

Using data from the internet and social media in research: ethics & consent

This is general guidance for researchers working with or collecting data from the internet and social media. Inevitably, it is a broad framework for thinking through ethical issues rather than a narrow prescription of what is expected in every case. As such, it is subject to dialogue, discussion and change as new modes of social media data come into play and/or legal and ethical frameworks are updated.

It was, and often still is, widely assumed that research on texts (without interaction with human subjects) raises no ethical concerns. Increasingly, however, we need to pay attention to the ethics of using data that has been generated by, and belongs to, private individuals and/or groups. This concern may hold, even if that data has appeared in the public sphere.

Much data posted on social media platforms is deemed to be in the public domain as users nominally agree to let third parties use their data when they agree to the *terms of service*. However there are strong *ethical* considerations (such as, ‘what were the users’ expectations about privacy, when they posted their data?’ and ‘might my use of this data be endangering to someone?’) that need to be taken into account.

Much publicly available social media data is actually *produced and owned by individuals* and even if they have ceded their ownership rights to platforms and/or third parties, reusing some of the data in research could cause them to become identifiable in the wider public sphere, when they were previously relatively low-profile, and therefore put them at greater risk of trolling and other kinds of violence than they would previously have been. This is especially the case in political situations where unintentional identification could have serious consequences, and where platform owners or controllers may not necessarily be reliably protecting users’ privacy.

Even where data is publicly available, it is usually ethically advisable to anonymise any information that could identify the poster, including by not using direct quotes (e.g. by paraphrasing), where this is compatible with the integrity of the research. Cases where this *is not, in general, necessary* include: if the person who has posted the information is a public figure (e.g. a politician, a celebrity, etc.); if the content of the postings are, in the main, non-contentious, or by contrast, if the posted content contains dehumanising, hateful and/or other material that it is in the public interest to draw attention to.

Do I need to obtain consent?

To help guide researchers we have listed below different types of texts/data under three broad headings of cases where: 1) informed consent is not required; 2) informed consent should usually be sought from a moderator/author/platform; and 3) informed consent (and/or copyright use permission) should be sought from each relevant individual. Researchers should, however, also refer to the note (§4) regarding researcher risk and mitigation.

1. Texts for which informed consent from the author/originator is NOT REQUIRED:

- Books (either print or online)
- Magazines (either print or online)
- Print or online newspapers
- Videos or transcripts of political speeches in the public domain
- Electronic news programmes on TV or the internet
- Fiction films and televised fiction programmes, game shows, chat shows etc.
- Documentaries released for public consumption (as long as one remains within fair-use copyright regs)
- Large and widely-cited political blogs (you just need to acknowledge the author)
- Letters published in newspapers
- Film reviews, Games reviews and other product reviews in the public domain by influencers or professional reviewers doing this for a living
- Ads on online platforms
- Other publicly available promotional and advertising material
- Memes and GIFs which are in the public domain
- Short video platform clips from public accounts of news organisations or other large governmental or non-governmental bodies (unless these have been taken without permission from private accounts)
- Short video platform clips from public accounts of politicians, celebrities and influencers (unless these have been taken without permission from private accounts or are recirculating data belittling or identifying private citizens who have not consented to the clips being shown)
- Data scraped from public feeds where all user metadata is redacted and tendencies are only summarised in aggregate form rather than by quoting individual comments which can lead to de-anonymisation.

For material that is in copyright, please keep in mind the requirement that re-use is fair and does not infringe the rights of the copyright holder (a concept known in the UK as fair dealing).

2. Texts/data for which informed consent should usually be sought either from a moderator or from an author

In the cases below, ethical implications should be considered and, where there is a risk of re-identification, attempted consent should usually be sought through contacting a moderator/author/platform (we can consider this consent to be collective - ie. a note on the site* or permission from the site or an "opt-out note" identifying oneself as a researcher and explaining what one is using the data for, which may need to be a 'pinned' post or repeated during the research process to ensure widest reach):

- Data drawn from small public blogs, particularly blogs by those in political circumstances, (especially where their writings could result in incarceration or state sanction for calling attention to injustice or hate).
- Data taken from photoblogs which contain identifiable images of third parties

- Large datasets of politically relevant private messages (e.g. WhatsApp discussions) containing identifying information if these are already in the public domain (Panama papers, Wikileaks)
- Data drawn from online platforms where quotation could lead to de-anonymisation and trolling of individuals. Particular *care needs to be taken with reposts or quotes to contact the original poster. Even anonymised posts can be searched through language searches and lead back to the original poster.*
- Data drawn from public platforms such as Facebook/ Instagram/Reddit (care might need to be *taken with re-posts or shared posts to contact the original poster if their posts are being used. Many posts can be searched through language searches and lead back to the original poster.*
- Comments published on public newspapers online, discussion boards or on platforms/ websites where users might or might not have pseudonyms. *Be aware that in some cases an attempt might need to be made to contact even pseudonymous users for consent.*
- Data drawn from private discussion boards or forums for which a person needs to sign up
- Film reviews, Games reviews and other product reviews in the public domain by individual social media users who are not professional reviewers
- IMDB, Rotten Tomatoes, Good Reads and other review site reviews (with metadata redacted).

*However, researchers should reflect on whether a note on a forum announcing the researcher's presence might have an effect on conversations in the forum, and if it may then be unethical to proceed. For example, a Reddit forum that is a safe space for a marginalised community might be seriously disrupted by the announcement of data collection. In these cases, the researcher might consider pursuing the research without consent, although they would need to consult with the Research Ethics Committee to ensure that they are effectively protecting the community both from identification and from any exploitation resulting from the research process itself.

3. Informed Consent should be sought individually in the case of:

- Anything taken from private Instagram/Facebook/Twitter/TikTok/Wechat/Redbook other platforms which has subsequently been made public but without consent of individuals
- Messages posted to boards of any semi-public semi-private online groups where the users are in a protected category or the topic is sensitive (e.g. BLM/Trans Rights/political opposition in some countries)
- Messages saved from Snapchat, Telegram, Viber, Signal, Vkontakt etc. (*Except in the case of fully anonymised messages where analysis contributes significantly to the public good over and above any potential detriment to the government/party/company/group/individual – e.g. Hate speech/ death threats/ sexual harassment*)
- Photo-sharing sites where human subjects are visible
- Messenger messages and Individual WhatsApp/Telegram/Signal/Vkontakt forwards (*Except in the case of fully anonymised WhatsApp forwards or Messenger posts where analysis contributes significantly to the public good over and above any potential detriment to the government/party/company/group/individual – e.g. Hate speech/ death threats/ sexual harassment*)

- Content creators who are at risk/vulnerable (especially where visibility of their material could result in incarceration or state sanction or calling attention to injustice or hate).

The social media environment is rapidly shifting, and research in the public interest can sometimes contravene the terms of service of platform companies. The researcher should consider their responsibility to the people creating the material that they draw on, and to themselves.

4. Researcher Risk and mitigation

Please ensure that in contacting individual users, groups or companies to ask for permission or consent you do not inadvertently expose yourself or your research team to harm or harassment. This means not using your personal or work email address if contacting organisations that are known or suspected to be involved in the circulation of race and death threats, Nazi, far right supremacist or other violent and hateful content, and extreme racist or misogynist material. If you have any doubts or concerns please consult the Research Ethics Committee (via research.ethics@lse.ac.uk) before you make any contact. Always have a separate contact email: We suggest a Protonmail account, and being circumspect about any data or contact details you divulge about yourself in online forums.

As noted at the end of section 2, there may be some circumstances where it may be safer and/or more ethically appropriate to *not* make contact for consent. Researchers should contact the Research Ethics Committee for advice.

Useful links

[LSE Social Media, Personal Data and Research Guidance](#)

University of Aberdeen: [Social Media Research: A Guide to Ethics](#)

(Note that this Guide pre-dates GDPR but is nevertheless a very useful resource)

Ahmed, W., Bath, P. and Demartini, G. (2017) Chapter 4: [Using Twitter as a Data Source: An Overview of Ethical, Legal and Methodological Challenges](#)

Williams, Burnap and Sloan, 2017: [Towards an Ethics Framework for Publishing Twitter Data in Social Research: Taking into Account Users' Views, Online Context and Algorithmic Estimation](#)

(Has a useful 'Decision flow chart' for publication of Twitter communications)

Association of Internet Researchers (AoIR) internet research ethics 2019: [Internet Research: Ethical Guidelines 3.0](#). Available at: <https://aoir.org/ethics/>

Intellectual Property Office: [Exceptions to copyright guidance](#) (Includes a section on fair dealing.)

Further resources

Foucault Welles, B., González-Bailón, S., & Hancock, J. (2020). The Ethics of Digital Research. In [The Oxford Handbook of Networked Communication](#): Oxford University Press.

Hewson, C. (2016). Ethics issues in digital methods research. In H. Snee, C. Hine, Y. Morey, S. Roberts, & H. Watson (Eds.), [Digital Methods for Social Science: An Interdisciplinary Guide to Research Innovation](#) (pp. 206–221). London: Palgrave Macmillan UK.

Hoser, B., & Nitschke, T. (2010). [Questions on ethics for research in the virtually connected world. Social Networks](#), 32(3), 180–186.

Zimmer, M. (2018). Addressing Conceptual Gaps in Big Data Research Ethics: An Application of Contextual Integrity. *Social Media + Society*. <https://doi.org/10.1177/2056305118768300>

Technical help/training: [LSE Digital Skills Lab](#)

See also [LSE data management guidance](#)

Contacts

Data management: contact the research data librarian via datalibrary@lse.ac.uk

Research ethics: contact Lyn Grove / Myriam Fellous-Sigrist via research.ethics@lse.ac.uk

Copyright: contact Wendy Lynwood via w.j.lynwood@lse.ac.uk