

# SOCIAL MEDIA AND RESEARCH



Background image by rawpixel.com - www.freepik.com



**seriss**  
WWW.SERISS.EU

SYNERGIES FOR EUROPE'S  
RESEARCH INFRASTRUCTURES  
IN THE SOCIAL SCIENCES

This project has received funding from  
the European Union's Horizon 2020  
research and innovation programme  
under grant agreement No 654221.



# 10 Legal and Ethical issues to consider

## 1 Do I need to obtain research ethics or legal approval?

Research involving human subjects often needs to be subject to a research ethics review. Requirements for an ethics review vary between countries, institutions and disciplines. Large scale studies involving harvesting from social media platforms raise the question as to whether the data gathered are to be treated as text (no human input) or as human subjects' data. In general, if there has been human input at some point in the process of creation, then the consequences for individuals and communities should be considered, even if there is no requirement for a formal research ethics review.

## 2 Can the harvesting of social media data qualify as "processing personal data"?

Yes. In addition to processing names, social media platforms will often include text, videos, audio and images etc. that can be entered into a search engine and traced back to a living individual, and will thus be subject to the General Data Protection Regulation (GDPR). If data contains personal information, make sure to be compliant with the privacy rules.

## 3 Do I need to obtain informed consent from the social media users?

It will depend on the research question and nature of data collection and analysis. If information is collected directly from individuals the research will be understood as involving humans and thus ethically often requires informed consent. E.g. if linking survey data to Twitter accounts, informed consent will be necessary. There are some exceptions in the GDPR that permit harvesting without consent. The ability to seek informed consent for large scale data sets should be weighed against its feasibility, the expected benefits of the study and data security measures in place. Informed consent in research ethics terms differs from consent as a legal basis in the GDPR'.

## 4 What should the informed consent cover?

Depending on the nature of the research project and the data collection activities of the researcher, social media users should know what information will be collected, for what purpose and security arrangements. Information about whether users can access and have the right to delete the data gathered and the right not to take part should be given. It is good practice to have separate consent for research, dissemination, and data sharing. The GDPR lists the categories of information that must be provided to a data subject regarding the processing of their personal data where it is collected from the data subject or obtained from another source.

## 5 How can I assess if the source is public or private?

Requirements to register, the presence of a moderator, password protection, and sensitivity of content etc. all suggest some intent toward private communication. Open access, institutional accounts, and broadcast messages all suggest more public intentions. Whether the users generating social media content consider their posts to be public or private will be one of the factors determining whether informed consent for use of the content is required.

## 6 Do Terms of Service (ToS) address the issues of research use?

This depends on the platform in use. Please note that jurisdictions within and outside the EU/EEA may have different rules for using the data. E.g. Twitter has different rules for US and non-US users. Other services have different conditions for different countries

incorporated into one document. Researchers are advised to check the platforms' ToS.

## 7 Are the data collected protected by IPR?

Individual records can be protected by copyright. The whole database or its parts can be protected by sui generis database rights. Some data are not protected at all and can be harvested at will. EU law allows certain statutory exceptions for non-commercial research purposes. Compiling, validating or structuring the data can give rise to a completely new IPR owned by the consortium or its individual members. Adding content to the database usually requires consent from the IPR holder. If the new IPR was created in the phase of research, the decision to publish must not violate the rights of any co-author. Researchers should be aware of the citation requirements for the used in order to avoid a charge of academic misconduct. Note that there is variation in national laws.

## 8 What about possible risks, sensitive topics and vulnerable groups?

Particular attention must be paid to vulnerable categories of individuals such as children, patients, minorities, people unable to give consent, etc., and also to sensitive topics (e.g. health, religion, political views, sexuality etc.). When harvesting/scraping data, the inability to determine the mental capacity and to verify age and vulnerability of social media users should be explicitly considered and safeguards of their interests specified. The risk of harm to the participants as a result of participation or of the inclusion of information that identifies them being included in the research should also be considered. You must always consider if the potential benefits of the research offset the additional risks.

## 9 Do I need to anonymise the collected social media data?

Not necessarily. If participants consent to their data being shared then anonymisation may not be required. Justification of why the research data will not be anonymised/pseudonymised should then be provided and confidentiality assurances should take account of the challenges of maintaining confidentiality. If not relying on the legal basis of consent, one should take into account whether social media data are considered 'public' or 'private' by the user or whether topics are considered 'sensitive' before assessing whether to anonymise or not.

## 10 Can I archive and share social media data?

Social media data are hosted by social media sites which have different ToS, privacy policies, and developer agreements that influence how social media data can be archived and shared. There are exceptions for research, e.g. Twitter permits sharing of Tweet IDs. It is recommended to study in detail the social media platforms' ToS. Note that ethical requirements may prevent sharing, e.g., if a participant has consented, but the researcher believes s/he does not fully understand the risks of data sharing. When sharing data, it is necessary to consider aspects of the consent obtained, and further, to consider the possible risks of sharing, i.e. it can make data more visible, and thereby increase possibilities for linkage and disclosure. These risks need to be weighed against benefits such as open data, research transparency and enabling wider use of data.



**seriss**

SYNERGIES FOR EUROPE'S  
RESEARCH INFRASTRUCTURES  
IN THE SOCIAL SCIENCES

*Legal disclaimer: This does not constitute or should not be construed as legal advice and/or guidance.*