Investigating ethical issues in research concerned with data sources involving sensitive personal data

Sai Prashanth Josyula

June 18, 2017

Contents

1 Introduction

Computer science and its scientific artifacts have an impact on nearly every aspect of modern life. Computer science researchers are expected to ethically conduct and report their research [8]. Data analysis is typically required for scientifically validating results of a research study. From a legal perspective, when dealing with personal data, researchers are constrained to national and international laws. From an ethical perspective, researchers face several ethical dilemmas as they are often ethically obliged to certain guidelines though they are not required by the law.

When research involves collecting data from various individuals, researchers are expected to maintain high ethical standards as recommended by professional bodies and international organisations. The data collected from individuals may hold personal, sensitive and confidential information. This does not always mean that all the data obtained from the participants is personal or confidential. The collected data could be shared ethically and legally. We investigate the ethical issues in research dealing with data sources involving sensitive personal data. This document is intended to assist a researcher aspiring to conduct ethical research when dealing with sensitive personal data.

2 Principles for conducting ethical research

2.1 Planning to conduct ethical research

The European Commission provides a framework for conducting research in ethical way. To ethically conduct research involving personal data sources, the researcher should consider the following questions at the time of the research proposal. The following answers to crucial questions [3] assist researchers dealing with personal data to plan their research.

1. *What kind of data are involved within the research?*
   The required personal data can be categorized as previously collected data with usage history or data that is to be collected exclusively for the current research.

2. *How to get the necessary legal permission to obtain and process the data?*
   Researcher should investigate the necessity for informed consent when he/she is dealing with personal data sources. In cases where the data has already been collected (e.g. in a preceding research project), obtaining a new informed consent might be necessary. These details should be discussed with local or the national data protection agency.

3. *Data preservation and usage*

   (a) *How long will the collected data be stored and when will it be irreversibly destroyed?*
   The researcher should decide how the data is conserved and stored. The cost for conversion and destruction should be considered beforehand.

   (b) *For how long will the collected data be used?*
   The researcher needs to decide the usage time beforehand. Typically, the data should be stored as along as the research project lasts. If researcher wants to use the data beyond the lifetime of the project, the data usage must be closely supervised.

4. *Data Security*

   (a) *How will the collected personal data be securely accessed?*
   The policies of security should be analysed well in advance and the process of secured access (eg: passwords, data encryption etc) should be determined before data collection.

   (b) *Which data structure and format will be used to securely store the data?*
   The researcher should have good knowledge on different data structures (e.g. databases) and data formats used to secure the data storage.

   (c) *How will the data be securely stored?*
   The collected personal data should not be stored on a flash drive (USB stick) or other media that are easily to lose or access. A computer server that is not connected to a Wide Area Network (WAN) or a hard disk should be preferred over flash drives.

5. *Data transfer*

   (a) *How will data transfer be monitored?*
   Personal data transfer within the research project, especially to a researcher based in a non-EU country must be dealt carefully. The EU legislation requires such data transfers be undertaken only to places where the level of data protection is at least equivalent to that of the EU area.

2.2 Collection and preservation of personal data

The Organisation for Economic Co-operation and Development (OECD) articulated principles of privacy and data protection to protect the privacy of personal data. Though the articulated principles

are not directly enforceable, nearly all existing privacy and data sharing laws and policies have adopted at least some of those principles [1]. We present an abridged version of those principles which can be adopted by a researcher in order to ethically deal with collection, preservation and sharing of personal data sources during his/her research studies.

1. *Collection Limitation Principle*: There should be limits to the collection of personal data and it should be obtained lawfully and ethically.
2. *Data Quality Principle*: The collected personal data should be relevant to the research. The collected data should be accurate, complete and kept up-to-date.
3. *Purpose Specification Principle*: The purpose of the collection of personal data should be informed to the subjects, and any change in the purpose should be notified.
4. *Use Limitation Principle*: The collected personal data should not be disclosed. It can only be disclosed after obtaining appropriate consent or by the authority of law.
5. *Security Safeguards Principle*: The personal data obtained by the researcher should be safeguarded from damage, disclosure, and misuse.
6. *Openness Principle*: There should be a general policy of openness on the development, practices and policies related to Personal data.
7. *Individual Participation Principle*: The individual participants should have the following rights:
   (a) To obtain information regarding the presence of data relating to him/her with the data controller.
   (b) To have communicated to them, their personal data in a reasonable manner and in a form that is readily intelligible to them.
   (c) To be given reasons if a request made as per 7a or 7b is denied and to be able to challenge such a denial.
   (d) To challenge the data related to them and if successful, to have the data erased, corrected, completed or amended.
8. *Accountability Principle*: A data controller should be accountable for complying with measures which fulfil the above-mentioned principles of privacy and data protection.

2.3 Sharing of personal data

When a researcher deals with sensitive personal data or confidential data as a part of his research, there can be a perceived tension between sharing and protecting the personal data. However, typically, such data can be shared while conforming to principles of research ethics and without violating data protection rules and regulations [2]. Research Ethics Committees are organisations that help researchers to conduct the research in an ethical way. They give the following advices to the researcher dealing with personal data: [2]

1. It is crucial to distinguish between collected personal data and research data in general.
2. Most research data obtained from individuals can be shared without breaching confidentiality.
3. Personal data should not be published or shared unless the required consent has been obtained.
4. Information that could directly or indirectly lead to identifying an individual may be excluded from sharing.
5. Personal sensitive data can be shared if suitable procedures, precautions and safeguards are followed.
6. Anonymized data does not come under the data protection laws.

2.4 Anonymization of personal data

Personal data should be anonymized in order to protect the data from disclosure. Few ways to anonymize the data are [2]:

i) Removal of the direct identifiers (e.g. name or address).

ii) Reduction of the precision of information or variable (e.g. replacing date of birth by age).

iii) Usage of pseudonyms.

In order to manage anonymization of data, the researcher should [2]:

a) Retain the original versions of data.

b) Create the logs of the replacement made to the data.

c) Store these logs separately.

3 Challenges in conducting Ethical Research

Researchers often face issues and challenges in conducting research in an ethical way. The different methodological and legal concerns [4, 6] in conducting ethical research are given below:

Methodological Perspective
Challenges to ethics due to practicalities

a) Research dealing with large number of subjects makes it difficult to obtain the informed consent. In some cases, a single case may require more than one informed consent. It is very tedious for the researcher to get the consent if the subject is not available (moved temporarily or unable to access the person).

b) Different organisational barriers lead to insufficient access to data.

Challenges to research due to ethics

a) When a substantial portion of the subjects do not provide the consent or refuse the consent, researchers do not have enough data available with them. In such cases, it is not always possible for the researchers to obtain the evidences in order to reach conclusions. Thus, insufficient data is available for analysis due to refused consent from the individuals. This is a threat to the scientific validity of the study.

b) In registry-based research, researchers rarely request informed consent from the participants [6]. This is because in the context of large-scale observational research, such requirements on the study lead to severe selection bias [6].

Legal Perspective
Challenge to ethics due to ambiguous definitions

a) The *Data Protection Act* mentions that if something is done in public interest, then informed consent does not need to be taken. But the terms "public interest" and "substantial public interest" have not been defined in the Act [4].

3.1 Challenges to Ethical Research in Nordic Countries

A researcher should make sure that he/she does not collect data that is already available. In Nordic countries, due to the use of unique personal identifiers, individual data available through various means are linkable [7], and such data is of a high scientific value. In these countries, for certain types of research, a researcher has the opportunity to use data from national registers in contrast to explicitly collecting data from individuals [7]. The Ethics Committees pertaining to Nordic countries are generally acknowledged to represent the public. In registry-based research, consent from the Ethics Committees can substitute the individual consents from study participants [6].

The principles of equality and justice stipulate that individuals should not be discriminated because of their inability to give consent. In registry based research in Nordic countries, researchers rarely request for an informed consent from the participants [6]. Thus, in practice, the stipulated ethical principles are not always followed. The different arguments [6] for non-conformance to certain ethical principles are presented below:

a) The requirement of informed consent would render many research proposals infeasible due to the large number of study participants. The costs for obtaining the consent of all the participants would be exorbitantly high and would render research involving such participants infeasible.

b) Requesting individual consent would drastically reduce individual participation rates and statistical power especially for nationwide population-based studies. This leads to validity threats in the research study (e.g. loss of generalizability).

c) It is difficult to obtain consent from individuals with limited knowledge of the communicating language (e.g. children and immigrants with limited or no knowledge of Nordic languages). Also, some individuals are dead (in registry based research) and hence their consent cannot be obtained.

4 Data protection regulation in Sweden

The *Data Protection Regulation* is a new regulation by the European Commission that is going to be effective from May 25^{th} 2018. This new regulation is going to replace the Data Protection Act in Sweden (*Personuppgiftslagen*). The regulation deals with similar rules and laws that are governed by the Data Protection Act.

According to the new regulation, the organisation or authorities which deal with personal sensitive data should appoint a data protection officer to monitor the personal data. The data inspection board can penalize the authorities or organisations who violate the laws of the regulation. The Data protection regulation is applicable to all the researchers, organisations, which deal with processing of the sensitive personal data (data collectors and personal data assistants). An important construct of the new regulation concerning the transfer of personal data is as follows:

- When data is obtained from the subjects by informed consent, subjects involved in the research have rights to be get informed about the data transfer.

5 Conclusion

As a computer science researcher in a Nordic country, when my research requires data analysis involving personal data, I will first ensure if explicit data collection from participants is necessary or if the required data is already available. Before making use of the personal data in my research, I would consult the local Ethics Committee. Depending on the size of the dataset, with the guidance of the ethics committee and my peer researchers, I would make a decision regarding the acquisition of consent from the participants.

Researchers should carefully examine and understand various ethical issues related to protecting the personal data of the individuals. This is important as the protection of sensitive data stands as an integral part in protecting the rights of individuals involved in research. Researchers should also focus on various challenges related to conducting research in an ethical way and mitigate these challenges to conduct the research fairly and lawfully.

Appendices

A Definitions

1. **Personal Data:** Data related to an identified or an identifiable person is called as Personal data. An identifiable person is one who can be identified directly or indirectly with the available data. Typically, the person is identified by reference to an identification number (e.g. *Personnummer*) [3].

2. **Anonymization:** Anonymization deals with severing the connection between the data and the individual, so that the data no longer reveals anything about that person [2].

3. **Privacy:** Right of individuals to control the data collection and use of their personal information [5].

4. **Confidential Data:** The data given in confidence or kept as a secret between two individuals [2].

5. **Sensitive Personal Data:** Data pertaining to person's race, ethnic origin, political opinion, religious or similar beliefs, physical or mental health or condition, sexual life, and similar sensitive data pertaining to an individual [2].

6. **Informed Consent:** Research involving methods that affect the research subject physically or psychologically requires the researcher to obtain consent from the participants. They should inform about how the data is shared, stored and managed. They should inform the individuals regarding how their confidentiality will be maintained. The consent should be obtained verbally or by written means. The research participant needs to be aware of:

    - Various methods used for handling personal data in the research.
    - Justification for data collection.
    - Data usage and tenure of the data storage.
    - Concerns related to rightful usage of data.

References

[1] Edward S. Dove and Mark Phillips. "Privacy Law, Data Sharing Policies, and Medical Data: A Comparative Perspective". In: *Medical Data Privacy Handbook*. Cham: Springer International Publishing, 2015, pp. 639–678. ISBN: 978-3-319-23633-9. DOI: 10.1007/978-3-319-23633-9_24.

[2] Veerle Van den Eynden et al. *Managing and Sharing Data; a best practice guide for researchers*. 2011. URL: http://www.data-archive.ac.uk/media/2894/managingsharing.pdf.

[3] Caroline Gans-Combe. "Data Protection and Privacy Ethical Guidelines (Version 5)". In: *European Commission, Sept* 18 (2009), p. 2009. URL: http://ec.europa.eu/research/participants/data/ref/fp7/89827/privacy_en.pdf.

[4] David Hayes and John Devaney. "Accessing social work case files for research purposes: Some issues and problems". In: *Qualitative Social Work* 3.3 (2004), pp. 313–333. URL: http://journals.sagepub.com/doi/pdf/10.1177/1473325004045667.

[5] Sandra C. Henderson and Charles A. Snyder. "Personal information privacy: implications for {MIS} managers". In: *Information & Management* 36.4 (1999), pp. 213 –220. ISSN: 0378-7206. DOI: https://doi.org/10.1016/S0378-7206(99)00019-1. URL: http://www.sciencedirect.com/science/article/pii/S0378720699000191.

[6] Jonas F. Ludvigsson et al. "Ethical aspects of registry-based research in the Nordic countries". In: *Clinical Epidemiology* 7 (2015), pp. 491–508. DOI: 10.2147/CLEP.S90589. URL: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4664438/.

[7] Jørn Olsen. "Register-based research: some methodological considerations". In: *Scandinavian Journal of Social Medicine* 39.3 (2011), pp. 225–229.

[8] David R Wright. "Research ethics and computer science: an unconsummated marriage". In: *Proceedings of the 24th annual ACM international conference on Design of communication*. ACM. 2006, pp. 196–201.