

Possibilities and limitations of Internet research: A legal framework

Katrine Utaaker Segadal

Norwegian Social Science Data Services (NSD)

katrine.segadal@nsd.uib.no

As the data protection official for research for some 150 Norwegian research and educational institutions, NSD has noticed an increase in research conducted on data harvested from the Internet in recent years. Today, the Internet is an important arena for self-expression. Our social and political life is increasingly happening online. This will have a major impact on how we understand the society in which we live and the opportunities for future generations to reconstruct the history of the 21st century.

Thus, data generated by the growth in electronic communications, use of Internet and web-based services and the emergence of a digital economy are increasingly valuable resources for researchers across many disciplines. At the same time there is a great need for knowledge and awareness of both legal requirements and ethical challenges related to the use of these new data sources, and for an understanding of the data's quality and scientific value.

In addition to the increased volume of this type of research, we have also seen a shift in focus. At first, the Internet and social media were studied mainly as a tool. The studies often concentrated on how the Internet worked as an instrument in e.g. education, health services or online dating. The methodological approach was usually interviews or surveys based on informed consent from the research subjects.

Today, the trend is to study the Internet as an arena for expressing or negotiating identity, often through projects of a sensitive character (e.g. political opinion, religious beliefs, health). Data are usually collected from social media such as blogs, social networking sites or virtual game worlds. These sources are publicly available, and often research is conducted without informed consent from the persons being studied.

This development raises questions such as: Which rules and regulations apply to research on personal data collected from the Internet? In which cases is it legal and ethical to conduct research on such data without the consent of the data subjects? When is it necessary to inform the data subjects of their involvement in a research project and when should this information be accompanied by an opportunity to refuse to be the object of research? These issues will be discussed in further detail in the following.

New European legislation in the making

The use of new types of data, such as those collected online and so-called Big Data, rank high on the international agenda. The OECD Global Science Forum points out the challenges related to the large amounts of digital data that are being generated from new sources such as the Internet although these new forms of personal data can provide important insights,

the use of those data as research resources may pose risks to individuals' privacy, particularly in case of inadvertent disclosure of

the identities of the individuals concerned. There is a need for greater transparency in the research use of new forms of data, maximizing the gains in knowledge derived from such data while minimizing the risks to individuals' privacy, seeking to retain public confidence in scientific research which makes use of new forms of data.¹⁶

To address this challenge, the forum recommends that research funding agencies and data protection authorities collaborate to develop an international framework that protects individuals' privacy and at the same time promotes research.

The European Commission has proposed a comprehensive reform of the EU's 1995 data protection rules,¹⁷ and we might see the results of this in the relatively near future if and when the new General Data Protection Regulation is implemented in Norwegian law. EU Justice Commissioner Viviane Reding said on the occasion of the legislative proposal:

17 years ago less than 1 % of Europeans used the Internet. Today, vast amounts of personal data are transferred and exchanged, across continents and around the globe in fractions of seconds. The protection of personal data is a fundamental right for all Europeans, but citizens do not always feel in full control of their personal data. My proposals will help build trust in online services because people will be better informed about their rights and in more control of their information¹⁸.

We will not go further into this, but just briefly mention that the new digital media, and the Internet as an increasingly significant

16 OECD Global Science Forum (2013): «New Data for Understanding the Human Condition: International Perspectives», page 2.

17 http://ec.europa.eu/justice/newsroom/data-protection/news/120125_en.htm

18 European Commission – IP/12/46 25/01/2012 – Press release: «Commission proposes a comprehensive reform of data protection rules to increase users' control of their data and to cut costs for businesses» http://europa.eu/rapid/press-release_IP-12-46_en.htm?locale=en

data source, are important reasons why the EU is currently upgrading the data protection regulation from directive to law. A regulation is a binding legislative act and must be applied in its entirety across the EU. **Directives** lay down certain results that must be achieved by all EU countries, but the individual Member State is free to decide how to transpose directives into national laws.

The demand for harmonization of rules and practices is high, particularly related to the use of data generated by or in relation to global communication networks such as the Internet. This type of network weakens the significance of national borders and the impact of national policies and legislation on the protection of personal data.

NSD's general impression of the Commission's initial proposal was that it would not lead to any dramatic changes for Norwegian research. The reason is primarily that Norwegian data protection legislation and the way this legislation is practised in relation to research are stringent, and that we have a high degree of protection of personal data in Norway. However, some of the recently proposed amendments to the Commission's proposal made by the European Parliament may have negative consequences for parts of the research sector if being transposed into EU legislation. There is a clear tendency in this proposal towards strengthening the right to personal privacy and control of own personal data at the expense of researchers access to such data.

The current Norwegian legal framework

In Norway there are primarily three laws (i.e. the Personal Data Act, the Personal Health Data Filing System Act, and the Health Research Act) that regulate the use of personal data for research purposes. In cases of collecting research data from the Internet, it is mainly the Personal Data Act that applies, so our focus will be on this.

Although the regulations are not always crystal clear, they provide important guidelines on how data initially produced for other purposes can be used for research purposes. The regulations may set limitations on usage, but they also provide many opportunities for valuable research.

The Personal Data Act is technology-neutral, although it is not necessarily adapted and updated with regard to technological development. The law applies to the processing of personal data, irrespective of source. It is applicable regardless of whether the data are self-reported, collected from a confidential source or gathered from a public registry. This implies that a research project is subject to notification to the Data Inspectorate or Data Protection Official when personal data are processed by electronic means, even if the information is gathered from a publicly available source on the Internet.

Data protection principles online

The purpose of the Personal Data Act is to protect the individual's privacy from being violated through the processing of personal data.¹⁹ Key principles of data protection are the need to protect personal integrity and private life, to ensure individuals' control of their own personal data and to guarantee that personal data are of adequate quality. These important principles are the basis for the interpretation of other provisions in the Personal Data Act, and place restrictions on research on information obtained from the Internet. They are closely related to essential principles of research ethics such as the demand to respect human dignity, integrity, freedom and right to participate, and the obligation to prevent harm and suffering.²⁰

19 Act of 14 April 2000 No. 31 relating to the processing of personal data, section one.

20 The National Committee for Research Ethics in the Social Sciences and the Humanities (NESH) (2010): *Guidelines for research ethics in the social sciences, law and the humanities*, chapter B.

These data protection principles are applicable irrespective of methods for data collection and data sources involved in the research. Consequently, they also apply to data collection online. However, handling these fundamental data protection principles in this context presents the researcher with certain challenges. Should one expect those who express themselves online to understand that their personal data may be used for purposes other than those originally intended, such as research? Have they given up control of their personal data when publishing on the Internet? And how does the availability of the data affect the researchers' duty to protect the privacy and personal integrity of the persons being studied? As a researcher it might be helpful to consider the following when trying to figure out these issues.

First of all, from what type of medium are the data obtained? Data collected from a public forum for debate will probably require fewer safeguards than a Facebook page with access restrictions. Second, does the data have the character of a public statement or is it reasonable to presume that the information is meant to be of a private and personal kind? And further, should the information be safeguarded considering the data subject's best interests, irrespective of medium or the author's assumptions? Sensitive data (e.g. information related to health) might require a high level of protection, even though it is published as part of a public statement at an open webpage. One might claim that the researcher has a special responsibility to protect sensitive personal data although the subject has disclosed it voluntarily, bearing in mind that the person might not view the full consequences of publishing the information online.

A fourth important factor is whether the data subject is a child or an adult. Information concerning children is subject to strict regulations. In 2012 a new provision of the Personal Data Act

was implemented. The Act states that «[p]ersonal data relating to children shall not be processed in a manner that is indefensible in respect of the best interests of the child».²¹ In the draft bill this provision is partly justified by the challenges associated with children's use of new technology. The ministry especially points out problems related to adults' attitudes towards publishing images of and information about minors. The most serious violations of children's privacy are increasingly committed by adults. This provision could also apply to a website's use of material voluntarily published by children themselves, if this use is indefensible.²² In this case further use by researchers might be illegal and unethical.

Furthermore, in relation to this, one should consider whether the information is published by the data subject itself or by a third party. If there already has been a breach of data protection principles, which may be the case when it comes to information published by another person than the data subject, researchers should be particularly careful.

Research without consent

As a default rule, personal data cannot legally be used for purposes other than the original one, unless the data subject consents.²³ And it is fair to assume that those who have published data on the Internet have not done so with the purpose of being the object of research. However, the Personal Data Act includes a number of exemptions from the general rule for research.

21 Act of 14 April 2000 No. 31 relating to the processing of personal data, section eleven, third paragraph.

22 Prop. 47 L (2011–2012) Proposisjon til Stortinget (forslag til lovvedtak) Endringer i personopplysningsloven, Chapter five.

23 Act of 14 April 2000 No. 31 relating to the processing of personal data, section eleven, first paragraph, litra c.

An important provision in this respect is that

subsequent processing of personal data for historical, statistical or scientific purposes is not deemed to be incompatible with the original purposes of the collection of the data, cf. first paragraph, *litra c*, if the public interest in the processing being carried out clearly exceeds the disadvantages this may entail for natural persons.²⁴

Thus, research activities are, per definition, not considered incompatible with the original purpose. Science is afforded a special position in the current legal framework, and this provision might be seen as a fundamental principle guaranteeing further use of data for research purposes regardless of the original reason for their production. This leaves open the possibility to conduct research on information obtained online without consent.

Having said that, as a general rule personal data may only be processed when the data subject has freely given an informed consent.²⁵ When designing a research project, the starting point should always be to consider whether consent should and could be obtained prior to the collection of data.

However, another provision offers a direct exemption from the main rule. Even sensitive personal data may be processed if this «is necessary for historical, statistical or scientific purposes, and the public interest in such processing being carried out clearly exceeds the disadvantages it might entail for the natural person».²⁶

Firstly, this entails that the planned processing of personal data must be required to answer relevant research questions. The

24 Act of 14 April 2000 No. 31 relating to the processing of personal data, section eleven, second paragraph.

25 Act of 14 April 2000 No. 31 relating to the processing of personal data, section eight, first paragraph and section nine, *litra a*.

26 Act of 14 April 2000 No. 31 relating to the processing of personal data, section nine, *litra h*.

researcher has to make it probable that the planned harvesting of data from the Internet is absolutely necessary to achieve the purpose of the study.

Secondly, if the necessity requirement is met, the law requires a balancing of interests between the project's societal value and any possible inconvenience for the individuals who are subject to research. It is crucial that the research will benefit society in some way or at least be an advantage for the group that is being researched. When assessing the probable disadvantages for the data subject, relevant factors are the degree of sensitivity, the author's presumed purpose of publishing (e.g. private or freedom of expression), the source (e.g. forum with restricted access or publicly available), who the data subject is (e.g. child, vulnerable/disadvantaged individual, adult) and the degree to which the data subject is identifiable.

Another important aspect to keep in mind in deciding for or against the processing of personal data for research purposes without consent is whether or not it will be possible to publish the results anonymously. This may be a challenge if one wishes to publish direct quotes, as these will be searchable on the Internet. It is also important to note that pseudonyms or nicknames may be identifiable because they may be used in various contexts online and hence function as a digital identity.

Moreover, an important factor is whether the data subject is informed of the research project. Having information and the opportunity to object to being included in the research will limit the disadvantages because the individual will then be able to exercise control over his or her own personal data. This may be a weighty argument for exempting a research project from the consent requirement. However, the right to object is not in itself considered a valid consent under the Personal Data Act. A valid consent must be a freely given, active and specific declaration by

the data subject to the effect that he or she agrees to the processing of personal data relating to him or her.²⁷

If a research project includes the processing of highly sensitive data (e.g. from blogs about personal experiences with eating disorders, self-harm or the like), and the information being processed is detailed enough to make the bloggers identifiable (a factor one generally must take into account), it may be difficult to exempt from the requirement for consent. This holds particularly if publishing direct quotes is deemed necessary by the researcher, so that it will be hard to guarantee anonymity in the publication. If the authors are minors, the threshold for not obtaining consent should be even higher. Adolescents over the age of sixteen will often be considered mature enough to give an independent consent in such cases. However, when obtaining consent online, it might be a challenge to be certain of the actual age of the person granting consent.

In the case of research on utterances from Twitter, which involves thousands of people, that focus on e.g. elections (which in the legal sense may be sensitive information about political views), there will clearly be legitimate reasons not to obtain consent from the data subjects considering the public character of both the source and content of the data.

In between these two rather clear-cut examples lies a range of grey areas which require concrete assessments in each case. My main message is that it certainly can be legal to conduct research on personal information obtained from the Internet without consent, as long as the researcher can justify the necessity and the benefits for the public clearly outweigh the disadvantages for the individual. The violation of personal privacy is often minimal when data is harvested on the Internet for research purposes. However, research

27 Act of 14 April 2000 No. 31 relating to the processing of personal data, section two, number seven.

on social media with restricted access differs somewhat from most other contexts in this respect. It is plausible that individuals who publish information about themselves under such circumstances might think that they are acting on a «private» arena, and that their purpose is to interact with a closed group of people. This indicates that the threshold should be slightly higher when considering not obtaining consent in such cases.

Obligation to provide information

The general rule is that the research subjects should be informed about the research. This is the case even if the exception clause from the requirement for consent applies. The basis for this rule is the fundamental right to exercise control over one's own personal data, and the assumption that the data subject should have the right to object to the intended processing of her personal data. However, a relevant exemption provision allows for research to be conducted without informing the data subjects: «The data subject is not entitled to notification [...] if [...] notification is impossible or disproportionately difficult».²⁸

If it is not feasible to get in touch with the affected individuals because it is not possible to obtain contact information or to communicate through the website, there is of course no way to provide those individuals with information.

Relevant factors in the assessment of whether it is disproportionately difficult to provide information are, on the one hand, the number of data subjects and the effort, either in terms of time or money, that providing information would entail. However, technological developments are and will most likely make it increasingly easier to distribute information to thousands of individuals at the

28 Act of 14 April 2000 No. 31 relating to the processing of personal data, section 20, second paragraph, *litra b*.

same time at no extra cost. The violation of personal privacy is not automatically less because the data subjects are numerous.

On the other hand, one should consider what use the data subject will have of being informed of the research project. Is it likely that the research subjects would wish to object if they had the opportunity to do so? If that is a reasonable assumption, information should be provided. Another important question is to what extent the research subjects will benefit from being able to refuse to be part of the research project. This will depend on the type of data being processed and how sensitive the information is. If what is at stake is very sensitive information, data protection principles indicate that information should be provided. This holds independently of whether the data subject initially has made the information publicly available.

Legally, the obligation to provide information is met only if the researcher gives individual information in such a way that the information is certain to reach the intended receiver. But in some cases, it may be appropriate to provide public information instead. This may be done through collective information published on the website from which the data is collected. It is not guaranteed that this information will reach everyone in the same way as when it is communicated directly by mail, email or other channels, but public information is nevertheless a measure that, to a certain extent, can justify exemptions from the requirement of individual information.

Conclusion

The Personal Data Act is applicable irrespective of the data source. The regulations do not distinguish between data harvested from the Internet and other sources (such as administrative registers). However, the legal framework leaves open a range of possibilities

for conducting research on information obtained online. It might be challenging, though, to apply the rules in this context.

The main rule is that the processing of personal information should be based on informed consent. But a number of exemptions make it possible to conduct research on personal information obtained from the Internet without consent, as long as the researcher can justify the necessity, and the benefits for the public clearly outweigh the disadvantages for the individual. The violation of personal privacy might often be limited when data is harvested on the Internet for research purposes.

References

European Commission – IP/12/46 (25/01/2012) – Press release:

«Commission proposes a comprehensive reform of data protection rules to increase users' control of their data and to cut costs for businesses» http://europa.eu/rapid/press-release_IP-12-46_en.htm?locale=en (Accessed 13 February, 2014).

Justis- og politidepartementet (2011): Prop. 47 L (2011–2012)

Proposisjon til Stortinget (forslag til lovvedtak) Endringer i personopplysningsloven.

Ministry of Justice and Public Security: Personal Data Act, Act of 14

April 2000 No. 31 relating to the processing of personal data.

The National Committee for Research Ethics in the Social Sciences and

the Humanities (2010): «Guidelines for research ethics in the social sciences, law and the humanities».

OECD Global Science Forum (2013): «New Data for Understanding the Human Condition: International Perspectives».