

Use of Social Media Data in Research Projects

This guidance document focuses on the data protection and legal aspects related to the use of [personal data](#) from social media platforms. Some ethical considerations are also introduced; however, this document is not intended to be exhaustive regarding the ethical aspects. If you have doubts, please contact the [ethics board](#) from your faculty.¹

This document applies to all data collected from social media platforms. Examples of social media platforms are Facebook, X (formerly Twitter), Instagram, Snapchat, Youtube, Reddit, etc.

Terms and conditions of the platform

Have you checked the terms and conditions of the social media platform concerning the gathering and use of the data?

Before starting to collect data from a social media platform, it is essential to verify the terms and conditions defined by the platform concerning the gathering, processing, and sharing of data. Terms and conditions of platforms commonly used in research:

- [X](#)
- [Meta](#)
- [Youtube](#)
- [Reddit](#)
- [TikTok](#)

Sometimes it can be hard to collect data that is still meaningful, whilst also complying with the requirements of these platforms. There are exceptions to these requirements for research, and the DCC can help you balance fulfilling the platform's requirements while still collecting significant data for your research.

¹Keep in mind that some of the recommendations in this guide might not apply to [medical research](#). In medical research other laws may apply which overrule instructions in this document, please contact the Bureau Ethische Toetsing Klinisch Onderzoek (BETKO) UMCG for more information (metc@umcg.nl)

Copyright law

Articles 3 and 4 of the new [EU Copyright Directive](#) provide exceptions to the copyright default rule for Text and Data Mining (TDM) research. The new directive allows researchers to make copies of data, as long as the access is lawful, and the project is carried out for the purpose of scientific research.

The copies should be stored with an appropriate level of security and may be archived after the end of the research project for scientific purposes, including verification. The European legislature also encourages the Member States and institutions to define commonly agreed good practices. In the Netherlands, the directive is implemented in articles [15n and 15o](#) of the Dutch Copyright Act.

Ethical considerations

The European Commission (DG Research and Innovation) in the document [Ethics and data protection](#) identifies the use of social media networks as “processing operations that may entail higher ethics risks”. For information, see section VII pg. 12-14.

Have you given due consideration to the level of risk the research poses to social media users, e.g., any potentially sensitive subject matter or potentially vulnerable social media users?

The [guidance document](#) on Ethics in Social Science and Humanities published by the European Commission enforces the sensibility of some of the personal data that may be disclosed on social media platforms: Researchers should “...consider the potential sensitivity of the data and whether the users could be harmed if their data are exposed to a new audience. Sensitive data postings relate to criminal offenses, use of illegal drugs, financial problems, mental health issues and suicidal feelings, extramarital sexual activity, controversial political opinions and activism”.

In line with this policy, X (formerly Twitter) does not allow the use of X-data to profile people and their potential behavior on sensitive topics, such as health (including pregnancy); negative financial status or condition; political affiliation or beliefs; racial or ethnic origin; religious or philosophical affiliation or beliefs; sex life or sexual orientation; trade union membership; alleged or actual commission of a crime” ([Sensitive information](#)). In addition, researchers are not allowed to perform off-X matching without consent from the X-user. More specifically, if you would like to connect X-data to another source (e.g., other public sources about this person or data you collected yourself), you are always obligated to ask for [informed consent](#) directly from the user.



Do you have a good understanding of the extent to which social media users are likely to perceive their posts to be public or private?

[The European Commission](#) underlines the relevance of the reasonable expectation of privacy of the data subjects. "Ascertain whether the data you intend to access is public (open platform vs password protect fora); if the forum is closed, contact the site or group administrator to see whether the forum is open to anyone who subscribes. Closed group and fora: if there is an expectation of privacy, seek permission from users to use the data and obtain their [informed consent](#)".

Did you know that you need to submit your project for ethical assessment if you use social media data in your research?

If you collect personal data for your research, it is necessary to submit your research project for ethical review. Because social media data often directly or indirectly identifies a person, social media data is almost always considered [personal data](#). This means that you should submit your project for ethical review to your faculty ethics board before you start collecting the data, even though the data is not collected directly from your data subjects or participants.

Privacy & Security

The GDPR does not distinguish between public and private data, only between personal and non-personal data. All personal data processed systematically falls under the GDPR. Most social media data can be considered as personally identifiable data under the GDPR.

The GDPR provides a framework for the protection of personal data and certain obligations for organisations (and their employees) that process personal data, however, it also recognizes the importance of scientific research. Therefore, the GDPR defines exceptions to certain obligations for research purposes ([art. 85](#) and [art. 89](#)). For researchers to meet the criteria for these exceptions it is necessary to demonstrate that the data are really necessary for academic expression and that the [principles of the GDPR](#) are taken into account.

Have you made an informed decision about the need to ask consent from the data subjects about the use of personal data?

In most social media studies, it is possible to contact data subjects directly, and ask them for consent to use their personal data in a research project. However, there are also situations in which asking for consent is not as straightforward or might even be impossible. Before you start collecting data without asking the

consent of the data subjects, make sure to [seek advice from a data steward](#) or [P&S coordinator](#). They could help you assess possible reasons not to ask consent. This is especially important if you collect [special categories of personal data](#).

When you are using personal data provided to you by a third party (e.g., X, Meta, etc.) and the data subjects have not expressly consented to its use in research projects, you should inform them that you have acquired the data and what you will be using them for ([Ethics and data protection](#), European Commission). This information should include ([art. 14 GDPR](#)):

- The contact details of the controller.
- The purposes of the processing for which the personal data are intended.
- The legal basis for the processing.
- The categories of personal data concerned.
- The recipients or categories of recipients of the personal data, if any.
- Where applicable, the intention to transfer personal data to a recipient in a third country or international organisation. This should include the existence or absence of an adequacy decision by the Commission, and appropriate safeguards to ensure that adequate protection of the personal data.

[Here](#) you can find more information on how you can inform your data subjects.

Have you considered the derogation in the case of the use of personal data for research purposes?

In the case of data processing for research purposes, the GDPR admits an exception to the obligation to inform your data subjects directly. The exception only applies when providing the information to the data subjects requires a disproportionate effort (the number of data subjects, the age of the data, and any appropriate safeguards adopted should be taken into consideration, [rec. 62](#)) or when the obligation is likely to render impossible or seriously impair the achievement of the objectives of the processing (in this case being able to perform the research project).

If this exception applies to your research, you should still invest effort in making sure that your data subjects could find information about your research project. This may include making the information that should be provided to the data subjects publicly available (for example having a project webpage).

Have you given full consideration concerning storage, sharing, and archiving of the social media data?

The public information gathered from social media platforms may contain personal and sensitive data. For that reason, they are subjected to the GDPR. Before processing personal data, you must guarantee that you have all the technical and organisational measures in place needed to ensure the freedoms and rights of the social media users. More information on standard [UG IT solutions](#) for data protection.

FAIR data

Did you consider how to make social media data FAIR?

FAIR stands for **F**indability, **A**ccessibility, **I**nteroperability, and **R**euse of data. The [FAIR principles](#) are the standard for responsible data management and practising open science. FAIR does not necessarily mean that your data is openly available to everyone. If you consider making the social media data you collected FAIR, it is important to think about how you want to make the data available and under what conditions.

The new [EU Copyright Directive](#) as well as the GDPR provide exceptions for research, making it possible to share data from social media platforms with other researchers, without asking for explicit consent from data subjects. Researchers are not allowed, however, to share this data for other purposes than academic research. To facilitate data sharing, the DCC can help you design [a procedure](#) for restricted access that includes the [terms and conditions](#) under which the data can be accessed by third parties, as well as an assessment procedure to verify the responsible use of the data by third parties.

An example of how you can make social media research data FAIR is the dataset of UG Faculty of Arts researcher dr. Tommaso Caselli: "[DALC - Dutch Abusive Language Corpus](#)".