Personal Data -
Definition, Storage and Reduction

The processing of personal data ((Art. 4 para. 1 DSGVO) is necessary for scientificpurposes in many research projects. This is usually done by observing individuals or when information is requested through surveys. This fact sheet provides an overview of the most important legal aspects of handling personal data, including definitions, an excursion on informed consent as well as listing regulations for storing and archiving such data. Furthermore, this fact sheet shows how the risk of personal identification can be reduced through pseudonymization and anonymization. This is explicitly not a legally binding document, but only a recommendation. If you have detailed questions, please contact your institution‘s legal department or its data protection officer.

What are personal data?

The legal basis for the handling of personal data is the General Data Protection regulation (GDPR) of the European Union, which is supplemented and specified by the German federal Data Protection Act (BDSG)) and the data protection laws of the federal states, such as the Thüringer Datenschutzgesetze (ThürDSG). Personal data is protected according to Art. 4 (1) of the GDPR. It states that personal data is "all information relating to an identified or identifiable natural person. A natural person is considered to be identifiable if he or she can be identified directly or indirectly by means of association through names, location data or the like. This includes characteristics "which reveal the physical, physiological, genetic, mental, economic, cultural or social identity of the natural person". Art. 9 of the GDPR additionally defines special categories of personal data, which are often colloquially referred to as sensitive personal data. Higher technical and legal requirements are imposed on the processing of such data.

Examples of affected data according to the GDPR:

Personal data

- name
- age
- educational career
- address
- nline identification
- phone number
- employer and workplace
- activities, hobbies and behavior patterns
- photos

Sensitive personal data

- racial and ethnic origin
- political opinions
- religious or ideological convictions
- membership in trade unions, associations or organizations
- genetic/biometric data
- health data
- sexual preferences or sexual orientation
- criminal convictions and offences[1]

[1] according to Art. 10 DSGVO processing only under official supervision

Basis of the storage and the declaration of consent

According to Art. 6 GDPR, personal data may be collected and processed if, among other things, the data subject has consented to the data processing, the data processing is necessary for the performance of a contract or due to legal obligations, or if there are legitimate research interests. However, the latter is always a matter of assessment associated with legal uncertainty, where the research interests must outweigh the fundamental rights and freedoms of the data subject.

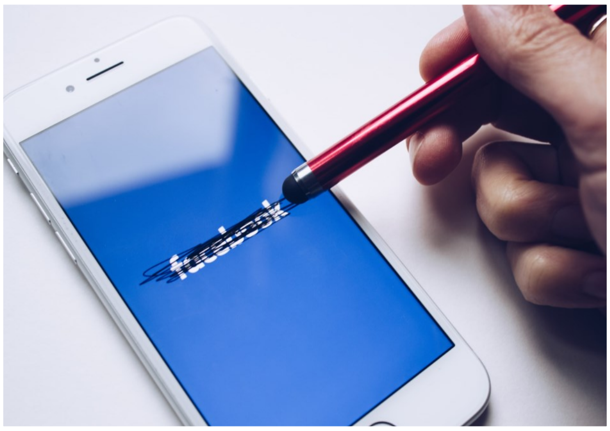
In order to avoid legal problems and to ensure the transparency of the research project, it is recommended to obtain a declaration of informed consent from the person concerned for the collection, storage and, if necessary, publication of the data. This consent is understood to be a voluntary, informed and unequivocal expression of will (Art. 4 (11) GDPR). Although consent can be given verbally, it is recommended to obtain it in writing to avoid future issues. According to Art. 13 GDPR, the declaration of consent should contain the following information:

- names of contact persons of the project (if applicable the data protection officer)
- purpose of the processing of the data or the legal basis thereof
- information about the recipients of the data (persons, organizations, countries, ...)
- information about (or criteria for) this, how long the data is stored
- information as to which data of non-data subjects are processed and from which sources they originate (Art. 14)
- Information on data subject rights such as: right to access (Art. 15), rectification (Art. 16), erasure (Art. 17), restriction of processing (Art. 18), data transmission (Art. 20), opposition (Art. 21), prohibition of automated decisions and profiling (Art. 22), appeal and withdrawal of consent

Not all points need to be recorded if they either do not apply to the research project or would negatively influence the collection of data (e.g. through bias of the participants when answering questions). Self-formulated forms or pre-existing templates may be used, but should be reviewed by the local data protection officer before use.

The obligation to inform data subjects about the collection of their personal data may be waived if, for example, this proves impossible or requires disproportionate effort (Art. 14 (5) (b) GDPR and §27 BDSG). In this case, however, technical and organizational measures must be taken to minimize the risk of data being compromised (Art. 89 (1) GDPR).

Personal Data - Definition, Storage and Data Economy



Since personal data are subject to a purpose limitation and storage restriction (Art. 5 (1) GDPR), they may only be stored for as long as necessary to fulfil the (research) purpose. As soon as the purpose is fulfilled via the end of the research project or if a data subject revokes his or her original consent to data processing, the personal data must be deleted in accordance with Art. 17 GDPR. However, in order to comply with good scientific practice, archiving of personal data is still permissible for this specific purpose for up to 10 years as long as appropriate security measures are taken[1].

Personal data must be adequately secured by organizational and technical measures both during the project duration and, if necessary, after the end of the project, if the data continue to be stored. This usually includes storing the data in a secure and lockable room. It is also recommended to encrypt the data and to use secure passwords. Ideally, the password should be randomly generated, documented and changed every few years. To avoid losing access to the data due to a chain of unfortunate circumstances, two or more people should always have access to the data. It is recommended that a security officer be appointed to organize and manage access to the data both during and after the project.

Data Economy: Pseudonymization and Anonymization

Personal data should be minimized at the earliest possible stage of the research project after collection, as far as this is possible and does not restrict analysis. In this way, they can be handled more openly and are also subject to fewer restrictions when transferred to an archive or published. The two main methods of data minimization are pseudonymization and anonymization.

Pseudonymization

According to Art. 4 (5) of the GDPR, "pseudonymization is the processing of personal data in such a way that the personal data can no longer be attributed to a specific data subject without the need for additional information. This additional information must be kept separately. In addition, technical and organizational measures are necessary to ensure that the personal data can not be assigned to an identified or identifiable natural person".

In other words, pseudonymization is the replacement of names or other identification features with a code that only allows conclusions about the actual person with the help of secure information. To what extent the pseudonym itself must be protected depends on the effort required to identify the person. These can be objective factors, such as required costs, time, technology and development.

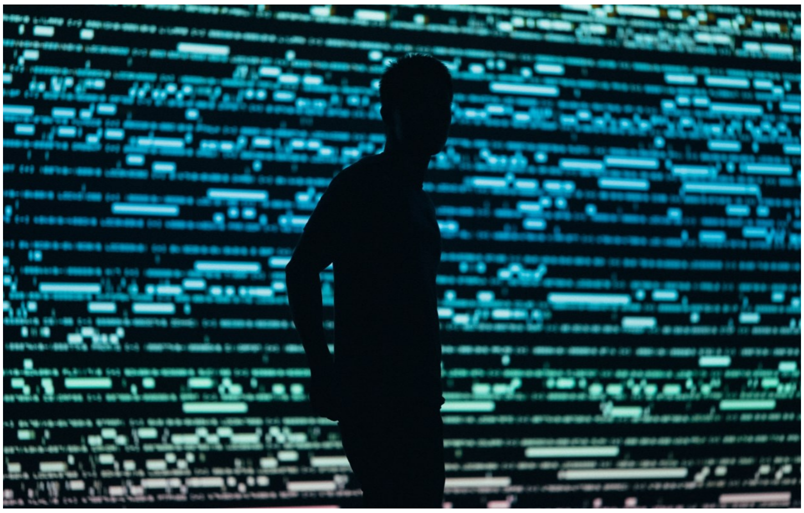
> **Example:** If an exam is being evaluated, participants can give an imaginary name (pseudonym). This name will be used in connection with the individual result when it is published, so that only the examiner and the participating person know the connection.

[1] Wirth, Thomas (2020), Die Pflicht zur Löschung von Forschungsdaten - Urheber- und Datenschutzrecht im Widerspruch zu den Erfordernissen guter wissenschaftlicher Praxis?, Zeitschrift für Urheber- und Medienrecht (ZUM) 64(8/9), 585-592

Anonymization

Anonymized data does not contain any personal reference, which is why, according to recital 26 of the GDPR, these data are no longer subject to the GDPR and no definition of these data is formulated there. In general, personal data is considered to be anonymized if persons can no longer be identified via the data and no conclusions can be drawn about the data to be protected (as in the case of pseudonymization). The process of anonymization is therefore irreversible.

In practice, personal data is made anonymous by deleting the unique identifier (such as the name of the person or another key value) and by not storing certain values in the form of a coherent data entry when this is no longer necessary for analysis (such as place, time and activity, which when taken together can enable identification). Another approach is the aggregation of data. When aggregating data, sets are grouped in such a way that they no longer allow any conclusions to be drawn about the individual person. This can be done by rounding or generalizing values, e.g. by grouping different ages into age groups. Another method would be masking, in which the original values are filled with random (but technically valid) new values.



> **Example:** Secret elections are based on the principle of anonymization, i.e. the results are collected and evaluated, but it is not possible to draw conclusions about the voting behavior of individuals. In databases, anonymization is generally done by deleting the connecting key between the tables (e.g. a customer number). The remaining data is then incoherent and can no longer be assigned to a specific person.

Which measures with regard to personal data are appropriate or even necessary should be determined in cooperation with the local data protection officer on a case-by-case basis..

If all references to an individual person have been removed from the research data, thus making it anonymous, the data and the resulting results can usually be published without any further restrictions, provided that no interests of third parties oppose this (e.g. licenses or in the case of commissioned research).

This fact sheet offers an insight into the handling of personal research data. If you have further questions on this topic, you can contact your legal department, the local data protection officer or our network at: https://forschungsdaten-thueringen.de/kontakt.html kontaktieren.