Guidance for the use of social media data in research

This guidance should be read by researchers for the purpose of gathering and analysing social media user data. For information on using social media for recruitment, please access the handbook.

1. Classification

For the purposes of this guidance, **social media data** refers to any information generated, shared, communicated, or otherwise made available by individuals or organisations through social networking platforms, online communities, discussion forums, blogs, microblogging services, content-sharing platforms, messaging applications, and other digital environments that facilitate social interaction. Social media data may include text, images, audio, video, metadata, user profiles, interactions (e.g., comments, likes, shares, and reactions), and behavioural data associated with platform use. Please note that because the landscape of social media platforms is continually changing it is not possible to issue platform specific guidance for all or even most platforms. These guidelines should therefore be interpreted with the specific platform(s) used in mind.

Publicly available social media data: content that is openly accessible without registration, membership, or approval, and where users would reasonably expect their content to be publicly visible.

Restricted-access social media data: content accessible only to members of a group, followers, subscribers, or other authorised users, where individuals may have a greater expectation of privacy**.**

2. University research ethics approval: when is ethics approval **not** required?

Research involving social media data may be considered **ethically exempt** (i.e. not requiring University ethics review) if it meets one of the following scenarios.

Scenario 1: Social media data, including direct quotations

Your research may be ethically exempt if **all** of the following apply:

a) The data are obtained from publicly accessible websites, pages, forums, or groups, or from restricted-access groups where you have obtained explicit written permission from the group moderator or site owner.

b) The data do not relate to vulnerable or dependent groups, including children and young people. [1]

c) The data are collected and used in accordance with the terms and conditions of the relevant social media platform, website, or forum.

d) The research **does not** involve sensitive, confidential, restricted, or private topics (e.g. mental health conditions, political opinions, religious beliefs, criminal offences, gender identity, sexual orientation, or other sensitive personal information).

e) Direct quotations from users may be included in research outputs, provided the above conditions are met. [2]

Scenario 2: Social media data on sensitive topics

Your research may also be ethically exempt if **all** of the following apply:

a) The data are obtained from publicly accessible websites, pages, forums, or groups, or from restricted-access groups where you have obtained explicit written permission from the group moderator or site owner.

b) The data do not relate to vulnerable or dependent groups, including children and young people. [1]

c) The data are collected and used in accordance with the terms and conditions of the relevant social media platform, website, or forum.

d) The research concerns sensitive topics (e.g. mental health, political views, religion, criminal offending, gender identity, or sexual orientation) but **does not** include direct quotations or other potentially identifiable content from individual users. [3]

If there is any uncertainty about the sensitivity of the data, the expectations of privacy held by users, or the potential for identification of individuals, you can book an appointment at an ethics clinic**.**

Researchers are required to adhere to all University best practice guidelines, including data protection, even where formal ethics approval is not required. Any exemption remains subject to compliance with applicable legislation (including UK GDPR and the Data Protection Act 2018) and with any access conditions, copyright restrictions, or usage agreements imposed by the archive or repository. Researchers (staff or PGT/PGR only) can generate a formal letter of ethical exemption in the ERM system.

Important ethical considerations

[1] **Note on vulnerable or dependent groups**: Researchers are not expected to demonstrate that no vulnerable or dependent individuals are present within a social media platform, forum, or online community, as this will often be impossible to determine. Rather, the key consideration is whether the research specifically focuses on, targets, or is likely to elicit responses primarily from vulnerable or dependent groups, including children and young people under the age of 16. Research may be considered ethically exempt where the data do not have a specific focus on such groups and there is no reasonable expectation that they are highly represented within the dataset. Examples of research that would not normally meet the exemption criteria include studies of online communities specifically for children, young people, individuals with mental health conditions, people receiving social care services, or other groups who may be considered vulnerable or dependent

[2] **Note on use of direct quotations:** Researchers should carefully consider whether the inclusion of direct quotations from social media users is necessary to achieve the aims of the research. While the use of verbatim quotations may be appropriate in some circumstances, particularly when analysing publicly available content relating to non-sensitive topics, it is important to recognise that quotations can increase the risk of individual identification through online searching or contextual information. Researchers should also be aware that content posted in online spaces may remain identifiable even when usernames or other obvious identifiers are removed. As such, the inclusion of direct quotations requires careful ethical consideration, particularly where users may not reasonably expect their contributions to be reproduced in research outputs.

When deciding whether to include direct quotations, researchers should consider:

- The sensitivity of the topic under investigation.

- The reasonable expectations of privacy held by users, including whether content was posted in a public or restricted-access online space.
- The vulnerability of the individuals or communities represented in the data.
- The likelihood that a quotation could be traced back to an individual through search engines or other means.
- Whether the research objectives could be achieved using paraphrased extracts or summaries rather than verbatim quotations.

Where direct quotations are used, researchers should take reasonable steps to minimise the risk of identification. This may include:

- Removing or pseudonymising usernames, profile names, and other identifying information.
- Avoiding the inclusion of contextual details that could enable identification.
- Considering whether quotations can be paraphrased, partially quoted, or modified using synonyms while retaining their original meaning and analytical value.

[3] **Note on sensitive topics and non-quotation exemption:** It is recognised that research involving sensitive topics may, in some cases, meet ethical exemption criteria where direct quotations are not used. However, researchers should exercise caution in applying this principle, especially where data are drawn from online communities that may be publicly accessible in a technical sense but are not reasonably perceived as "public" by their members.

Some online spaces, particularly those associated with marginalised, niche, or identity-based communities, may not fall within conventional definitions of vulnerable or dependent groups, but may nonetheless involve participants who would not reasonably expect their content to be used for research purposes. There is evidence that such communities can experience research use of their content as intrusive or exploitative, even where data use is technically permissible under platform terms and conditions.

3. **University research ethics approval:** when **does** research involving social media data require ethical review?

Ethical review is required where research involving social media data presents more than minimal ethical risk. This includes situations where one or more of the following apply:

a) The data are sensitive, confidential, restricted, or private (e.g. information relating to mental health, political opinions, religious beliefs, criminal offences, gender identity, sexual orientation, or other sensitive personal matters) and the study **does** plan to include direct quotations or other potentially identifiable content from individual users.

b) Individuals are identifiable, or there is a reasonable risk that they could be identified, and anonymisation is not possible (not including public figures, politicians, celebrities, etc).

c) The research may reveal or lead to the disclosure of criminal activity. [4]

d) The data relate to, or are likely to include, vulnerable or dependent individuals, including children and young people.

e) The project involves developing, testing, or applying methods to re-identify previously de-identified data.

f) The use of the data could expose individuals to harm, distress, reputational damage, discrimination, or adverse personal, social, or economic consequences.

[4] **Note on risk of disclosure of criminal activity**: as all research carries some form of risk, this is asking you to consider whether any such risks are merely hypothetical or real. If a real and probable risk exists that you will uncovering information that you would be required to disclose to the authorities, the study will require ethical review.

.