University of Essex

# Research data and social media

## Some general principles


Authors: Research and Enterprise Office



Publication date: July 2020

Amended: December 2024

Review date: December 2025

# Table of Contents

# Introduction

1. Social media is increasingly used in research. It can be perceived as risk free, but social media isn't simply a vast collection of interesting data: it is a set of places, communities and gathering points for people. Most of the people participating aren't there with any intention of providing data for research.
2. This doesn't mean that social media shouldn't be used in research. There are many ways in which social media can be a useful tool for research, or an interesting topic for research. Social media is an important part of modern life.
3. Although it exists in the virtual space most of the laws, rules, ethical principles and commons sense that apply in everyday life apply to social media too.
4. Because social media is groupings of people then research using it is likely to be research that requires ethical approval.
5. The ethics of social media research is a maturing area. Here we offer some general principles to help guide those considering the use of social media for their research.

# Consider whether social media is the best place to collect your data

6. Consider whether social media is the right place to carry out your research. Can you identify real benefits from using social media that you won't have access to elsewhere? Social media data shouldn't be used simply because it's convenient.
7. Social media users don't reflect the entire population, so research using social media data only could introduce unwanted bias into your sample. On social media people can present themselves as someone other than who they are – and some interactions on social media may be bots or algorithms and not people at all.

# The internet isn't a public place

8. There is a tendency to consider that anything on social media is public. Research shows that users often don't consider themselves to be in a public place when they participate in social media, in particular when the group is discussing a minority interest topic where the participants' anticipation may be that the group is only used by a small number of likeminded people.

9. Look for signs that the group may not consider itself to be having a public discussion. A group that requires registration, a password, or has a moderator which controls access to it is perhaps more likely to consider itself to be operating in a private space.
10. Consider the topic of discussion. Greater care should be taken where discussion centres around sensitive topics such as mental health, drug taking or sexual abuse.
11. Questions of whether online postings are public or private are determined to some extent by the online setting itself, and whether there is a reasonable expectation of privacy by users of a platform (as indicated by the British Psychological Society, 2013) – for example whether users expect their content not to go beyond certain circles: 'private' Facebook group (where one asks to join) can be considered private – even if the group has 300 users-, whereas an open Facebook page that anyone can join might be considered public by users; an open discussion on Twitter/X or Instagram in which people broadcast their opinions using a hashtag (in order to associate their thoughts on a subject with others' thoughts on the same subject) can be considered public.
12. Examples of open forums: Facebook open groups or pages; hashtag-led discussions on X/Twitter by public accounts; business/product pages on Instagram; YouTube videos and channels. These can be considered public forums as they are online forums where no sign-in or pre-approval of moderators is required to read the posts or watch the content.

# Consent

13. If you are asking for consent from participants then consider the extent to which it's possible for them to exercise their right to have data withdrawn. If you're using social media to interview individuals that might be easy, but if their data is a set of points within a very large data set it may not be possible. This should be made clear in your participant information.
14. Questions of whether the data is public or private relate to the extent to which researchers are ethically bound to seek informed consent from social media users (as stated in the ESRC Social Media Research: a guide to ethics). Individual users on public platforms (who are not public figures) should still be subjected to anonymisation or pseudonymisation.
15. In line with other institutions[1] and with GDPR's requirement, the data collected from social media must be justified for the study. Under the GDPR, consent may not be needed if other reasons can justify the processing/collection and use of data; however, consent for ethical purposes might still be necessary. Under the GDPR, the reasons to collect and use data can be under

[1] (see York)

Article 6(1) GDPR but also can be Article 89 GDPR. GDPR article 89 allows for collection and procession of data for scientific purposes or for historical research with some safeguards under the principle of data minimisation, that is the principle of using only data strictly necessary to the research purposes. When Article 89 GDPR applies, GDPR consent might is not needed. Consent for research ethical purposes, however, might still be necessary.

16. Acquiring informed consent is problematic with large data set and can seem virtually impossible in aggregate data containing thousands or even hundreds of thousands of data units. Additionally, in some cases, a social media user's data is accessed (viewed) without consent having been sought because the content is publicly accessible. 'Participants' in such research are rarely aware of their participation.
17. To help guide researchers we have listed below three broad headings of cases where: 1) consent should be sought from each relevant individual; 2) consent should be sought from a moderator, author, or platform owner; and 3) consent from an individual, author, or originator is not needed. Researchers should, however, also refer to section 29 "mitigation strategies".

# When consent should be sought individually

18. Researchers should seek consent for purposes of research from individual users for specific posts if any of the following applies**:**
    - When the user account is private, and the researcher had to actively ask and be accepted by the user for connection*.
    - If the content of their posting (on Twitter/X, Instagram, Snapchat, Facebook) is specific to the individual's beliefs or opinions (thus something they might have expressed in an interview).

    *If the user account is private but they posted on an open page/forum where registration is not necessary, opt-out consent might be preferrable (see paragraph 20).

# When consent should be sought through contacting a moderator, author, or platform owner

19. In the cases below, ethical implications should be considered and consent should be sought through contacting a moderator, author, or platform owner:
    - YouTube/TikTok videos that contain personal opinions but are clearly aimed at being disseminated for information purposes (i.e., videos of academic or journalist giving an informed opinion on historical event).
    - Open Facebook groups and pages where opinions of users might be shared to inform on specific topics (e.g. users commenting Police forces' Facebook pages).
    - Non-individual Twitter/X/Instagram user/pages that express opinions on specific topics.
20. Forms of opt-out consent should be considered on an ad-hoc basis in these cases, when the content relates to personal opinions and ideas. For example, the researcher can either post on

the page or send posters/users information identifying oneself as a researcher, explaining what one is using the data form (with information on anonymity and confidentiality) and setting a date to opt-out, after which date the consent is considered implied.

## When consent from the individual user, author, or originator is not needed

21. Consent is not needed in open forums[2] when the following apply:
    - The user's account is public and it can be followed without approval.
    - The user's post is linked to a hashtag or an open conversation thread, thus the user has renounced expectations of privacy and implicitly agrees to be scrutinised in public debates, and/or benefits from, or aims at, being part of public discussions for public interest, including scientific or historical research, ex article 89 GDPR.
    - If there is no previous interaction between the research and the users/participants, it may not be practical nor possible to ask for consent from many users: for instance, Twitter/X may classify requests for consent as spam and even Facebook or Instagram might flag any message from unconnected contacts under "Other/Hidden".
22. In line with article 6(4) GDPR on further processing, the context in which the data is initially found/collected can be considered in the decision to use that data for further processing whilst considering further safeguards such as pseudonymisation (article 6(4e)). When there are expectations for the posts and the accounts online of reaching out a larger audience, data can be used for further processing for the purposes of scientific and historical research.
23. Some open/public forums are particularly relevant for research on brands, corporate communication, advertising, where individuals who manage accounts may be difficult to contact or may not respond when contact has been made thus stalling the research. In these cases, ethical consent might be considered implied by the accepted and sought-after publicity of the page which indeed opens the page to a variety of public scrutiny of public interests, thus including scientific and historical research.

[2] Examples of open forums: Facebook open groups or pages; hashtag-led discussions on X/Twitter by public accounts; business/product pages on Instagram; YouTube videos and channels. These can be considered public forums as they are online forums where no sign-in or pre-approval of moderators is required to read the posts or watch the content.

# Balancing the risks and mitigation strategies

24. Be aware of the risks to participants that your research might pose, especially if you are not openly declaring your interest, or are not seeking informed consent, or where people are expressing opinions or experiences that they may not expect to be shared or made public. Wherever possible you should anonymise your findings.
25. Consider the impact of your research on young people or vulnerable adults. Even where a social media platform bans participation by those under a specified age (for many social media services this is 13 but always check as sites vary) it is not likely to have robust controls in place to absolutely prevent any children from taking part.
26. You should also have regard for any potential that your activity may have to damage the reputation of the University. If people feel that University of Essex researchers cannot be trusted, that will have a real impact on the ability of Essex researchers to recruit participants or attract future funding.
27. No activity is risk free, but as with any research the important thing is that you identify and acknowledge the risks and do as much as you can to mitigate those risks. Where the risks are greater than the benefits it may not be ethical to proceed.
28. There are three scenarios where the use of mitigation strategies is required:
    - Where consent is not being sought.
    - Where consent is implied, for example obtained through opt-out formulas for the collection of data.
    - Where there is passive online research, that is research that comprises passively 'scraping' or collecting data from social media sites, or reusing social media data collected by others, without direct interaction with participants.
29. Consider the following mitigating strategies:
    - Anonymise or pseudonymise social media users including direct and indirect qualifiers. Note that direct quotations scraped from social media sites (e.g. Twitter or Instagram posts) cannot be considered as anonymous data and cannot be pseudonymised as they can always be linked back to their author via simple web searches. Special consideration therefore needs to be given to protecting the identity of social media users.
    - Drawing inspiration from Bruckman (2004) as suggested by the code of ethics of the British Sociological Association - researchers might consider adopting a “moderate disguise”, whereby verbatim quotations can be avoided, no names nor pseudonyms, nor identifiable details are given, and some terms are purposefully changed from quotes to make attribution more difficult.

- Take care to ensure any additional rights of the social media users in their research (e.g., the right not to have words being used out of context).
- Consider what social media users might think about researchers repurposing their data without permission) and operate always along the lines of the “no-harm” principle.

30. For an indicative list of low risk research activities involving social media refer to Annex B. Even in low-risk cases, it is best practice to consider anonymisation or pseudonymisation through camouflage for example.

# Always read the small print

31. Social media platforms all have terms and conditions of use and may have policies on what material may be posted and for how data may be used. Terms and conditions may require you to sign up with your own name and to use or not use data in particular ways. You should respect these terms as providers may be able to take legal action against you if they believe you have breached them.
32. You should also be aware that terms can change with little or no notice, so ensure you check for changes that could affect your research. Providers may consider that you are agreeing to their terms simply by virtue of the fact that you are interacting with their platform.

# Watch out for deleted posts

33. From time-to-time individuals may remove or delete material they have posted. Even where the material had been posted in a way that made it clear it was intended to be public, removal of that material should be taken as a withdrawal of consent for that to be used. Check that material isn’t withdrawn during the course of your research.

# Protect yourself

34. Engaging in activity on social media is not risk free. Consider what would happen if you were identified as a researcher, or if people object to comments you make.
35. If you are participating in discussions, think about whether the account you are using can be traced back to other social media accounts that identify who you are, where you work or live, and details about your family. Consider adjusting the privacy settings of your own social media accounts to protect your personal information.
36. Make sure you have a plan for dealing with the impact of receiving high levels of abuse and threats. Identify in advance where you will seek help and support from, if you need it.

# Ethical approval

37. Ethical approval is likely to be required for some research that uses social media data. Approval may be granted by ethics officers, at sub-committees, or at Ethics Committee, depending on the level of risk involved, and how novel the use of social media is.

# Help us learn

38. Use of social media in research is growing and changing as social media grows and changes. Ethical opinion will also change as we come across more case studies. Help us develop a robust and realistic approach by discussing novel uses with us. Contact the REO Research Governance Team reo-governance@essex.ac.uk or discuss with your department's ethics officer.

# Further reading

39. ESRC Social Media Research: a guide to ethics
    University of Sheffield. Research Ethics Policy note no 14: research involving social media data
    UKRIO. Good practice in research: Internet-mediated research