Processing of personal data in student theses

2023

Contents

Introduction

These guidelines aim to explain the most important data processing principles that students should consider when writing their theses.

Data protection is a broad field, and advances in technology are making it increasingly difficult to keep data secure. Therefore, before collecting data, it is important to get acquainted with the legislation in force and think carefully about what purposes and how to process the data. As a general principle, it is better to use as little personal data as possible – exactly as much as is needed for the thesis.

Personal data processing is governed by two main legal instruments – General Data Protection Regulation (GDPR) of the European Union and the Personal Data Protection Act (IKS) applicable in Estonia. If personal data need to be processed to achieve the purpose of the thesis, both legal acts must be complied with. Compliance with personal data processing requirements will primarily protect the individuals whose data are processed (clause 4 of the preamble to GDPR). Therefore, when you choose your thesis topic, it is worth considering if you need to process personal data in the research phase.

More information is available on the university’s IT helpdesk wiki pages, where you will find the cybersecurity and data protection guidelines.

1. Definitions

1.1. Data subject

A natural person whose data are processed.

1.2. Anonymous data

Data that cannot be associated with any single person and that do not enable direct or indirect identification of a person. Anonymous data are not treated as personal data under the GDPR.

1.3. Personal data

Any information about a data subject. The person can be identified directly, for example, by a personal identification number or a name, or indirectly – for example, if you write “Kalmer, a maths teacher from Kambja", it would be possible to identify the person in question if there is only one male maths teacher in Kambja (Article 4 (1) of the GDPR).

When collecting data for your thesis, you should make it very clear to yourself whether it is possible to identify the respondents directly or indirectly.

1.4. Processing of personal data

Any act performed with personal data, including viewing, collecting, recording, organisation, storage, alteration and disclosure, using, forwarding, cross-use, combining, erasing or destruction, granting access to the data, and making enquiries and extracts of them, or several aforementioned actions at a time, independently of the method or the means used.

1.5. Consent

A freely given, specific, informed and unambiguous indication of the data subject's wishes, by which the data subject clearly expresses (for example, by a written statement) agreement to the processing of their personal data (Article 4 (11) of the GDPR)

1.6. Pseudonymised data

Personal data that are processed in such a manner that the personal data can no longer be attributed to a specific data subject without the use of additional information. Data should be pseudonymised, for example, if data that were previously collected about the data subject in a personalised form are no longer needed in the analysis stage of the thesis.

When using pseudonymisation, the data allowing identification of a person are replaced by a code, and, as a result, the person cannot be identified. Pseudonymisation is reversible – if you know the code and the original data, it is possible to establish a link between the data and the data subject. Therefore, the code key must be kept separate from the original data and destroyed at the end of processing.

If the same personal data need to be processed for the purposes of further research or other projects (for example, larger datasets are used in several projects), the data subject's consent must be obtained. When requesting consent, the purpose of using the personal data must be formulated; when the goal is achieved, the processing must stop. There must be a (new) legal basis for processing the same personal data for a new purpose.

2. Principles of processing

The processing of the data used in the thesis must comply with the principles set out in Article 5 of the GDPR.

- The **principle of lawfulness, fairness and transparency** – the author of the thesis must prove that the processing of personal data complies with legislation and is understandable to the data subject, and that data subjects are treated equally in the process. There must be a legal basis for processing personal data (legislation or consent, see also cl. 4).
- The **principle of purpose limitation** – data may only be processed for a specific purpose and only if it is not possible to complete the thesis otherwise.
- The **principle of data minimisation** – as little data as possible should be collected to meet the purpose of the thesis. Data must not be collected 'just in case'. When planning the thesis, the author should write down what data are needed and why.
- The **principle of accuracy** – the collected data must be accurate and up to date.
- The **principle of storage limitation** – no set of personal data should be kept forever. When planning the thesis, you should decide from the beginning how long the personal data will be kept and what the purpose of storing is. The data must not be kept longer than what you promised to the people who gave you the consent. It is recommended that you inform the data subject of the storage period on the consent form. After the purpose is met and the thesis has been completed, the personalised data and consent must be destroyed.
- The **principle of integrity and confidentiality –** unauthorised and unlawful use of personal data is not allowed. Personal data must be kept securely, and their confidentiality guaranteed. When writing a thesis, the author must ensure that third parties cannot access the data, that authorised access is required, and that the data is not accidentally lost, destroyed or damaged. Confidential information must not be shared with friends or relatives. It is inappropriate, for example, to promise confidentiality to the respondents in the introductory part of a survey but later share the most interesting answers with friends or acquaintances. If the thesis contains sensitive information about an individual, access to the thesis must be restricted. Also, the thesis defence must be closed to the public.
- **Accountability** The controller of personal data, or the student writing the thesis, is always responsible for actions with the personal data at their disposal and must be prepared to provide proof of due processing. According to Article 15 of the GDPR, data subjects have

the right to request information from the controller about processing their data. According to Article 12 (3) of the GDPR, the controller has one month to respond to such a request.

3. Planning the thesis

3.1. Collecting and analysing data

When processing the data needed for your thesis, you should prefer the environments provided by the university. If you use Google Forms or a transcription software, the owners of which have not made an agreement with the University of Tartu, it must be clearly indicated on the consent form. If you use a non-university application, please read its data protection policy.

For information on various technological possibilities, contact your institute, college, supervisor, or the university's IT helpdesk.

Please also note that anonymising or pseudonymising data in the analysis stage is preferable. The data subject's name and other information that could identify the respondents should be removed. After the thesis has been defended, all personal data must be destroyed.

3.2. Data storage location during analysis

All students have a user account ending in @ut.ee, which enables them to use the university's applications. It is recommended to use the university's OneDrive to store the personal data collected for the analysis in the thesis. OneDrive user instructions can be found on the university's wiki pages. Also, the consent should be stored in the same environment as the data.

Whether you store the personal in a university-managed environment or elsewhere, you must ensure it is not accessible to third parties. For example, if you keep the data on a home computer, other family members must not see them. It is important to be careful and ensure that the data are not accessible to third parties when using cloud services (for example, when backing up or automatically uploading photos and videos). In addition, the computer must be protected by antivirus software.

If you have collected data by email, via forms or recording, you should consider how to safely forward the data to another application or to third persons (for example, for expert assessment).

When using a flash drive, you must ensure that third persons cannot access the data and that the data are destroyed later.

4. Legal basis for data collecting

One should not collect any data before it is clear on what legal basis the data are being collected. The legal basis can be, for example, a person's consent or a legal obligation.

4.1. Data processing based on consent

Before asking for consent, consider for what purpose and for which personal data it is needed. For example, you should decide whether the personal data are collected for analysis purposes only and the thesis will describe the results anonymously, or the personal data are also disclosed in the thesis. On that basis, you should ask the consent.

Preferably, the consent should be in writing, as according to the GDPR, the data controller has an obligation to prove having obtained the consent. If you plan to record interviews (including audio recordings), you must inform the respondent in advance. A person's voice does not change after a certain age, so the data subject can be identified by voice. If you use personal data for personal purposes only (for example, in the research diary), you do not have to inform people about recording.

The person whose consent is asked must clearly and unambiguously understand what they consent to. Leave enough time for the respondent to read the consent form and explain for what purposes you will use the person's data. The consent form must also include contact information, so that data subjects can request more information if necessary. Each item for which consent is asked must be provided on a separate line so that the person giving consent can clearly indicate what they agree to. Implied consent, which requires a person to take steps to opt out, is prohibited.

If an individual's personal data are disclosed in a thesis, the consent must be included in an annex.

4.1.1. A person has the right to withdraw their consent **at any time** – even after the thesis with personalised data has been defended and made public. The data processing prior to the withdrawal has been carried out based on the consent; no action is required in this case. However, from the moment of submitting the notice of withdrawal, the data processing must stop immediately and the person's data must be erased. If the thesis has already been made public in the university's web environment, public access to the thesis will be closed. In agreement with the author, either the thesis is made public again after removing the personal data, or it remains closed, and no one can access it.

4.1.2. If **children take part** in the study, written consent must be obtained from the parent or legal guardian, even if the author of the thesis is the children's teacher or supervisor. If the parent or legal guardian does not give consent, the child may participate in the study, provided that the child's data are not recorded or used in the study (for example, if interesting games are organised for the study at the kindergarten, the children whose parents do not consent can also take part).

Generally, the parent's or legal guardian's consent is needed for children under 18. Information society services that require the internet for their provision and use (online search engines and sales portals, public sector e-services, e.g. e-services of the Estonian Tax and Customs Board or the Estonian Transport Board, digital signature) are an exception. For those services, processing of the child's personal data under the child's consent is permitted if the child is at least 13 years old (Personal Data Protection Act §8 (1)).

4.1.3. If the **student him- or herself is the subject of research**, the good practice is to inform others whom the study may concern. If the self-analysis in the thesis uses the personal data, opinions, thoughts and assessments of third persons, their consent must be obtained.

4.1.4. **Covert observation** is a method which is generally non-consented research in which respondents do not have the option to decide whether they want to take part in the study. In covert observation, it is necessary to ensure that no personal data are processed in the written paper. Until the data have not been placed in a data set, there is no data processing, and the GDPR requirements do not apply. The processing of personalised data is not allowed in the case of covert observation.

4.2. Data processing under ethics committee resolution

If special categories of personal data (e.g., data concerning health, religious beliefs, political opinions, etc., according to Article 9 (1) of the GDPR) need to be processed for the thesis, the data subject’s consent is insufficient. In this case, the research ethics committee must make the decision (Article 6 (4) of the GDPR). For example, the approval of the ethics committee is mandatory for processing health, biometric or genetic data, the release of personal data from the health information system, or the clinical trial of medicinal products. The ethics committee’s approval is an additional safeguard for the processing of special categories of personal data.

Before the ethics committee’s approval has been obtained, it is not allowed to start collecting the data.

4.3. Compliance with a legal obligation

Sometimes there are legal restrictions on the use of certain personal data. For example, if the data analysis in a thesis is related to traffic fines that are about to expire, personal data relating to these fines must not be disclosed. In order to comply with such restrictions, it is important to familiarise yourself with relevant legislation before starting your thesis.

5. Analysis and results

If personal data have been collected properly and there is a legal basis for processing (for example, consent), further work should follow the principle of data minimisation: only data necessary for the purpose are processed. Personal data should be kept securely; for example, they can be anonymised or pseudonymised for analysis.

5.1. Personalised original data and interviews

Before analysis, the interview responses and transcripts of audio and video recordings containing personal data are preferably pseudonymised (in which case the GDPR applies), or anonymised (in which case the GDPR does not apply). The author of the thesis must assess the volume of data which allows the identification of an individual.

5.2. Data transmission

If data are sent for analysis to a third party, all data that could identify a person must be removed. Also, text must be anonymised (except if consent for transferring such data has been obtained from the data subject). When sending the draft thesis to the supervisor, all data that could identify a person must be removed unless consent has been obtained.

5.3. Annexes to thesis

The student should take care to remove personal data from annexes to the thesis. For example, if an excerpt from the research diary or transcripts are attached, it is necessary to check that they do not contain personalised information (including references to someone recognisable through description). If an annex to the thesis includes a questionnaire, personalised data should also be removed from the questionnaire. Whether or not to disclose the author's own personal data is up to the student to decide.

5.4. List of references

Even if personal data from a public source have been used in the thesis, and the source has been cited accurately, this may not always be the proper behaviour. For example, if you mention someone by name (inmate of Viru Prison, with first name and surname), you must ensure that you cite the correct source (it is not an illegal website) and that the personal data can be used for the purpose of the thesis. To use previously published data in the thesis, you must have a purpose and a legal basis.

According to the university's study regulations, the thesis defence is a public event. If the thesis contains personal data, the student must apply to the institute to have the defence closed for the public.

6. After thesis defence

6.1. If the thesis includes anonymous data, the data must not be published at the thesis defence, either public or closed.

The theses are made public on the university's website (generally on the institute's or college's website), and after the defence, in the university's digital archives in DSpace.

6.2. The theses are published under the **non-exclusive licence** CC-BY-ND-NC for public use. Further information on the licence is available on the institute's or college's website. If the author of the thesis has obtained the data subject's consent to use their personal data, but not to publish it, they must request permission from the institute or college to restrict access for a certain period or request that the thesis is not published at all. Restriction of access can also be requested for a thesis that deals with trade secrets. For that purpose, a non-exclusive licence to reproduce the thesis must be submitted using the form provided. In this case, the thesis is stored until the end of the copyright period, but it is not published.

6.3. The personal data collected and analysed for the thesis must be **destroyed** after the thesis defence from all locations the data were processed. After that, you can destroy the consents based on which the data were processed.

6.4. When writing the **acknowledgements**, note that if you have used anonymous data in the thesis, you should not thank an anonymous source by name.

7. Checklist

Question	Answer
Are you going to use personal data in the thesis?	
What is the purpose of using personal data?	
What is the legal basis for the processing of personal data?	For example, consent
Who is being researched?	
What personal data are collected?	
What the personal data are used for?	
Is it necessary to apply for the Ethics Committee's approval for the research?	For example, when special categories of personal data are used.
Which personal data are processed in the stage of collecting or receiving data?	For example, are the data personalised or pseudonymised?
Which software do you plan to use for data collection?	
Which personal data are processed in the data analysis stage?	For example, personalised data, pseudonymised data
Which software do you plan to use for data analysis?	
Who has access to the personal data of this study?	For example, who processes them for the purpose of the study?
Do you plan to transfer the data used in the study to a foreign country?	For example, to get an expert assessment
In what form are the personal data used in the thesis? ☐ personalised data ☐ pseudonymised data ☐ anonymous data ☐ aggregated statistics	

Are paper documents containing personal data generated in the course of the work?	
Where are the personal data stored?	
How are personal data stored?	
How long are the personal data stored?	
What will be done with the research material after the end of the research?	For example, destroyed, returned to the provider of data, safely stored with the researcher

Further information:

Terje Mäesalu

Senior Specialist of Data Protection

andmekaitse@ut.ee