

# Wisdom of the **crowd**



US  
University of Sussex

DEMOS

CASM  
Consulting

Ipsos MORI

## #SocialEthics a guide to embedding ethics in social media research

Harry Evans; Steve Ginnis; Jamie Bartlett

12 November 2015

# Contents

<b>Summary</b>	<b>5</b>
<b>1 Introduction and background</b>	<b>19</b>
1.1 Background to the Wisdom of the Crowd project	19
1.2 Context and scope	20
1.3 Methodology	20
1.3.1 Stage 1: Secondary research	20
1.3.2 Stage 2: primary research	21
1.3.3 Structure of this report	21
<b>2 The key challenges</b>	<b>24</b>
2.1 Defining social media research	24
2.2 Legal considerations	25
2.3 Industry regulations	26
2.4 Public expectations	28
<b>3 Research with the public</b>	<b>30</b>
3.1 Attitudes and awareness of social media research	30
3.1.1 Attitudes and awareness among the general population	30
3.1.2 Attitudes and awareness by sub-group	32
3.2 Conjoint analysis development	33
3.3 Conjoint analysis findings	34
3.4 An ethics board or <i>your</i> data	37
3.5 Anonymity	39
3.6 The context	39
3.7 Consent	40
3.8 Acceptability among key groups	40
3.8.1 Twitter users	41
3.8.2 Frequent contributors	41
3.8.3 Personal information sharers	42
3.8.4 Age differences	42
<b>4 Conclusions and Recommendations</b>	<b>45</b>
4.1 Social media analysis in context	46
4.2 Recommendations for ethical research	48
4.3 Awareness of social media research	49
4.3.1 Low awareness of social media data usage	49
4.3.2 Perceptions of terms and conditions	49
4.4 Consent and opting out	51
4.5 Data minimisation	54
4.6 Safeguarding young people	57

<b>4.7 Re-publishing social media content</b>	<b>59</b>
4.7.1 Publication of content	59
4.7.2 Research with official accounts	60
4.7.3 Developer guidelines on publication	60
<b>4.8 Expectations of users</b>	<b>63</b>
<b>4.9 Seeking ethics approval as standard</b>	<b>64</b>
<b>5 Appendices</b>	<b>67</b>
<b>5.1 Survey on attitudes towards ethics in social media research: topline results</b>	<b>67</b>
<b>5.2 Conjoint technical details</b>	<b>73</b>
5.2.1 Analysing the Conjoint Exercise	73
5.2.2 Key Assumptions	73
5.2.3 Model Structure	74
5.2.4 Fitting the Model	74
5.2.5 Additional Tests and Sensitivity Analysis	74
<b>5.3 Conjoint results overview</b>	<b>75</b>

## List of Figures

<b>Figure 3.1 – Awareness of the usage of social media data</b>	<b>31</b>
<b>Figure 3.2 – Unacceptable purposes for social media data sharing</b>	<b>32</b>
<b>Figure 3.1 – Relative utilities of different attributes</b>	<b>35</b>
<b>Figure 3.2 – Index of relative importance in driving approvability</b>	<b>36</b>
<b>Figure 5.1 – Full list of attributes in the conjoint exercise, with their relative importance in driving approvability</b>	<b>75</b>
<b>Figure 5.2 – Full list of levels within attributes in the conjoint exercise, with their relative importance within attribute</b>	<b>75</b>

## List of Tables

<b>Table 3.1 – The most and least acceptable research project scenarios</b>	<b>36</b>
<b>Table 3.2 – Percent of people rating a research project with ‘your data’ less/more acceptable than when asked to imagine being on an ethics board</b>	<b>38</b>
<b>Table 4.1 – Recommendations by issue, audience and stage of research</b>	<b>48</b>

---

# Summary

---



# 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.<sup>1</sup>

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<sup>2</sup>. 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 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.

---

<sup>1</sup> 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 (EPSRC) and Economic and Social Research Council (ESRC).

<sup>2</sup> 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:
  - 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 the brand guidelines of social media organisations. Is there a need to differentiate publication of social media content between individuals and companies or organisations?
  - 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

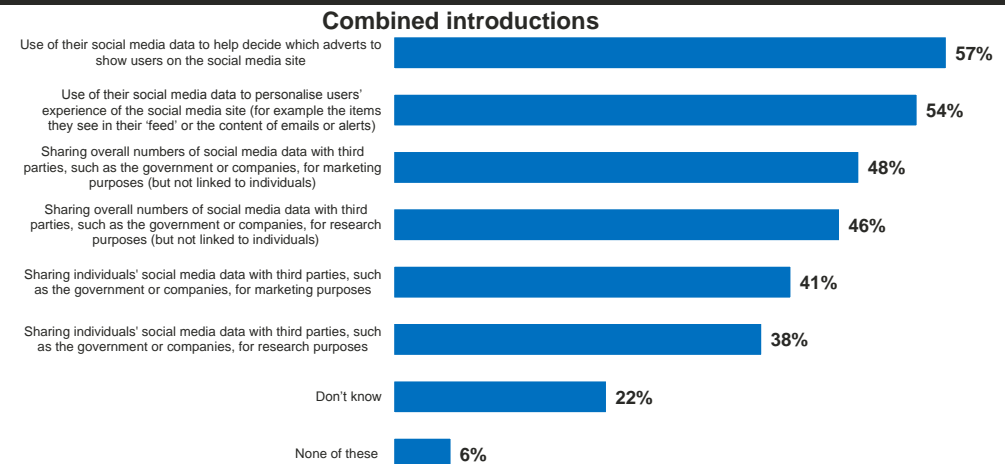
It was felt that in order to fundamentally future proof and inform our understanding of the ethical issues at hand it was vital to conduct research with the public. Three kinds of research were carried out with the public in order to support these objectives:

- 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%).

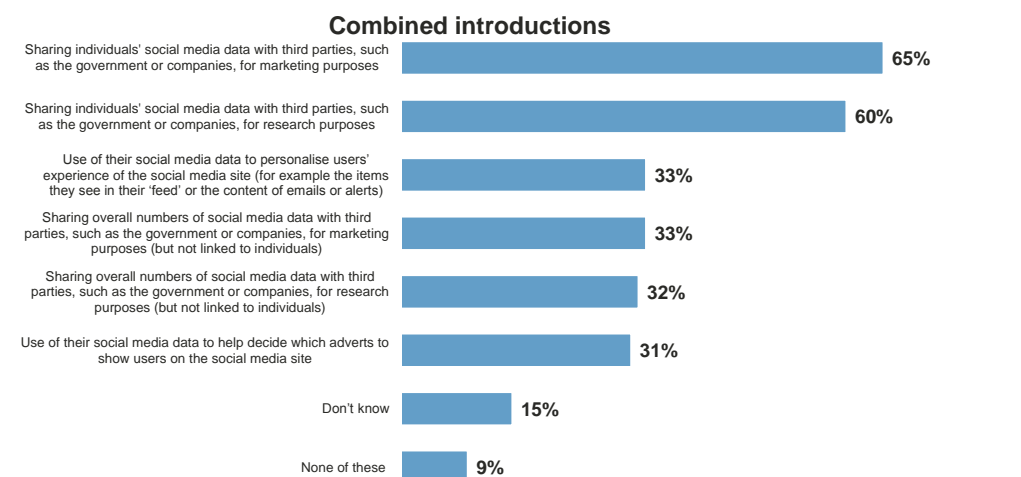
**Which of the following examples of using people's social media data do you think currently happen under the terms and conditions of social media sites?**



Base: 1,250 GB adults

Source: Ipsos MORI

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



Base: 1,250 GB adults

Source: Ipsos MORI

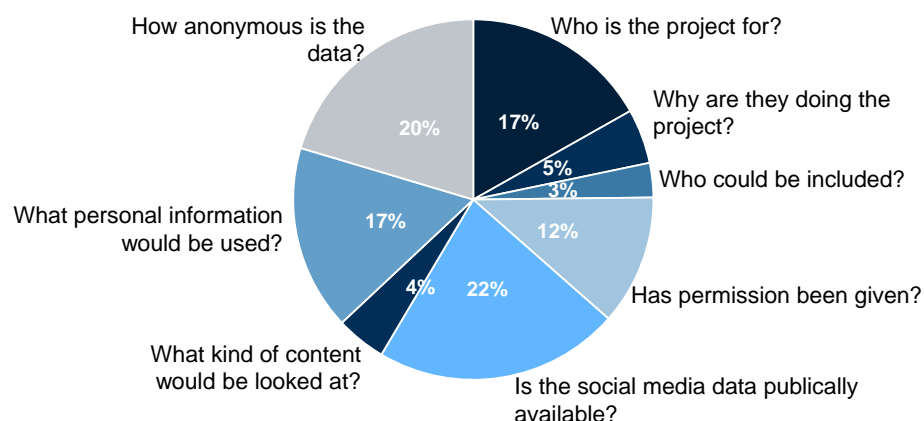
- When asked to review how likely they would be to approve a social media research project,<sup>3</sup> the average approvability score for a project was 5.02 out of 10. Furthermore, 41% gave an average score of 4 or below, which is a broad proxy for unlikely to approve social media research projects. 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.

<sup>3</sup> On a scale of 1-10 where 1 is 'would definitely not approve' and 10 is 'would definitely approve'.



- 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 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:
  - The commissioner of the research;
  - the purpose of the research;
  - what kind of data will be collected and how; and
  - 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*

- 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:

- removing the author's name and @tag from the researcher's sight;
- stripping out other data that is downloaded in the content of a social media post, such as named persons or place names;
- 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;
- 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 by 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.

---

# Introduction and background

---

# 1 Introduction and background

## 1.1 Background to the Wisdom of the Crowd project

This paper forms 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, CASM Consulting LLP and the University of Sussex have come together to critically examine the commercial possibilities for social media research.<sup>4</sup>

The project has assembled leading academics, technologists, thought leaders and insight specialists over the course of the year to address the technical and ethical issues caused by the huge growth in social media data. This project aims to give credibility to the large scale analysis of online conversations by placing non-technical analysts (researchers) at the centre of the data exploration process, with the tools and experience necessary to understand, inform and develop client expectations. 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.

As well as ensuring that the welfare of subjects remains a primary consideration of any research project, research ethics also considers the honesty, integrity, objectivity and transparency of research practices more widely. 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 therefore vital for researchers to step back and consider the views of the people behind the research.

A key strand of the Wisdom of the Crowd project has therefore sought to investigate ethical uses of personal data on social media and review how organisations can balance uses of information with people's right to privacy. The overall aim of this strand has been to develop a series of sensible, positive, and practical recommendations for improving ethical standards in social media research.

---

<sup>4</sup> Further information about the Wisdom of the Crowd project can be found on the project website: [www.ipsos-mori.com/wisdomofthecrowd](http://www.ipsos-mori.com/wisdomofthecrowd)



## 1.2 Context and scope

The project partners believe that there is currently greater clarity in guidelines for conducting ethnographic and co-creational based social media research. This project has therefore intentionally sought to address the current gaps and contradictions in guidelines for conducting ethical large scale social media analysis.

Moreover, the specific recommendations for improving ethical standards have been set within the context of the market research industry in the United Kingdom, and thus under the regulation of the Market Research Society (MRS) and the Data Protection Act (DPA). Wherever possible, the project has sought to consider the legal and regulatory frameworks outside of the UK, and ethical guidelines from other disciplines outside of market research. It is thus hoped that the suggestions put forward in this report will also have a wider positive impact on other organisations and individuals conducting social media analysis.

**Data Protection Act**  
The Data Protection Act is the primary piece of legislation that defines UK law on the process of data pertaining to identifiable living people.

## 1.3 Methodology

The ethical review of this project consisted of three stages:

- 1** Secondary research: a review of the current literature around social media research and its ethical, legal and regulatory implications. This part of the research was to understand the literature landscape before designing stage 2.
- 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** Recommendations: bring the findings of stage 1 and 2 together in order to make recommendations for best practice for researchers conducting social media research.

### 1.3.1 Stage 1: Secondary research

A scoping document was published in early 2015<sup>5</sup>. The purpose of this document was to understand what the current ethical guidelines are that govern and regulate social media research, as well as assessing what research had already been conducted to try and understand what public expectations are for use of social media data.

A fuller explanation of the findings can be found in chapter 3 of this report, but broadly, issues in the literature fell along two axes:

- Legal, regulatory and ethical – what level of regulatory authority covered research undertaken. We found that while many issues were either legal, or stipulated and additional regulatory best practice, but that there were some issues that were

<sup>5</sup> Ipsos MORI and Demos, *Unlocking the Value of Social Media Work Package 3: Ethics, Stage 1: Scoping*, June 2015, available from: <https://www.ipsos-mori.com/Assets/Docs/Publications/wisdom-of-the-crowd-social-media-ethics.pdf>



clearly not being dealt with by either the law or regulators, but were still areas of concerns in the wider public sphere.

- Data collection, analysis and publication – the stage of the research at which an ethical issue was relevant. Different issues were relevant at different stages – anonymity of the data subject for instance is relevant for all three, but gets more important as you move from the collection of personal data to the publication of it.

### 1.3.2 Stage 2: primary research

The primary research stage of the project consisted of three stages:

- Interviews with experts: the project team spoke with experts across nine organisations, each with a particular interest in social media research. The purpose of these consultations were to gain clarity on some of the technical points which the scoping report raised, as well as ensure that we were covering all avenues of exploration.<sup>6</sup>
- Qualitative research with the public: two workshops were conducted with adults aged 16+, and one with 13-15 year-olds to better understand the views of native users of social media. These workshops were conducted in July 2015 and recruited to be broadly reflective of the UK population. This was conducted to fulfil a need identified in the literature for a better understanding of what is the most important to people who share information through social media,
- Quantitative research: an online survey of 1,250 UK adults aged 16-75. Fieldwork took place between 7-13 August 2015. The survey explored people's attitude towards social media research and included a module of conjoint analysis which aimed to identify which attributes of a research project were key drivers of ethical best practice.

A summary of the data from the quantitative survey can conjoint analysis can be found in the appendices to this report.

### 1.3.3 Structure of this report

- Chapter 2 reviews the key challenges in navigating the legal and regulatory frameworks for ethical social media research. These provide the base for the scope of the primary research conducted and the focus of later recommendations.
- Chapter 3 outlines the findings from the online survey and conjoint analysis, providing an evidence base on public expectation of how social media projects should be conducted.

---

<sup>6</sup> In total, individuals from nine different organisations were interviewed. These organisations fell into five broad categories: regulatory experts, client and government representatives, technology companies, academics and legal bodies.

- Chapter 4 details the final conclusions and recommendations of the ethics strands. The 19 recommendations provide guidance on how best to embed ethics in social media research, and are targeted at researchers, regulators and social media organisations.

---

## **The key challenges**



## 2 The key challenges

This chapter presents a summary of the key findings from the scoping report, which can be found in full under separate cover<sup>7</sup>.

### 2.1 Defining social media research

Social media research covers a very wide range of different types of methodology, with often quite different types of data, data collection methods, modes of analysis and presentation. Each social media research methodology brings with it different ethical challenges that must be addressed and each is governed by separate regulations.

Social media research sits as distinct from other, traditional methodologies, but as with any research, it must adhere to existing legislation and relevant professional regulations. In order for social media research at this scale to take place, it is clear that existing guidelines on research methodology and ethics need to be re-thought in relation to how they may apply to social media research. The changes required go to the very core of current consensus on ethical best practice, such as the need to ask for informed consent from research participants, the ability to guarantee anonymity, or the special protection granted to those under the age of 16.

The secondary research and the interviews also made clear that social media research needs to develop its own ethical framework given it is so different to more orthodox kinds of research. The unique nature of social media requires a unique approach, and it will have its own ethical challenges. The need to measure engagement and interactions means that the identifiable individual becomes more important to some kinds of social media research. If a certain individual is very effective at engaging in a social media communications campaign, then that person may be someone that a government department or organisation may want to get in contact to improve their communications effectiveness in the future. This is part of the nature of using social media, and social media research becomes an expedient and effective way of finding these things out. Despite this, it brings up its own ethical considerations about the nature of identifying individuals – what, for instance, is the difference between identifying an individual in social media research who has retweeted a public health message to the most followers, and sitting down and manually clicking through the Tweets?

Another core issue that became evident during the interviews was about the potential for social media research to move in different directions, and at a rapid pace. Facebook Topic Data<sup>8</sup> is one of these new ways of doing research, providing anonymised, aggregate figures for a certain search query – for instance, how many people mentioned a certain brand in a specified time frame. This clearly provides additional safeguards in

---

**Social media monitoring**  
Here referred to as social media research or social media analysis, social media monitoring is a methodology for analysing high volumes of social media data using tools to categorise content in pre-defined ways.

<sup>7</sup> Ipsos MORI and Demos, *Unlocking the Value of Social Media Work Package 3: Ethics, Stage 1: Scoping*, June 2015, available from: <https://www.ipsos-mori.com/Assets/Docs/Publications/wisdom-of-the-crowd-social-media-ethics.pdf>

<sup>8</sup> Facebook, 'Topic Data: Learn what Matters to your Audience' (accessed 25/08/15), <https://www.facebook.com/business/news/topic-data>

terms of keeping individual's data safe and secure, as well as anonymous. However, the promise of anonymisation means that the quality of the data cannot be assured, and organisations are reliant on unregulated bodies to draw together this research. The research sector, in receiving this data, has strict quality assurance processes and has a reputation that is built around the quality of the analysis it produces – no such guarantees or expertise can be provided by social media organisations.

In addition, receiving aggregated, anonymous figures from social media organisations is not necessarily preferred by the public, who we found to be uncomfortable with social media organisations, as they were with other organisations doing this research. The Facebook Topic Data solution to ethical issues only makes the compliance burden smaller, without making the public feel safer about their data. It does this at a cost to regulation of the research procedures and by sacrificing any opportunity to verify the data.

Considerations for social media analysis were found to fall primarily into three domains: legal, regulatory and ethical. In the UK, the source of the **legal considerations** is primarily the Data Protection Act, while **regulatory oversight** is provided by the Market research Society (MRS) and other membership bodies that demand certain standards from their members. In addition, the need to explore the ethical situation going beyond the regulators and the law by examining how using social media data can contravene **public expectations** of what was fair.

## 2.2 Legal considerations

As outlined in the scoping report, there is a clear legal definition of personal data, and all social media content that contains personal data is subject to the regulations set out in the Data Protection Act (DPA). Removing the unique username or given name that is attached to the social media data may not be sufficient to guarantee that the social media data is free of personal information, for example the post may contain a name or other identifying features.

Social media research using personal data can still take place under the DPA, but the data must be processed fairly and lawfully.<sup>9</sup> In the absence of informed consent, the research will need to establish the fair and lawful basis for collecting personal data from social media sites. This will normally be established by the wording of the social media site's terms of use and/or privacy notice. Where the terms of use or privacy policy make explicit that data can be collected for additional purposes, then research may take place. This lays down a legal justification for the use of *public* Twitter and Facebook posts for the purposes of research. The DPA does prohibit the researcher from conducting social media research under circumstances where this is *not* allowed under the terms and conditions – where users have not been fairly notified that their data might be used in this way.

---

**Personal data**  
Personal data is data that relates to a living individual that can be identified from the data or identified when that data is combined with other kinds of easily acquirable data. Data that is classed as personal data is subject to the Data Protection Act.

---

**Research exemption**  
The research exemption in the DPA provides organisations with a limited exception to process personal data for reasons other than the initial collection of the data. In practice, this means that data collected for one purpose can be used for the additional purpose of research (strictly defined).

<sup>9</sup> Section 1, Part 1 of Schedule 1 of the Data Protection Act (as amended)

These issues of privacy are compounded when dealing with sensitive personal information – data pertaining to things like sexual orientation or ethnicity. Sensitive personal information can, in the main, only be processed if it has been declared public by the author. This provides added complication to the considerations for technologies that try and derive characteristics about individuals using social media based on other content. It is also stipulated in the DPA that all personal information held about an individual must be accurate and up to date. This raises additional challenges for derived characteristics that only have a certain probability of accuracy and questions as to under what circumstances it is acceptable to use such information in social media research.

To add further complication to the legal situation, the forthcoming draft EU General Data Protection Regulation (replacing the EU Data Protection Directive) is widely assumed to have extensive ramifications on the use of data by data controllers. The Regulation is likely to put further constraints on companies wishing to retain personal data, as well as making the right to be forgotten explicit in legislation – potentially putting tougher rules on companies holding deleted social media data, particularly if used for ‘profiling’ purposes. The Regulation may also require old social media data to be removed by companies from their systems.

Timescales for the Regulation are currently unknown, but in any situation social media research needs to future-proof itself against potential constraints that new legislation may put in place.

While the legal situation is still confusing, there is the possibility for a new kind of social media research to be put in place where the DPA would not be relevant, thus freeing up the researcher to work with fully anonymised data. Facebook Topic Data<sup>10</sup> will allow researchers to ask key questions of Facebook, who will then come back with aggregated and anonymised results about what kinds of people are talking about certain words and terms. This would give the researcher only the basic counts of who was saying what, without individual-level data. However, as stated above using Topic Data might relinquish the researcher from legal considerations, but broader ethical concerns remain.

## 2.3 Industry regulations

For some time there was little guidance on how social and market researchers could use social media data. However, with the latest update to the Market Research Society (MRS) Code of Conduct, more guidance was provided for researchers using large datasets. Specifically this confirms that the fair and lawful use of personal data for research can be established the terms and conditions of a website, although the Code also makes it clear that informed consent should be sought directly wherever possible.

---

**Code of Conduct**  
The MRS Code of Conduct is the document that codifies the rules and regulations that members of the MRS must abide by.

---

<sup>10</sup>Facebook, ‘Topic Data: Learn what Matters to your Audience’ (accessed 25/08/15), <https://www.facebook.com/business/news/topic-data>

There is still a lack of guidance from regulators of research. The scoping report identified areas where further clarity could be provided by regulators.

- What data can be considered 'private'? Under definitions provided by regulators, data on Facebook could be private despite the fact that this Facebook is accessible to everyone (albeit everyone with a password). There needs to be added nuance to the distinction between 'public' and 'private' social media, especially as the requirement to have an account in order to view any Facebook data is more likely a business decision than an active attempt to protect people's data.
- As well as this, there is some additional confusion from the 'public'/'private' terminology, which mirrors the terminology used by some social media organisations. For instance, if an individual sets their Facebook settings to 'public', it is not necessarily 'public' according to the guidelines. The guidelines fit closer to an approximation of user expectation, than it does to the definition of these terms on social media sites themselves.
- While the regulations emphasise anonymisation, there is a question over the extent to which this is practically feasible? As noted above, even social media content not attached to a name could contain personal data. In addition, Twitter's terms of service for the Application Program Interface (API) require Tweets that are re-published to be presented as they appear on Twitter (i.e. with name and @tag attached).
- Regulations currently do not make a distinction between personal data provided in the collection process, and personal data visible to the researcher. Whilst many social media API's does not limit the personal data that can be *collected*, but it may be that additional safeguards can be put in place automatically before the researcher sees the personal data.
- There is also a lack of guidance on issues such as how to treat sensitive personal information and even what exactly constitutes personal data on social media. There is no distinction made between personal data collected through the API in different fields – such as name and location, fields that can be easily hidden from the view of the researcher – and the actual content published by the individual on the social media. The content may contain its own sensitive personal data, such as sexuality or political affiliation, which cannot be eliminated before the researcher sees it. There is currently no guidance on how to limit the potential impact of this, or even a recognition that it could happen in the guidelines.

Moreover, in the guidance that has been provided from regulators, there are some areas that need some kind of clarification.

- The industry regulations require anonymity in the publishing of findings or reporting this back to a client. This principle goes against the purposes of certain types of social media research, such as key influencer analyses, which attempt to

### 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.



identify important people in social networks. Much analysis of this kind already takes place within social research.

- Is it possible to mask social media contributions to guarantee an author's anonymity while at the same time adhering to copyright law and/or site terms of use? This is similar to the issue of whether tweets can be masked within the developer rules that Twitter outline.
- While the industry generally holds that research on young people should take place only with parental consent, there are currently no guidelines on whether or not this also applies to social media research. A great many under-16s use social media – both Twitter and Facebook have a minimum age requirement of 13 - and there is no way of adequately eliminating all under-16s from social media research.

## 2.4 Public expectations

While regulations and legislation exist to protect participants in research from possible harm, the scoping report recognised that this might only go so far in safeguarding social media users. Consideration was given as to what an 'appropriate flow of data' might be for the public. On the one hand, this is a way of safeguarding participants, but it also serves to protect the reputation of researchers. Examining what the public ethics looks like also allows future-proofing against potential changes to legislation.

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 that these are not necessarily fit to transpose in their entirety onto a commercial environment.

There has been little research 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<sup>11</sup>.

Many groups have called for more transparency in general in how and what extent social media data is collected.

---

<sup>11</sup> 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)



---

**Research with the public**

---

## 3 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. This was primarily to ensure that our recommendations for addressing the challenges of conducting ethical social media research (outlined at chapter 2) would be embedded in user expectations. The main research objectives of our work with the public involved:

- 1** exploring awareness of social media research, specifically looking at attitudes surrounding the publication social media content in the context of research, the vulnerability of under-16s on social media and the acceptability of social media research in comparison to other potential uses of social media data;
- 2** trying to identify which, if any, components of a social media research concern the public and where there is opportunity for researchers to allay fears; and
- 3** specific attitudes towards the collection of data and the research process, as well as safeguards that are available and ideas for improving transparency.

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 within the online survey cited above, 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.

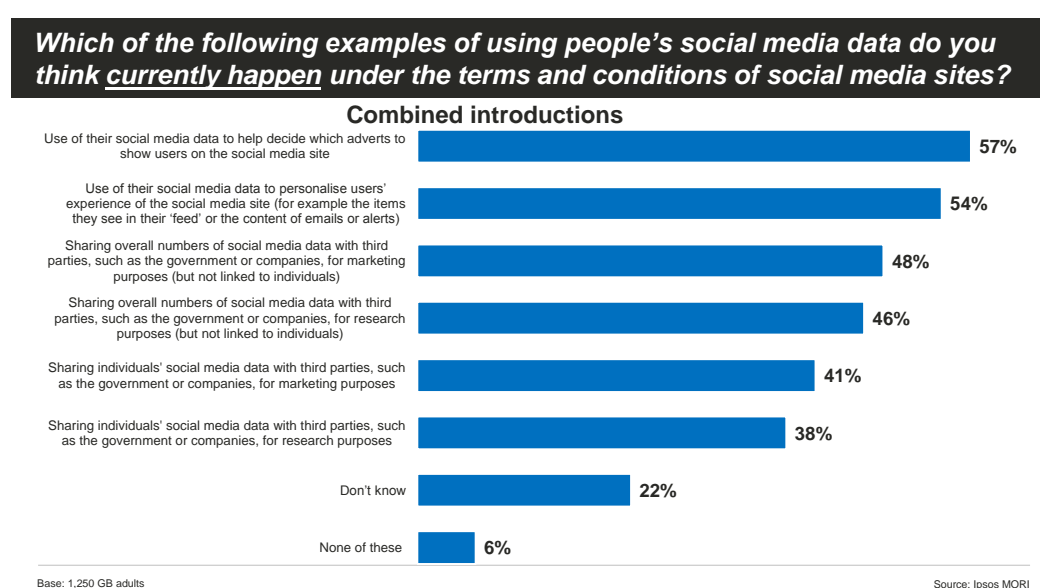
### 3.1 Attitudes and awareness of social media research

#### 3.1.1 Attitudes and awareness among the general population

Currently, the fair and lawful processing of social media content is relies on the agreements given to terms and conditions signed on joining social media platforms. The quantitative survey attempted to understand public awareness and attitudes towards the way social media data is being used through such terms.

The survey first asked people to consider what they believe currently happens with their social media data<sup>12</sup>. While the majority were aware that adverts are targeted using their social media data (57% said this currently happens), and that data being displayed in their social media feed was selective based on their social media data (54%), fewer believed that data is shared with third parties, such as the government or companies. Only two in five (41%) said that they thought that individual-level social media data is shared with third parties for marketing purposes. Awareness was similarly low about research. Fewer than two in five people (38%) said they thought sharing social media data with third parties for the purposes of research currently happens under the terms and conditions.

Figure 3.1 – Awareness of the usage of social media data

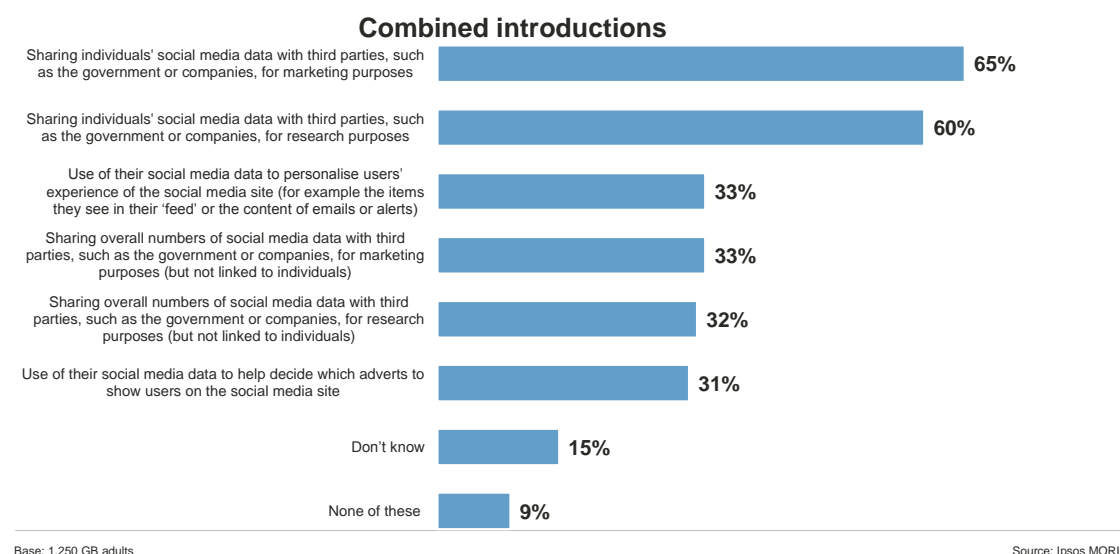


As well as a lack of awareness, research with social media data was also one of the least popular activities on the list. When asked to select activities that they thought *should not* happen, a majority (60%) thought that data should not be shared with third parties for research purposes under social media terms and conditions. Research purposes were only more accepted than sharing data for marketing purposes (65%). Other purposes, such as using data to target advertisements, were more accepted (31% said this should not happen). A third of people thought that sharing overall numbers not linked to individuals for the purposes of research should not happen (32%).

<sup>12</sup> Half of the sample was shown one introduction before these questions, and the other half was shown a slightly extended introduction, adding 'Many social media platforms are free to use. In return for using the service for free, social media sites make use of the data provided on the site by users.' This did not have a large impact on the results so all data quoted here is from the tables with both introductions combined.

Figure 3.2 – Unacceptable purposes for social media data sharing

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



It is clear from these results that there is scepticism among the public about social media research. This may be linked to a lack of awareness. With further analysis of the data, of those who thought that sharing data for research purposes currently happens, two in five (42%) believed that it should not happen, a considerably lower proportion than the public in general (60%).

The findings from the quantitative research were broadly backed up by the qualitative workshops, where many participants had not considered the potential of their data being used for research. When the idea was presented, there was a general scepticism from participants, who felt that this was another unnecessary intrusion into their private lives.

### 3.1.2 Attitudes and awareness by sub-group

While awareness of data being shared with third parties for social media research stayed typically low among most elements of the populations, it was slightly different among some groups. The youngest group surveyed, 16-34s, was slightly more aware (42%) than the older group of 55-75s (33%).

Twitter users were also more aware of this happening (43%, compared with 38% overall), possibly reflecting the openness of the medium, and the fact that most social media research to date has focussed on Twitter. Linked to this, daily contributors to social media (48%) and those sharing four or more types of data on social media (53%) were very aware of the data being used in this way. These findings are unsurprising in many ways as they demonstrate that people who use the medium most often are also most aware of research taking place.

There is a similar story for groups that think social media research *should not* happen. The older group, 55-75, are a more likely than 16-34s to say that social media research should not happen (65% versus 57%, respectively).

Conversely, Facebook and Twitter users are also more likely to say that social media research should not happen than people who do not use social media (63% of Twitter users say it should not happen, 62% of Facebook users, compared to 50% of people not on social media). In keeping with this finding, two thirds of users sharing 4+ types of personal information on social media (66%) said that sharing data with third parties for social media research should not happen (higher than the 60% overall).

While many of these differences are relatively small, they show that concern about social media research is a much more pertinent concern for users of social media, and not simply a concern *about* those on social media. It also shows that much of this concern is coming from the older group of social media users, who may have adopted some of these media later than the younger generation and who have not been brought up with new media.

These findings make sobering reading for researchers looking to uphold a reputation and conduct ethical social media research. There is clearly a need for change.

The next section, looking at the conjoint analysis with the public, delves deeper into the nuances in public acceptance of social media research projects, examining how the different variables of a research project can affect how approving people are of it.

## 3.2 Conjoint analysis development

'Conjoint analysis' was conducted in order to identify the different principles that are most important in deciding whether a social media research project would be likely to be approved by members of the public.

This analysis involved asking respondents to imagine themselves as a member of an ethics board with the power to approve or reject research projects. Respondents were then presented with six scenarios, each with eight different attributes that could sway their opinion on the scenario. Respondents rate on a scale of 1-10 how likely they would be to approve that project. The attributes are randomly rotated, meaning each of the possible scenarios is seen by a substantial proportion of the sample. The subsequent analysis determines which attributes are the most important in the decision making process and, within that, which variables impact that attribute the most.

This analysis is intended to provide an understanding of whether, for example, 'Who the project is for?' exerts more influence on decisions than the question 'Has permission been given?'. It is possible to look deeper into this attribute and see whether 'A central government department' is considered more suitable, i.e. more likely to be approved, to conduct the research than 'Researchers in universities and similar organisations'. Further technical details of the conjoint analysis can be found in the Appendix 5.2 of this report.

### Attributes and levels

In conjoint analysis, 'attributes' denote the categories within which the 'levels' fall. The attribute could be 'Who is the research for?', and the various levels within that would include 'Government departments' or 'Companies'.

The attributes for the conjoint, and their respective levels, were drafted based on findings from the literature review and the interviews to determine that we were asking about the most pertinent contexts for a research project.

These were then tested in six cognitive interviews with members of the public to ensure that the exercise was manageable and that there were no important factors missing from people's concerns about using social media data for research. The attributes and levels were adjusted as a result of this. As a final check, the scenarios were presented as an additional probe to participants in the workshop who undertook this exercise themselves.

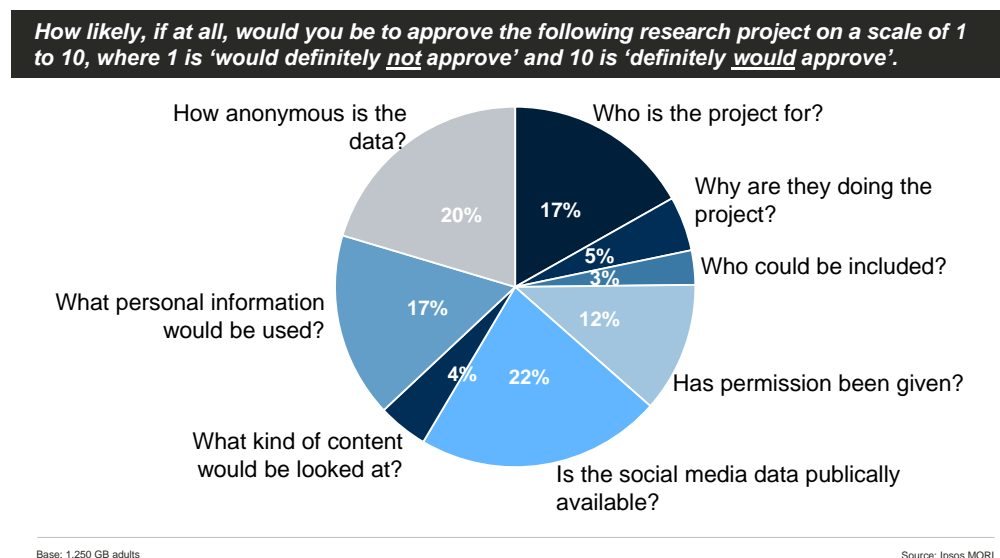
### 3.3 Conjoint analysis findings

While the conjoint is designed to give relative, rather than absolute, scores of likely approval, there was an average score across all participants. On a scale of 1-10, where 1 is 'would definitely not approve' and 10 is 'would definitely approve', the average score across all scenarios was 5.02. However, this misrepresents the range of views held by respondents: just 26% of respondents gave scores around the mid-point of 5-6, 41% gave an average score of 4 or below (broadly a proxy for unlikely to approve), 33% gave an average score of 7-10 (broadly a proxy for likely to approve). This shows that there was a spread of results, with many people clustering their scores around the bottom of the scale, and another cluster at around 7-8 approvability.

Figure 3.1 lays out the relative utilities of the different attributes that comprised every scenario. Within these attributes are different levels that are randomly selected for each scenario – for instance 'Who is the project for?' might appear with 'A central government department' or it might be 'A private company' as two examples.

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

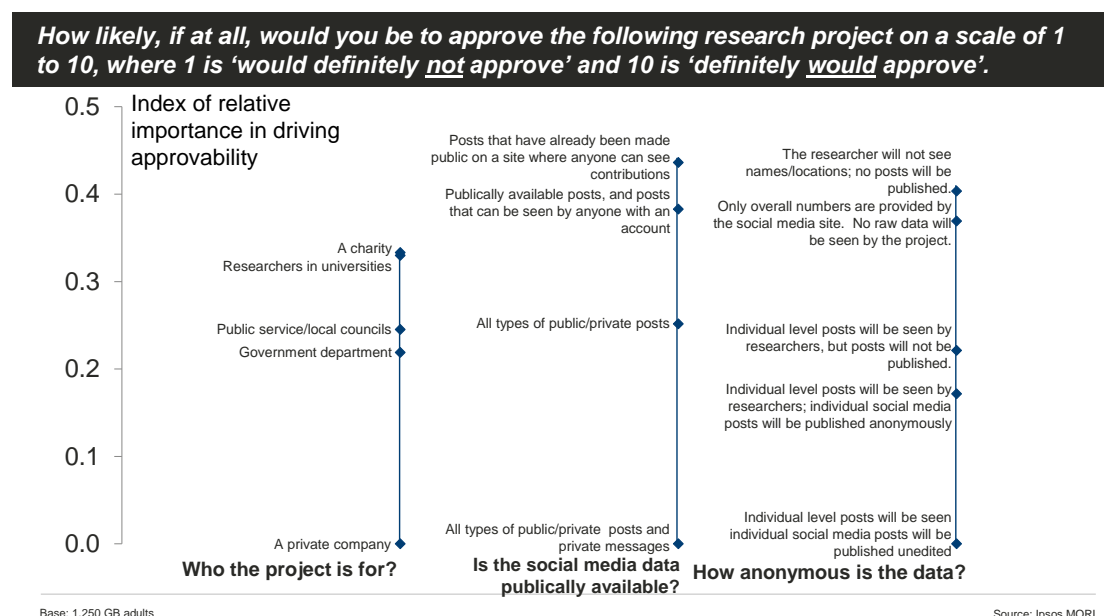
Figure 3.1 – Relative utilities of different attributes



However, within this, there is much more nuance. Figure 3.2 shows the importance of different levels within each of the most important attributes. Looking at the data like this can tell us whether or not the importance of the attribute is driven a constant improvement between attributes, or whether the importance is marked by a significant jump between two levels within the attribute.

Take, for example, 'Is the social media data publically available?' While this is the most important attribute, this is mainly driven by a large disapproval for projects that would analyse private messages. Once private messages are removed from the proposal, there is less of a distinction between the other three levels within the attribute.

Equally, there is little difference in likelihood of a project being approved between whether the project is commissioned by a Government department, public service, university or charity; however there is considerable drop in likelihood of the project being approved if it has been commissioned by a private company.

Figure 3.2 – Index of relative importance in driving approvability<sup>13</sup>


Below is a table that represents the *least* and *most* acceptable scenarios, based on choosing each of the least acceptable levels for each attribute. Under the least acceptable scenario, just 15% of the public would be likely to approve the project. In contrast 50% of the public would approve the most acceptable scenario. This is driven mainly by the opt-in mechanism, the fact that posts have already been made public prior to publication, and that no raw content will be published.

Table 3.1 – The most and least acceptable research project scenarios

	Least acceptable scenario	Most acceptable scenario
<b>Who the project is for?</b>	A private company	Researchers in universities and similar organisations
<b>Why are they doing the project?</b>	To identify the most active or most well connected social media users in a network	To review or act on comments about a product or service they deliver
<b>Who could be included?</b>	Anyone on social media who has been identified as visiting a broad location relevant to the project	Anyone on social media who has used a word, hashtag or phrase relevant to the project
<b>Has permission been given?</b>	All those who have agreed to the general terms and conditions of the social media site when they first signed up	Only those who have opted in to their data being used for this specific project

<sup>13</sup> Within each attribute, the levels have been rescaled so that the least effective level within an attribute is given the value of zero.



<b>Is the social media data publicly available?</b>	Collecting all types of public and private posts, including private messages between individuals.	Collecting posts that have already been made public on a site where anyone can see contributions regardless of whether they have an account
<b>What kind of content would be looked at?</b>	Purchasing habits, or information on the products or brands people like	Health behaviours
<b>What personal information would be used?</b>	Sensitive personal information (for example, sexuality and political affiliation) which is relevant to the project shall be used	Age, gender and broad location will be used to compare different groups of people
<b>How anonymous is the data?</b>	Individual level posts will be seen by researchers, individual social media posts will be published unedited (including author name and any other details posted by the author)	The researcher will not see names/locations and IDs during analysis; no posts will be published.
<b>AVERAGE SCORE ON 1-10 SCALE</b>	3.55	6.28
<b>% giving 1-4 score of likelihood to approve</b>	64%	23%
<b>% giving 5-6 score of likelihood to approve</b>	21%	28%
<b>% giving 7-10 score of likelihood to approve</b>	15%	50%

Source: Wisdom of the Crowd conjoint analysis

Base: 1,250 GB adults.

### 3.4 An ethics board or your data

The final scenario presented only to social media users was constructed slightly differently. Instead of asking participants to imagine they sat on an ethics board, this final question asked them instead to imagine that it was *their own* social media data being used for the research project. Respondents were presented with the same scenario they had been presented with at the beginning of the exercise, only this time asking them to consider if the project was to use their own data.

The differences between the participants' first response and last response were then analysed to see what difference it made to the likelihood of approving the project. The below table shows the percentage of people who found use of their own data more, less and equally acceptable by different subgroups. The general principle remains the same

however, in most cases, users were less likely to approve a project if it was to include their own data than somebody else's.

**Table 3.2 – Percent of people rating a research project with 'your data' less/more acceptable than when asked to imagine being on an ethics board**

	'Your data'		
	% less acceptable	% same	% more acceptable
Twitter users	48%	33%	18%
	49%	33%	18%
Daily users	48%	32%	24%
Weekly/Monthly users	47%	35%	17%
Less frequent users	53%	30%	17%
0 activities sharing personal information	52%	31%	18%
1-3 activities sharing personal information	49%	33%	19%
4+ activities sharing personal information	45%	35%	20%

Source: Wisdom of the Crowd conjoint analysis

Base: All social media users (1,121)

The table shows that there was a real split between those who marked the final scenario, with *your* data, as more acceptable, less acceptable or the same. The higher-end users of social media – both in terms of frequency of use and diversity of use – were less likely than the low-intensity users to find the use of their own data to be less acceptable than use for an ethics board. This suggests that the more engagement you have with a social media platform, the more lax you become about sharing data. Though, conversely, it may mean that if you do not share data on social media, you may be doing so precisely out of an increased privacy concern.

In general, however, people showed more concern when it was their own data at stake than if they were on an ethics board. It is difficult to know what the reasoning for this might be. On the one hand, it could be that people do not make the direct link between an ethics board approving the project and the possibility of their own data being used. On the other hand, it may be that when placed in the position of being the arbitrator of the decision, they place more weight on the importance of the research itself above their own point of view.

### 3.5 Anonymity

The conjoint 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.

These findings are coherent with the other public research carried out. Nearly three quarters of adults (74%) would prefer to remain anonymous if their social media content was published and this can be seen as a difference between the most acceptable and least acceptable cases of anonymity in the above scenarios. Just 10% of people would like to be attributed.

The qualitative research showed a desire for anonymity in the research, but there was less of a focus on whether it was possible for the social media content to be re-identified by researchers. Along with the conjoint findings, this suggests there may not be a great benefit in the public's eyes to social media organisations providing aggregated and anonymous figures versus the researcher analysing de-authored content. It was concerning that data was being *looked at*, regardless of who was doing the looking.

The quantitative research also showed that the majority (54%) agreed that all social media accounts should be given the same rights to anonymity regardless of whether it's a public institution, company or high profile institution. This surprising finding is backed up by the conjoint finding that the least acceptable reason for doing social media research is to identify key accounts in a social media network.

### 3.6 The context

The conjoint analysis reiterated the message that context is key when dealing with attitudes towards data – there was lots of variation between attributes, and between levels as well. 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.

The kind of content and the purpose of the research both came low down the list of priorities, and there was little difference in whether researchers were examining attitudes on social media or how and where services are being used. The *who* was more important, but this was mainly due to respondents being concerned about private companies – the other levels (charities, researchers in universities, and public sector) were all looked upon more favourably than private companies, but there was not a great amount of difference between the charitable and public sectors.

Other methodological contexts played a bigger role – the *how* of the research. The anonymity procedures were important, as mentioned above, and so was the sensitivity of the personal information being looked at. The other how was to do with whether or not the data was public, and this made a big difference to acceptability, especially if the research was looking at personal messages.

In the workshops, it was suggested that any research that would use public social media data should make sure it had a worthy reason for doing so. There was little expansion on this, though for some participants, it appeared to correlate loosely with charitable purposes, or for improving delivery of public services. This theme comes out of the conjoint in several areas. Firstly, trust in the organisation to use the data was a big driver – private companies doing this research were less acceptable than a charity or a public sector organisation. In addition, improving a product or service that the commissioner delivered was the most important driver within the purpose attribute.

Findings from the workshop showed that people cared about how precisely the research could work. They had a lack of awareness of how social media companies use their data – while they are certain they use it, they're not sure how and what safeguards might be in place to protect their data.

### 3.7 Consent

While the conjoint analysis did present permission to use data as an important 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.

The workshops also corroborated this message. People saw an opt-in as the gold standard that researchers should work towards, but did also give the options for an opt-out for specific projects as also good practice. Participants were keen to talk about ways that social media organisations could provide that option at the point of signing up, although also talked about the drawbacks of this approach. Drawbacks included not being sure what your data could be used for in the future.

Participants widely understood that there was a transactional element to signing up to a social media site, though talked about this in negative terms: 'giving up' their data. There was a definite feeling that once this initial transaction had taken place, there was no real control left over what happened with their social media data. This lack of control was reinforced with a lack of awareness of what the data would be used for.

The quantitative survey suggested that the majority (60%) of people believe that sharing individuals' social media data for research should *not* happen. This was nearly as many that thought that data being shared for marketing purposes should *not* happen (65%). These findings suggest that consent for social media research cannot be ignored out of hand, and that the public have serious concerns about this happening.

### 3.8 Acceptability among key groups

Within the conjoint analysis, we were also able to look at the differing responses from key groups. We decided that there were several groups that might hold interesting variation within them:

- Twitter users – Twitter is the most researched medium so far, and also the environment where users might have least expectation of privacy;

- Frequency of contribution to social media – in order to examine whether or not the amount of time spent engaging with social media impacted upon priorities for a research project;
- Variety of personal information sharing on social media – identifying whether or not users who give up a variety of different personal information have different priorities to users who are fairly narrow in their social media activity; and
- Age differences – we know that there are differences in attitude towards social media research in the different age groups from the quantitative survey, and the recentness of much of this technology makes age an interesting factor to examine.

### 3.8.1 Twitter users

The conjoint analysis also showed up some important differences between different sub-groups. At the attribute level, Twitter users placed less importance on the type of commissioner of the research, and were also less concerned about how anonymous the data was than social media users as a whole. Instead, Twitter users were slightly more likely to make decisions based on the sensitivity of the content, as well as the purpose for the project. This is perhaps reflective of Twitter users having more of an expectation of the publicity of their tweets.

Collecting posts that were already public, for instance, made a bigger positive difference to Twitter users than to social media users as a whole. Twitter users were also more positive about collecting public posts that were viewable by anybody with an account. Twitter users were happier with research projects using details such as age and gender in the analysis than general social media users too.

Expectations of the public nature of the Twitter medium can also be seen in the quantitative work. Twitter users were the group most likely to agree that social media accounts held by institutions, organisations and high-profile institutions have less of a right to anonymity than regular users (25% of Twitter users agreed with this versus 21% of social media users).

### 3.8.2 Frequent contributors

We were also able to look at how responses to the conjoint exercise differed by the level of a respondent's social media involvement. Daily contributors to social media put more of an emphasis on who the hypothetical research project was for than an infrequent user. The level of anonymity afforded in the research project was not as big a driver for daily contributors as it was for the less frequent users. The sensitivity of the personal information being used was more important for the least frequent group of social media contributors than for those who contributed once a month or more.

Those who contribute to social media least frequently were most swayed by research projects that looked to evaluate and improve services – such as projects reviewing comments on a service that is delivered, or identifying people who might benefit from a service. Daily contributors were much more likely than less frequent contributors to react

positively to projects where the purpose is to identify the most active, or well-connected individuals in a network.

Even though anonymity was not a major driver overall for daily contributors, within the attribute they were more likely to favour higher levels of anonymity than the less frequent contributor groups. Daily contributors were a lot less likely to support projects where their posts would be published de-authored – possibly reflecting a greater awareness of the ability to re-identify social media users despite efforts to de-author.

### 3.8.3 Personal information sharers

As well as measuring the frequency of contribution on social media, the conjoint and quantitative study also captured behaviours about how different levels of personal information sharing impacted on attitudes towards social media research. While similar to frequency of contribution, personal information sharers are categorised by the diversity of the information they have shared in the last years – engaging in 0, 1-3 or 4+ types of social media activity.

The anonymity of the data was a less important attribute for those who have posted 4+ types of post in the past year than those who had posted less than that. While some of the differences between the different levels of sharer were between the highest users and the lowest users, there were also differences between the moderate users and the highest users. For instance, the publicness of the information was a bigger driver for the 4+ group than the 1-3 posts group.

Interestingly, the kind of content that was being looked at was barely an issue for the low sharers, was of greater importance to those posting 4+ types of personal information.

Within the attributes, the 4+ group was the only group in the breakdowns we looked at to have an approval score of more than 0 for private companies, though all other commissioners were also generally higher for this group. This high use group were also more driven than the low use groups for research projects that collected public data.

### 3.8.4 Age differences

Priorities in the conjoint analysis also differed by age. The commissioner of the project was most important for those aged 55+, and least important to the 16-34s. The public nature of the social media was also less important for those over 55 compared with those under 55. Permission was also less of a drive in acceptability for over 55s than the younger group of 16-34s.

Within the attributes there were further noteworthy differences in the age groups. Young people were less likely to be swayed if it was a central government department that was commissioning the research. Researchers in universities added more acceptability for the 55+ group, while they were less acceptable as a commissioner for the 35-54s.

The young group of 16-34s were also more likely to approve a project where de-authored social media posts were published, while the other age groups were more suspicious of this.

There were also age differences in the quantitative survey. While nearly two-thirds (65%) of 55-75 year-olds thought that research using individual-level social media data should not happen, this fell to 57% among 16-34s. Using social media data to personalise a user's experience of the social media site was also more unacceptable among the 55-75s (42% thought this should not happen, versus just 27% of 16-24s).

This age distinction might in part be ascribed to an awareness divide. While awareness of social media research using individuals' data was low across the different age bands, it was highest among 16-34s (42% aware, versus 33% for 55-75 year-olds). Awareness of sharing overall social media numbers with third parties for the purposes of marketing or research were both highest among 16-34s (53% aware of marketing and the same proportion aware of research purposes) and lowest for the 55-75s (44% and 41%, respectively).



---

# Conclusions and Recommendations

---

## 4 Conclusions and Recommendations

Researchers seeking to conduct ethical social media research face a number of substantial challenges; both in navigating the conflicting guidelines that govern best practice, and in grounding their methodology in the expectations of users. This chapter draws conclusions from across the primary and secondary research conducted as part of the **Wisdom of the Crowd** project, and seeks to offer a series of sensible, positive, and practical recommendations for improving ethical standards in social media research.

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. We are therefore currently developing a code of conduct that draws on this report and makes it applicable for think-tank research.

These recommendations are intentionally aimed at the market research profession in the UK, and thus at companies practicing social media research under the regulation of the Market Research Society (MRS) and the Data Protection Act. However, it is hoped that the suggestions put forward in this report will also have a wider positive impact on market research agencies conducting social media research outside of the UK, and on companies using social media data outside market research. This includes those working in academia, government, journalism, and technology, and in particular companies providing social media analytics in the UK but outside of MRS membership.

Most of the recommendations outlined in this chapter provide researchers with guidance on how they can improve their ethical practice; however, it is also acknowledged that some principles require additional action from regulators and social media organisations to support researchers in this endeavour and realise the full potential of social media research.

In some areas, it is possible to offer recommendations of best practice that should be conducted as standard; however, other areas require a new set of starting principles, from which exemptions should be made on a project by project basis under an internal ethics review. All appropriate steps should be taken to meet these principles and to mitigate against risks; yet the extent to which they are applicable may depend on the objectives of the research.

## 4.1 Social media analysis in context

It is important to firstly set the context in which these recommendations have been made. The conclusions drawn below have helped shape our reasoning and generate workable solutions. As a direct result of the primary and secondary research conducted through Wisdom of the Crowd, the partners of the project believe that:

- 1 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.
- 2 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.
- 3 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.

- 4 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.
- 5 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.
- 6 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.
- 7 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.
- 8 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 programme interface (API).

## 4.2 Recommendations for ethical research

It is intended that these recommendations will help in overcoming some of the key challenges presented in Chapter 2. Each recommendation is preceded by a discussion of the challenge or issue it is seeking to address, and the rationale underpinning the recommendation being put forward. This staged process examines evidence from our interviews with experts, a quantitative survey with the public, qualitative workshops with adults and young people and a conjoint analysis undertaken on the survey findings. Further details on the methodology for this project, including topline results from the quantitative survey, can be found in the appendices to this report.

Recommendations have been made to all relevant parties, and across the research process: data collection, data analysis and publication. Table 4.1 illustrates how each of the recommendations fits into these categories.

Table 4.1 – Recommendations by issue, audience and stage of research

	Issue	Audience	Stage of research
<b>Recommendations 1-2</b>	Boost awareness to build trust	Researchers and social media organisations	Collection and analysis
<b>Recommendations 3-5</b>	Option to opt out	Researchers, regulators and social media organisations	Collection, analysis and publication
<b>Recommendations 6-7</b>	Minimising unnecessary collection of personal data	Researchers, social media organisations	Collection and analysis
<b>Recommendations 8-10</b>	Removing under-16s from social media research	Researchers, regulators and social media organisations	Collection and analysis
<b>Recommendations 11-16</b>	Permission for publication	Researchers, regulators and social media organisations	Publication
<b>Recommendations 17-18</b>	Defining private	Researchers and regulators	Collection and analysis
<b>Recommendation 19</b>	Establishing ethics reviews for social media research	Researchers	Collection, analysis and publication

## 4.3 Awareness of social media research

### 4.3.1 Low awareness of social media data usage

It is clear that the benefits of research on social media data are not widely understood, and that the idea is not currently trusted. Under two fifths of adults (38%) think that sharing individuals' social media data with companies for research purposes currently does happen; over half of adults (58%) believe that it should not take place at all. Researchers should therefore adopt transparent practices to ensure that awareness is bolstered, and that negative perceptions are challenged.

Based on discussions of the ethical challenges of social media research with legislators and regulators, it was clear that the best way of ensuring an ethical and legally compliant best practice is to maintain transparency. Any attempt to be as transparent as possible about what is being done with the data is likely to be perceived by regulators as a desire to be compliant.

The qualitative groups with the public reinforced this message that awareness of uses of social media is low among the public. Focus group participants were concerned that things were happening with their data that they were not aware of. Many participants in the groups were not aware that third parties are able to undertake research using their social media data, and this contributed to the overall feeling of a loss of control. It was expressed that people do too much with their data without their knowledge and that transparency was one way to counter this. It was stressed by some participants that, while suspicious generally, if researchers were more transparent then this would help to build trust.

The findings from the conjoint analysis also demonstrated that context of methodology is vital for improving the acceptability of a research project. Particularly, knowing if the project is for a charitable or public sector commissioner, and whether the data is in the public domain can have a big impact on the level of acceptability.

### 4.3.2 Perceptions of terms and conditions

It was apparent from the scoping report that there are significant concerns regarding the terms and conditions of social media websites. Specifically, whether terms and conditions provide a sufficient basis for research to take place without any further contact with the user? The law and the industry regulations accept that it does provide such a basis, but concerns remain among both users and stakeholders about the transparency of terms and conditions.

These concerns were also voiced in the groups with the public. The large majority participants said they had not even tried to read the terms and conditions of the social media sites they used. A culture has been built of regarding all privacy policies as dull and too long to contemplate. Participants admitted that they would probably still not read the terms and conditions even if they were far shorter than currently.

There also appears to be a heavily entrenched view that privacy policies exist to baffle users rather than helps them, and this means that most users will not even try to engage with them. Steps social media organisations have made to shorten their privacy documents have not yet overcome the hurdle of people's preconceptions of the comprehensibility and length of such documents. As an end result, social media users are often unaware that the data they share on social media sites could be used for market 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:

- The commissioner of the research;
- the purpose of the research;
- what kind of data will be collected and how; and
- 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.



## 4.4 Consent and opting out

During the fact-finding exercise with experts, it became clear that social media research with large datasets is distinct from other types of traditional research in that consent is simply not practicable. Regulators have made changes to allow social media research to take place, so long as the data is made public under the terms and conditions of the social media site.

Consent and permission were also important factors for the public. Even if they had not read the agreement in detail, participants in the qualitative workshops assumed that they were giving some level of permission for third party use of their data when agreeing to sign terms and conditions. They were most likely to cite use of their data for tailoring services to them or providing targeted advertising; in each of these examples they perceived that users would get some direct benefit. However, they did not feel that technical agreement through terms and conditions was sufficient to be considered as consent for their data to be used in a research project.

In the conjoint analysis, while *permission* was not one of the most important drivers of approving a project (5<sup>th</sup> of 8), the public were more likely to approve projects which had some form of opt out or opt in mechanism. Equally, they were least supportive of projects that relied purely on terms and conditions as permission to go ahead with social media research.

A system whereby an individual opts in to a specific research project was seen as the most ethical, while opting out of specific projects had a good degree of approvability. Many participants in the workshops said that they would want to be contacted about specific research projects to ask for consent; however others disagree and suggested that individuals might not want to be bothered each and every time a researcher was hoping to do social media research.

Participants also discussed the possibility of a tick box form of consent for research, which could be selected at the point of sign up or amended within profile settings at any time. While they thought that being approached for specific projects was fairer, there was an assumption that a tick box allowing people to opt out of their data being used for research purposes would be a more acceptable form of consent than simply allowing consent to come via the terms and conditions.

It is therefore clear that opt-out mechanisms for social media research warrant serious consideration for the market and social research industry. Google has recently put together an opt-out system for personalised advertisements<sup>14</sup>. The European Advertising Industry is also putting in place an opt-out system for behavioural advertisement – the practice of looking at internet browsing activity to target online ads at individuals<sup>15</sup>.

---

<sup>14</sup> Google (2015), 'Opt Out', <https://support.google.com/ads/answer/2662922?hl=en-GB> (accessed 15/09/15)

<sup>15</sup>Your online choices (2015), 'About' <http://www.youronlinechoices.com/uk/about-behavioural-advertising> (accessed 15/09/15)

## 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.

## 4.5 Data minimisation

Another common theme among both stakeholders and members of the public was the amount of personal data collected during social media projects. Regulators and legislators recommended a policy of 'data minimisation' and only collecting the data that is required for a certain research aim. The ICO recommends data minimisation as a policy of best practice when dealing with personal data and publishes a guide on the topic<sup>16</sup>. The primary aim of a policy of data minimisation is to reduce the risk of harm to the research participant.

The groups with the public also illustrated that data minimisation is a good course to follow. While there was a broad concern expressed in the groups that personal data could be used for malicious purposes, there was also a wish that any research that was done should only collect the data that was relevant to the study. Some participants maintained the belief that it must be possible to limit the *collection* of personal data, but admitted that, failing that, the information should be hidden from the researcher where possible.

The conjoint analysis reinforced the view implementing a policy of data minimisation would increase the likelihood that a social media research project would be approved. When asked to review a series of hypothetical projects, the type of personal information collected was an important attribute and, within this, sensitive personal information was the characteristic most likely to prohibit a hypothetical project. Using age, gender and broad location were the kinds of personal information that were more likely to drive acceptability.

While it can be tempting for researchers to collect as much detail as is possible, it is safer to try and limit the personal data that might be seen by the researchers. It is not always possible to limit *collecting* this data as the APIs of social media sites tend to either be *fully on* or *fully off*, but steps may be taken to ensure that the data the researcher has access to is limited.

The principles of data minimisation should also be applied to deriving characteristics about individuals. Regulators recommended that the researcher ask themselves why it is that they need that data and whether or not it is possible to conduct the research without it before this was done. The quantitative study also suggested that deriving characteristics may be undesirable for the public, where less than a fifth (18%) thought it was acceptable for researchers to estimate personal details about an individual based on other information. This gives reason to take strict care of when and where inferred characteristics might be used.

Legal bodies, too, have concerns about derived characteristics, and we are looking carefully at what safeguards may need to be in place to ensure that such methods are fully compliant – for example, ensuring that these metrics are accurate and up to date. There are also restrictions on deriving characteristics classed as sensitive personal

---

**Derived characteristics**  
Technologies are being developed that allow algorithms to estimate the age, gender, and even ethnicity of a social media user, based on other details they have provided – such as name or interests. These are derived characteristics and have varying levels of accuracy.

<sup>16</sup> Information Commissioner's Office (July 2014), 'Big Data and Data Protection', <https://ico.org.uk/media/for-organisations/documents/1541/big-data-and-data-protection.pdf> (accessed 15/09/15)

information. For these reasons, we advise researchers to be very careful when considering analysis using characteristics that are derived from other information a user has made public.

The use of metadata came out of the expert interviews as an interesting area for social media research. An example of this might be looking at tweets that are geo-tagged in a certain location, which allows researchers to look at Twitter users only from within a broad geographic area.

The qualitative work with the public, however, indicated that there appears to be little grasp of what kind of additional metadata is collected by social media organisations about them. Mostly references to personal data were related to names, photos etc. The 13-15 year-olds taking part in the focus groups had a better grasp of the possibilities of metadata, mainly due to their experience of apps which ask them if they are happy to share their location.

Participants in the focus groups did have an opinion on their location data being used when this was presented to them as a possibility. There was a consensus that a broad location was acceptable for the purposes of research, but publication of more specific locations were seen as 'creepy'. The example given in the groups was that while a London Borough might be an acceptable level of detail for location, a specific street location or postcode would not be.

---

**Meta data**  
**Attached to any social media content is additional information about the user – such as the time of the post, for instance, or, in some cases, their location at the time of posting.**

## 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:

- removing the author's name and @tag from the researcher's sight;
- stripping out other data that is downloaded in the content of a social media post, such as named persons or place names;
- 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;
- creating generalised groupings of data instead of analysing specific data. For instance, generalising locations by cities instead of exact street locations; and
- 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.

## 4.6 Safeguarding young people

The literature review conducted previously highlighted a gap in guidance relating to whether it is ethical to conduct research on social media given the fact that young people are openly using the services, but are not identifiable. The regulatory framework in traditional market and social research demands that researchers do not undertake research with under-16s without prior parental consent and it was apparent from the interviews conducted here that regulators have not yet reached a decision on how to deal with this issue. This is an important safeguard to ensure that research is conducted within the best interests of young people under the age of 16.

The qualitative work with the public found that they were also concerned about this. Participants voiced fears that young people would not understand the extent of data that was being shared and would be left vulnerable to harm because of it. There was also a fear that the terms of use for social media services are even less likely to be read and understood by users aged under-16.

We also talked to 13-15 year-olds as part of this study. This group is old enough to use most social media sites, but they are under the age that researchers would conduct research without a parent's consent. It was clear in these groups that their understanding of how much data, and how much personal data, they share on social media was fairly advanced and often more savvy than older participants; however they had less awareness of the potential consequences of sharing this data. These young participants had not given much thought as to what might happen to their data once it had gone online, and the concept of social media research itself was somewhat alien.

When presented with information about different types of social media research, the response from young people was mixed. Many did not have a strong opinion on whether the research should or should not happen, while others asked whether it was right that researchers could look at their social media data without their consent.

While it is clear from talking to experts that there is the possibility for excluding 13-15 year-olds from datasets by estimating their age, this will only ever be probabilistically correct. This means that estimating age will not be able to adequately avoid the fears of the adult workshop participants who did not feel that young people had a good enough understanding of how their data might be used to be a participant in research.

Currently, social media APIs do not allow for a selective download of data – in the case of Twitter, the hose of Tweets is either on or off. This means that researchers who are conscious about downloading data from young people have no way of not downloading that data. Despite this, social media organisations do have this data available to them about users, so ethical researchers would appreciate a function that allowed a selective download of users.

## 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.



## 4.7 Re-publishing social media content

### 4.7.1 Publication of content

One of the greatest areas of dispute surrounds the re-publication of social media content as part of a research project. Some argue that research agencies should be able to re-publish content that has already been made public. Others suggest that the principle of anonymity (on which consent for participation has traditionally taken place) should be upheld in social media research.

In order to adhere to anonymity, some industry regulators suggest masking social media posts, yet it is also accepted that this may not always provide a guarantee of anonymity. For example for public Tweets it would be possible to identify the author through using internet search engines, even if the author had been masked in the publication of the research. This is complicated further by the brand guidelines of some social media organisations which govern the re-publishing of social media content. For Twitter for example, there is a requirement to publish any tweets in full, including the user's @tag. In addition, any posts that are published must be deleted from the publication if they are deleted by the user from Twitter. Furthermore, it is also plausible that some people might actually wish to be attributed in the republication of social media posts.

However, there was a strong feeling from the focus groups that if given the choice participants would opt to remain anonymous in publication unless their consent is sought. Indeed, the quantitative research found that most people would opt for anonymity in publication (74% agreed with this, versus 10% who said they would want to be attributed).

Whilst participants were happy to be re-published by other social media users (where the comment remains in context), they are less comfortable with the principle that they will be attributed to content that has been processed and analysed as part of a research project, where they perceive that the content will be judged to hold a particular opinion, attitude, behaviour and/or demographic. They were most concerned about the possibility of a comment being taken out of context and whether they would be held account to things they said in the past that may no longer be true or where their views might have changed. This was particularly pertinent to examples cited where potential employers had used people's social media as a way of filtering out applicants for vacancies – a practice that some participants believed to be unfair.

Ethical consensus in this area already seems to be moving towards seeking consent where possible. Some journalists are starting to ask permission to re-use content from social media<sup>17</sup>. COSMOS – a social media research institution at the University of Cardiff – also practice seeking consent for all social media content that they wish to publish<sup>18</sup>. However, where the content of the post is not considered sensitive, COSMOS will still publish posts where they do not receive a response to their request for publication.

---

<sup>17</sup> See an example of best practice guidelines from the American Press Institute (September 2014), 'Practice ethical curation and attribution', <https://www.americanpressinstitute.org/publications/reports/strategy-studies/ethical-curation-attribution/> (accessed 15/09/15)

<sup>18</sup> COSMOS, 'Ethical statement', <http://www.cs.cf.ac.uk/cosmos/cosmos-ethics-statement/> (accessed 15/09/15)

#### 4.7.2 Research with official accounts

Ensuring the anonymity of participants/users during reporting the findings of social media analysis is still held highly by regulators. While researchers ought to make efforts to provide anonymity when undertaking many types of social media research, there are other times when this is either not practical or desirable. For example, there is currently a popular form of research which identifies key individuals in a client's social media network. This type of research allows a client to understand which social media users play an important role in their network.

It is vital to this kind of research that individuals can be identified to the client, even if their specific posts are not published. Currently the regulations would require anonymity and therefore erode the usefulness of this type of research.

However the medium of social media is clearly very different to traditional forms of communication; there is currently a clear definition of when an individual on social media becomes a brand. Take analysis of the election for example, whilst the official account of the Prime Minister (@Number10gov) would obviously count as a public organisation, and thus not subject to the DPA, what about the account of the leader of the opposition, a parliamentary prospective candidate seeking election, or a journalist commenting on the campaign? Moreover, if a Twitter user is retweeting information about a communications campaign to several thousand followers, do they have a legitimate expectation of privacy about this fact? In addition, all social media users have the freedom to conduct their own kind of key influencer analysis simply by understanding who is sharing their content via social media platform plugins; and publically accessible tools such as Twitter Advanced Search can easily identify individuals within simple searches. Why should organisations not be allowed to undertake a more rigorous form of this kind of analysis?

The public groups did not have much to contribute when asked about this topic, although some did accept that individuals might at some point become a brand if they had a large enough base of followers. However, it was apparent that they, themselves, were content to “have a nosy” at other people on social media who they might know through friends. In the quantitative survey, a small majority (54%) felt that institutions, organisations and high profile individuals have the same rights to anonymity as any other social media account.

#### 4.7.3 Developer guidelines on publication

The current guidelines for the display of tweets stipulate the following requirements, which are necessary for use of the API:

- Do: ‘Show name, @username, unmodified Tweet text, profile picture (where possible), timestamp and the Twitter logo nearby’
- Don’t: ‘Modify the Tweet text with the exception of removing hyperlinks’<sup>19</sup>.

---

<sup>19</sup> Twitter, ‘Display Requirements’ (accessed 25/08/15), <https://about.twitter.com/company/display-requirements>

Both of these principles go against an ethical researcher's attempts to anonymise Tweets in the publication of them. Regulatory guidelines in place do not permit the identification of social media users at the point of publication and so Twitter's limitations on anonymisation of their content is problematic.

Brand guidelines force researchers to look for consent for each piece of content that is re-published. It would be easier for the researcher and safer for the respondent if fully anonymised social media could be published, involving both de-authoring the content and masking it, by taking steps to reduce the risk of re-identification through modification of the social media content.

## 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 by 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.

## Recommendations 11-16 continued

### *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.

## 4.8 Expectations of users

The classification of social media data as either ‘public’ or ‘private’ has a number of important consequences for social media analysis, both in the extent to which users expect private content to be analysed for research, and in the extent to which private sensitive personal information can be processed under the terms of the DPA.

The conjoint analysis demonstrated the public/private nature of the data is an important factor in determining the public acceptability of a research project. Whilst use of public social media content was very likely to make a research project more acceptable, use of private data was considered fairly off-limits. As might be expected, research involving private messages was likely to eliminate all prospect of approving a project to go ahead.

Participants in the qualitative workshops were aware of privacy settings, and many had used these to tailor their account; however they were uncertain how this aligned with what data was and was not available for social media research. Participants noted that everything on Twitter is public (unless an account is protected), but also held the expectation that Facebook data was viewable by all, unless they adjusted their privacy settings. They also noted that direct messages on Facebook, or similar applications like WhatsApp and Snapchat held greater implicit privacy; many participants suggested that these should therefore remain off-limits to researchers, and even social media organisations themselves.

The scoping report and subsequent interviews established that regulators would technically consider a forum like Facebook to be private, as it requires a password to view information. Private forums require explicit consent from members/users before research can take place compliantly within them. However, it was not clear whether Facebook can be seen as a great deal more private than Twitter – especially as the APIs do not have access requirements that restrict who can use the data at the back-end. It was clear from talking with legislators, that under the DPA, public forums are whenever the data has been made public under an agreement such as the terms and conditions of a website.

## 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.

## 4.9 Seeking ethics approval as standard

Interviews with experts in the field highlighted a fear that even though uses of social media data are safe, they can in some cases lead to a potential harm to participants. In all forms of research, regardless of whether it is market, social or academic research, one of the fundamental principles is that of avoiding harm to participants. Where consent is not sought directly for a research project, the onus to avoid harm is therefore even stronger – although it is also clear that a participant having given consent does not remove the researcher’s responsibility to protect participants from harm.

The conjoint analysis conducted for this project points to the importance of context in helping decide whether social media projects would gain approval from the majority of the public. Context is paramount both in terms of the purpose of the research, and in the methodology used to conduct the project.

A number of the recommendations in this report suggest that approval from an ethical review process is required before it would be appropriate to proceed. This is particularly important where the project seeks to include data from those aged under 16, analyse aggregated and anonymous private data, or publish raw unedited verbatim content. An ethics review is also crucial to ensuring that the maximum amount of data minimisation will be applied to the project.

## 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.



---

# Appendices

---

## 5 Appendices

### 5.1 Survey on attitudes towards ethics in social media research: topline results

This survey was conducted as part of the wider Wisdom of the Crowd project, looking at the feasibility of conducting social media research. The Wisdom of the Crowd project is an undertaken by Ipsos MORI in collaboration with the Centre for the Analysis of Social Media (CASM) at Demos, and the University of Sussex. The project receives part-funding from Innovate UK.

Results are based on a total of online 1,250 interviews with adults aged 16-75 in Great Britain. All interviews were completed online between the 7 and 13 August 2015.

Base sizes of less than 100 should be treated with caution and these results seen as indicative only. Bases of less than 30 should be treated with extreme caution and be seen as indicative only.

The data has been weighted by age, gender, region, social grade and working status.

**Q1. Which of the following have you visited or used in the last 3 months?**

	%
Facebook	74
YouTube	67
Twitter	35
LinkedIn	23
Instagram	21
Pinterest	18
Google+ (The social networking site from Google, not the search engine)	12
Snapchat	12
Tumblr	8
Other social media site such as Vine	4
None of these	10
Don't know	*

Q2. How often, if at all, do you personally do each of the following:

	Several times a day %	Once a day %	A few times a week %	A few times a month %	Hardly ever %
Login to check your social networking accounts?	46	19	15	7	13
Contribute or share content to a social networking site?	12	13	25	19	32
Read Tweets that other people have written	35	18	26	12	9
Send a Tweet you have written yourself	9	9	21	18	42
Re-tweet content someone else has written	10	8	21	19	42

Q3. People use social networking sites for lots of different reasons. Which of the following, if any, best represent your main reasons for using social networking sites?

Please think about all of your activity across all of your social networking accounts such as Facebook, Twitter, Google Plus, Pinterest and LinkedIn.

	%
To engage and have conversations with people other than friends and family	55
To keep up to date with news or events	47
Sharing photos that I've taken	43
To keep up to date with companies, organisations or brands I like	26
To engage and have conversations with friends and family	25
To support and raise awareness of campaigns or issues you care about	20
To read what is written by politicians, celebrities, or other well-known people	19
To find new friends or contacts	19
To search for offers or reviews, or to purchase products or services online	16
To share an offline experience online, (e.g. discussing what you are watching on TV)	16

To report on things I've seen taken place or that are important or urgent	13
To review or rate products or services	12
To express political opinions	11
To promote or sell goods and services	8
To be noticed by politicians, celebrities, or other well-known people	3
Other (please specify)	7
Don't know	7

#### *INTRODUCTION 1 – ASKED OF HALF SAMPLE*

Before they can join a social media site users are asked to agree to terms and conditions, these outline what they can do on the site and how the information they give will be used.

There are a number of different ways social media sites could use the information shared by users on their platform. Data shared by users includes the attitudes, experiences, behaviours and personal details they choose to publish on the site.

#### *INTRODUCTION 2 – ASKED OF OTHER HALF OF SAMPLE*

Many social media platforms are free to use. In return for using the service for free, social media sites make use of the data provided on the site by users.

Before they can join a social media site users are asked to agree to terms and conditions; these outline what they can do on the site and how the information they give will be used.

There are a number of different ways social media sites could use the information shared by users on their platform. Data shared by users includes the attitudes, experiences, behaviours and personal details they choose to publish on the site.

- Q4. a) 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?  
 b) And which, if any, do you think the following examples should not happen?

	Currently happen %	Should not happen %
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	33
Use of their social media data to help decide which adverts to show users on the social media site	57	31
Sharing <b>individuals'</b> social media data with third parties, such as the government or companies, for <b>research</b> purposes	38	60
Sharing <b>overall numbers</b> of social media data with third parties, such as the government or companies, for <b>research</b> purposes (but not linked to individuals)	46	32
Sharing <b>individuals'</b> social media data with third parties, such as the government or companies, for <b>marketing</b> purposes	41	65
Sharing <b>overall numbers</b> of social media data with third parties, such as the government or companies, for <b>marketing</b> purposes (but not linked to individuals)	48	33
None of these	6	9
Don't know	22	15

Questions 5 and 6 were asked here and indicate the questions asked in the conjoint module. Please see next appendix for details.

- Q7. Which of the following statements comes closest to your view on the publication of social media posts?
- A. If one of my social media posts was used for research and was selected to be published, I would want to remain anonymous so that no one knew it was me
- B. If one of my social media posts was used for research and was selected to be published, I would like the post to be attributed to me so that people could see what I said

	%
Agree much more with A than with B	58
Agree a little more with A than with B	16
Agree equally with both / don't agree with either	10
Agree a little more with B than with A	6
Agree much more with B than with A	4
Don't know	6

- Q8. And which of the following statements comes closest to your view on conducting social media research?
- A. All social media accounts should be given the same rights to anonymity when used in social media research, regardless of whether the account is held by a public institution, private company or high profile individual.
- B. Social media accounts held by public institutions, private companies and high profile individuals should be treated differently to accounts held by members of the public; social media research involving these accounts should not be anonymous.

	%
Agree much more with A than with B	38
Agree a little more with A than with B	16
Agree equally with both / don't agree with either	15
Agree a little more with B than with A	13
Agree much more with B than with A	8
Don't know	10

Q9. Which of the following, if any, have you done on social media in the last year?

	%
Talked about your holidays or put up holiday photos	36
Said when your birthday is	30
Recommended a book, film, TV show or play to others	28
Talked about a major event in your life – for example wedding, becoming pregnant, giving birth, getting a new job, moving house, etc	27
Asked people to support a campaign or petition	24
Posted articles from newspapers you agree with	21
Talked about your children or put up photos of your children	17
Talked about voting for a political party	15
Talked or asked questions about a medical condition you or your family have	9
Talked about your religious beliefs	6
None of these	30
Don't know	2

Q10. To what extent do you agree or disagree with the following statements:

	Strongly agree %	Tend to agree %	Neither/ nor %	Tend to disagree %	Strongly disagree %	Don't know %
You can't live in the modern world without giving personal information to companies and government	15	45	18	11	7	4
I find it a bit creepy that government and businesses could know so much about me, even if it doesn't really harm me	37	32	15	8	3	4
I benefit from companies using my personal data – for example by getting a quicker service or receiving recommendations for products I wouldn't otherwise have thought of	2	17	30	25	21	5



I don't really think about what information or data the government and companies hold about me since there is nothing I can do about it	7	28	28	20	14	4
It is acceptable for a researcher to use computer programmes to estimate personal details about an individual, such as gender or age, from other information such as their name, topics they have posted about and so on	3	15	32	25	20	5

## 5.2 Conjoint technical details

### 5.2.1 Analysing the Conjoint Exercise

The conjoint exercise involved presenting each respondent with a series of six different scenarios, each scenario describing a possible social media research project. The respondent was asked to imagine that they were advising whether the project should be approved or not. Each scenario was built from eight components, or *attributes*, such as “Who the project is for”, with each attribute having a fixed set of possible answers, or *levels*. Each respondent gave their likelihood to approve each scenario on a 10 point scale. The purpose of the analysis was to determine which components of the scenario were strongest in driving stated likelihood to approve

The final scenario was then repeated with a slightly different question. The difference between the two responses for the same scenario was then considered. We considered fitting a separate model for the second question, but with only one rating for each respondent there was insufficient data for a robust result

### 5.2.2 Key Assumptions

There are some underlying assumptions behind the analysis of the conjoint exercise are as follows:

1. Each respondent has a level of “utility” for each scenario presented to them. This is an abstract concept which could be interpreted as a measure of acceptability of the scenario
2. Each of the 8 That the utility for the presented option can be considered as being the sum of the utilities of each of the components (the levels of each attribute)

3. That the utility can be related to a respondent's claimed likelihood to approve.  
The greater the utility the greater the chance of approval

### 5.2.3 Model Structure

Respondents were allocated to "blocks" with each block seeing a different set of scenarios. In total there were 40 blocks each with 6 scenarios, and so 240 different scenarios were shown in total with between 30 and 32 respondents seeing each. For each scenario we calculate an average likelihood to approve and we relate that to the utility for the scenario represented by the sum of the utilities of each of the components.

### 5.2.4 Fitting the Model

We have converted the 10 point likelihood scale into a probability using a linear transformation (10 =95%, 9=85% .....1=5%).

For the relationship between utility and probability we have used a logit model of the form:

$$p = 1/(1+\exp(U))$$

An initial set of parameters are then determined using a multiple regression which minimises the mean squared error between the actual rating for each scenario and the modelled rating. There are 39 parameters (8 attributes with between 3 and 6 levels).

The final model also includes an adjustment for each respondent to allow for the variation from individual to individual in terms of tendency to approve or not. This works well if there is great variation between respondents, but less well if each respondent shows variation in tendency between exercises. The final set of utilities derived from this approach was compared with those obtained from the initial model and found to be not significantly different

### 5.2.5 Additional Tests and Sensitivity Analysis

We tested the assumption of independence of attributes by including interaction effects between pairs of attributes. Each possible pair was tested separately as a potential incremental term, but none were found to have a significant effect. We concluded that the independence assumption was reasonable.

### 5.3 Conjoint results overview

The below table outlines the full list of attributes and levels within those attributes in the conjoint analysis. The importance of each attribute is represented as a **percentage** that shows the share of variation that each attribute explains, i.e., the relative importance of each attribute to acceptability. The utility scores for each level represent their importance, and a higher utility score means that the level is more likely to lead to a positive response to the hypothetical research project.

Within each attribute, the levels have been rescaled so that the least effective level within an attribute is given the value of zero.

**Figure 5.1 – Full list of attributes in the conjoint exercise, with their relative importance in driving approvability**

Attribute	Relative importance of attribute in driving approvability
Who the project is for?	16.8%
Why are they doing the project?	4.7%
Who could be included?	3.1%
Has permission been given?	11.8%
Is the social media data publically available?	22.0%
What kind of content would be looked at?	4.7%
What personal information would be used?	16.7%
How anonymous is the data?	20.2%

**Figure 5.2 – Full list of levels within attributes in the conjoint exercise, with their relative importance within attribute**

Levels within attribute	Relative importance of level within attribute
<b>Who the project is for?</b>	
A central government department	0.22
Public services or local councils	0.24
Researchers in universities and similar organisations	0.33
A charity	0.33
A private company	0.00
<b>Why are they doing the project?</b>	

To evaluate the success of a social media campaign	0.09
To assess public opinion on an issue	0.05
To review or act on comments about a product or service they deliver	0.09
To identify the most active or most well connected social media users in a network	0.00
To identify a group of people who could benefit from a service or product they deliver	0.02
<b>Who could be included?</b>	
Anyone who uses social media	0.01
Anyone on social media who has used a word, hashtag or phrase relevant to the project	0.06
Anyone on social media who has been identified as visiting a broad location relevant to the project	0.00
<b>Has permission been given?</b>	
All those who have agreed to the general terms and conditions of the social media site when then first signed up	0.00
Only those who have opted in to their data being used for research generally when signing up to the social media site	0.09
All excluding those who have opted out of their data being used for research generally when signing up to the social media site	0.03
All excluding those who have opted out of their data being used for this specific project	0.08
Only those who have opted in to their data being used for this specific project	0.23
<b>Is the social media data publically available?</b>	
Collecting posts that have already been made public on a site where anyone can see contributions regardless of whether they have an account	0.44
Collecting publically available posts, and posts where they can be seen by anyone who has an account for that social media site	0.38
Collecting all types of public and private posts, including private forums but excluding private messages between individuals	0.25
Collecting all types of public and private posts, including private messages between individuals.	0.00
<b>What kind of content would be looked at?</b>	
Purchasing habits, or information on the products or brands people like	0.00
Preferences and opinion of music, sport and entertainment	0.07

Attitudes on a particular social issue	0.04
Health behaviours	0.09
What types of information people need and are requesting	0.06
How and where services are being used	0.06
<b>What personal information would be used?</b>	
Personal information may be collected but will not be used	0.30
Only age will be used in order to remove posts from those considered to be under 16	0.31
User IDs and profile descriptions will be used to remove fake users and accounts from companies or institutions.	0.23
Age, gender and broad location will be used to compare different groups of people	0.33
As well as age, gender and broad location, other information will be used to help compare different groups of people (eg social grade or employment status)	0.21
Sensitive personal information (for example, sexuality and political affiliation) which is relevant to the project shall be used	0.00
<b>How anonymous is the data?</b>	
Only overall numbers are provided by the social media site to researchers. No raw data will be seen or published by the project.	0.37
The researcher will not see names/locations and IDs during analysis; no posts will be published.	0.40
Individual level posts will be seen by researchers (including author name and any other details posted by the author), but posts will not be published.	0.22
Individual level posts will be seen by researchers; individual social media posts will be published with IDs removed and content amended to reduce risk of re-identification	0.17
Individual level posts will be seen by researchers, individual social media posts will be published unedited (including author name and any other details posted by the author)	0.00

**Steve Ginnis**

Head of Digital Research  
Social Research Institute  
Ipsos MORI  
[steven.ginnis@ipsos.com](mailto:steven.ginnis@ipsos.com)

**Harry Evans**

Research Executive  
Social Research Institute  
Ipsos MORI  
[harry.evans@ipsos.com](mailto:harry.evans@ipsos.com)

**Jamie Bartlett**

Director  
Centre for the Analysis of Social  
Media at Demos  
[jamie.bartlett@democo.uk](mailto:jamie.bartlett@democo.uk)

**For more information**

Ipsos MORI  
3 Thomas More Square  
London E1W 1YW

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

**[www.ipsos-mori.com](http://www.ipsos-mori.com)**  
**[www.twitter.com/IpsosMORI](https://www.twitter.com/IpsosMORI)**

**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](http://www.ipsos-mori.com/wisdomofthecrowd)**