC – as part of the competition 'Data Exploration – creating new insight and value'.

www.ipsos-mori.com/wisdomofthecrowd