Guidance on how to process personal data for research

1. What is personal data?
Personal data is defined as information that can directly or indirectly identify a natural person. It can be a name, registration number of a car, or information about the course of an illness. It can also be an experience, interaction or something similar that can be attributed to a particular person. This means that pieces of information that can be linked to a particular person through context will also be personal data, although the individual information itself may not be personal data.

As such, a distinction is made between direct and indirect identification. Direct identification includes things like name or social security number. Indirect identification can be information regarding age, gender, education, and other information that may not directly identify a person but can nonetheless be used in conjunction with other information for identification.

There is no limit to what kind of information may be considered personal data, which means that it must be estimated on a case by case basis. For example, if you need to do a survey, each answer may well be anonymous, but in some cases the linking of answers will allow you to identify a particular person.

There may be questions about class levels, school, gender, hobbies, and whether their parents are divorced. The individual answers will not be personal data, but when the results of the individual form are compiled as a whole, the participant will be able to be identified in some cases.

Although there may be more students who have answered the same question and as such, for example, there may be two in particular that may be able to be identified, it will count as personal data since it is such a narrow group that it removes some doubt as to who it is. As such, you have to look at the information in a larger context, and not look at each question in isolation.

Direct identifiers are those that can immediately identify the person without further information. This can be, for example, a person's CPR number, where that one piece of data alone can identify the person.

Strong indirect identifiers can include email addresses or phone numbers, which do not immediately allow you to identify the person, but with further investigation they can be associated with the person. In this case, more than one piece of data isn't needed to find the person if you just search for the person who it relates to.

Indirect identifiers, however, are those that cannot stand alone if they are to be associated with a person. ZIP codes, for example, are not personal data per se, but may be if they are linked to other information and as such, in context, can allow a person to be identified.

2. What is meant by *processing personal data*?
All use of personal data is ‘processing’ in the legal sense. This means that any activity or series of activities involving the use of personal data is relevant.

It may be the collection, registration, organisation, systematisation, storage, adaptation or modification, retrieval, search, use, disclosure, dissemination or any other form of transfer, compilation or interconnection, restriction, deletion or destruction. Also, even looking at the data is covered by the concept of ‘processing’.

From this it is clear that the meaning of ‘processing’ is taken in the broadest sense.

2.1 Data Processor vs Data Controller
It is crucial to have data responsibility clarified when processing personal data. The Data Controller is the organisation, person, or company that has overall responsibility for the processing of data. It is the Data Controller who initiates the processing of the data and decides for which purpose and with which tools it should be processed. The individual researcher is not a Data Controller in the above sense, it is SDU.

The Data Processor is an external party (compared to the Data Controller) that processes the personal data on behalf of the Data Controller. As such, the Data Processor acts on the instruction, and also at the request, of the Data Controller. The Data Processor is responsible for storing and processing data in accordance with the agreement that establishes the framework for data processing. The agreement is called a Data Processing Agreement, and must be entered into by both the Data Controller and Data Processor. It is important to have the roles in place prior to starting processing.

The lawyers at RIO are responsible for negotiating the Data Processing Agreement. We have a standard agreement that all agreements are based on. It is also at RIO that all data processing agreements must be registered and signed. You should send an email to sdu.persondata@sdu.dk if you need assistance with a Data Processing Agreement.

3. Anonymisation vs. pseudonymisation
It is important to make clear whether data is anonymous or not. If data is sufficiently anonymised, the rules of the General Data Protection Regulation (GDPR) do not apply, and therefore data can be processed within a much more flexible framework. However, it is very difficult to anonymise information so that it is no longer personally identifiable. Anonymisation must also be irreversible.

#### 3.1 Pseudonymisation

Pseudonymous data is also personal data. Pseudonymisation is the process of making information not directly personally identifiable, but it also does not ensure total anonymity. The typical situation is one where you replace the name with another name or a CPR number with a code. If there is a key that can return data to direct personally identifiable information, your data is not anonymous, but pseudonymous. When the key file is deleted, the pseudonymous data becomes anonymous.

The important thing to note when working with pseudonymous data is if there is a key to decoding information. If it is possible to use the key and data together, it is possible to return data to a personally identifiable form. Note, however, that a data set is not necessarily pseudonymised even if you change the names and CPR numbers with a code. For example, if you have a questionnaire where job title is a free-text field, then the job title “Prime Minister” or “President of the Supreme Court” will mean that data is not pseudonymised.

Pseudonymisation is therefore the term for the replacement of the directly identifiable information with pseudonyms or codes. The codes must be stored separately and must also be protected with sufficient technical measures so that it is not possible for a third party to combine the codes and data, and therefore identify the persons within the dataset. As long as the key to pseudonymisation exists, it is pseudonymous data. It is only when the key file is deleted that the dataset becomes anonymous.

Where it is possible to pseudonymise the personal data, it should be done as soon as possible after the data has been collected. As mentioned previously, pseudonymisation does not mean that the data protection rules do not apply, but it is a very effective security measure, and should the information fall into the hands of unauthorised persons, the damage will not be as great as if direct personal data is accessed by the unauthorised person.

*Deidentified data*Statistics Denmark uses another concept: deidentified data. They define it as data where all formal identification such as name, CPR number, CVR number, and address are removed.This is not to be confused with pseudonymous data, since pseudonymity assumes that the data set is anonymous without the existence of a code. The deidentified data, on the other hand, is what we call indirectly identifiable data, so it is not possible to identify the participants from the data set immediately.
The deidentified data is thus still personally identifiable, although the participants cannot be found without connection to other information or further investigation. As long as it is possible to identify the participant in some way through a reasonable effort, it is personal data. As such, the deidentified data is personal data in the general sense and is neither pseudonymous nor anonymous.

#### 3.2 Anonymisation

Anonymous information is not personally identifiable and it is not possible to identify a person based on the information within the dataset. There are several tools available to anonymise data and ensure that data is being anonymised. If the information, with reasonable effort, can be used to identify a natural person, it is not anonymous information. What counts as ‘reasonable effort’ is somewhat more difficult to explain. It is an assessment of the technical possibilities available to identify the person. It involves taking account of all possible tools that can reasonably be used in trying to identify a person.

When it comes to other possibilities for identifiable information to be given in open question and the risk of the information becoming available, this risk must be assessed in each case based on the topic of the research project, and the underlying circumstances.

It would probably be difficult to make data completely anonymous, but data is considered anonymous if it is not possible, with reasonable effort, to identify a given person. When assessing whether it is possible to identify a person, all tools that could reasonably be used to identify that person should be taken into consideration. It must also be kept in mind whether others have information that makes it possible to identify the individual.

4. Choice of legal basisWhen dealing with personal data for research purposes, there are several choices in relation to which legal basis applies to your work. First, you need to make it clear whether personal data is involved. Next, you should assess which legal basis best suits your project in order to give your participants the best access possible to legal rights. There are differences between the rights and requirements in relation to the various legal bases.

Regardless of which basis you choose, the obligation to give the participants the information as stated in GDPR art. 13 or 14, must be fulfilled so that the participants know what their personal information is going to be used for. If you use SDU templates, you are sure that all the requirements are met.

In general, keep in mind that you must not process more information than is relevant to your project. Therefore, you should minimise the information that you are processing as much as possible, so that you only ask for the information that is needed. Other than that, you can minimise what you have to process by deleting/anonymising unnecessary information if you no longer need it. Of course, this is only an option in relation to the information you do not need in relation to the verification of results.

If you have to collect, use, etc., personal data for a research project and you know that the information will not be used later for anything other than research, it will often be an opportunity to use § 10 of the Data Projection Act and/or Article 6 (1)(e) of the GDPR. This avoids the risks that may be associated with using consent as a legal basis, which is discussed next.

The other option is consent in the classical sense, which is the consent to participate in the study and for SDU to process the participant’s personal data in accordance with the rules of the GDPR.

If you choose to use consent as a processing basis, participants have the right to withdraw their consent and request the deletion of their personal data. However, it might affect your project negatively if one of more participants withdraw their consent. If you can anonymise the personal data, they will no longer be subject to the data protection rules and therefore the participant would not be entitled to request the deletion of their anonymised data (after all, it is not possible to find the participants information as they are anonymised).

If you choose § 10/art. 6, there is no consent to withdraw and therefore you do not risk having to delete information for that reason.

We have created a template that allows participants to ‘consent’ to participate, but the consent does not cover the actual processing of information as the legal basis for the processing is provided for in § 10/art. 6. Participants must therefore ‘consent’ to participate, and may withdraw consent, which will result in no new information being collected, but the researcher may still process information collected before that ‘consent’ was withdrawn. As such, it is possible to continue processing the existing information, as collection, use, analysis, disclosure, etc., is based on § 10/art. 6.

It is important that you choose the correct legal basis before you start collecting personal data. Please contact SDU RIO on e-mail sdu.persondata@sdu.dk if you have any questions.

5. Anonymisation methodsThere are some techniques to anonymise data.First, you can think ahead before you collect your data to structure your data collection in a way that only the relevant information is collected, and in a form that is as least personally identifiable as possible. It can be, for example, through the use of close-ended questions where the participant is unable to add their own information. That way you won't get more information than necessary. You should therefore be careful about inserting questions where it is possible to write free text, but rather delineate the possible answers with pre-determined possibilities.

For example, it is possible to change the municipality of origin by dividing the answers into categories such as urban, semi-urban, rural, etc. This can minimise the risk of identification, but is not enough by itself. It must form part of a larger anonymisation process.

You can also add ‘noise’, so that extra information or details are added. This can be done, for example by changing the age of the participant by +- 2 years. Then, to some extent, data will be accurate but not immediately available to unauthorised people, if the addition of noise is part of a broader anonymisation process.

It may also be a technique to remove or change information, for example that people with AIDS are categorised as people with a severe, long-term illness, and other people are simply categorised as healthy or ill. In this way, the different diseases are not distinguished, but the participants are divided into groups that remain relevant to the research project.

You can also categorise the other information into groups, so that few personal details can be located. For example, this can be done by dividing participants into categories, e.g. in relation to age, place of study, place of work, place or residence, or household. These can be instead categorised as 41-45, long-term education, shop, town in Jutland, and wife and two children. In this way, the information is so general that it is difficult to identify the person, as the description can fit several different people.

Below is an overview of the type of information and what can be done to anonymise it. The **special categories** are marked in bold.

*Concrete examples*
In March 2019, the Danish Data Protection Agency issued serious criticism regarding a taxi company's processing of personal data. The company had ‘anonymised’ customer information by deleting customer names but not their phone numbers. As such, it remains possible to identify the individual customer based on information about the date of the trip, the start and end time of the drive, the number of kilometres driven, payment, and GPS coordinates. This means that the information alone is pseudonymous, and therefore still personally identifiable.

6. Notification of research project
Research projects must be reported to SDU's record of processing if personal data is processed within the project. This notification must be done via the university’s list of processing activities, which is managed by SDU RIO. SDU RIO will then be able to present a list of ongoing projects at the university where personal data is being processed upon the request of the Data Protection Agency.

In order to notify SDU RIO of the project, a notification form must be submitted. You can do this via SDU's registration page. You must answer several questions that relate to your project, such as whether you will process general or special categories of information, how long you expect your project to last, etc.

When you have completed the form, press submit. SDU RIO will then review the notified material in order to ensure that the data protection requirements are met. An opinion on these issues is then sent to the notifier.

7. Approval from other authorities
Some projects require further approval other than the one obtained from RIO. Health sciences projects involving biological material from humans must be approved by the National Science Ethics Committee. If you wish to obtain information in patient records without the consent of patients, this requires approval from the Danish Patient Safety Authority.

In addition, there may be cases where you wish to receive information from other authorities, companies, registries, etc., and here you need the issuing party’s approval or acceptance before you can receive the information.

Most have a disclosure system where you can apply to receive information. Others may want you to prepare a disclosure statement that you can send to SDU RIO for review. When you receive the information from the counter-party, SDU becomes the Data Controller for the information and it must be notified at the record of processing

8. Disclosure of research data
The lawyers at SDU RIO must be involved when research data is to be transferred or disclosed to other parties in order to ensure that the data protection rules are complied with.

Disclosure involves situations where you disclose personal data to external parties outside of SDU. It includes therefore persons, companies, municipalities, or others who are not affiliated with or employed by SDU.

Disclosure can also be made to parties outside of the EU – then it will be a transfer to third countries.

Making personal data available involves situations where you leave data to another SDU colleague, and the data thus remains internal to SDU. Typically, it will be a researcher who needs to use data for a new project that has not previously had access to the data.

You can find disclosure forms on SDU's website.

9. Do you have any questions?
Questions about the processing of personal data can be sent to sdu.persondata@sdu.dk.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Types of identification** | **Direct identifiers** | **Strong indirect identifiers** | **Indirect identifiers** | **Anonymisation method** |
| CPR no. | X |  |  | Deleting |
| Full name | X |  |  | Deleting/changing |
| Email address | X | X |  | Deleting |
| Telephone no.  |  | X |  | Deleting |
| Postcode |  |  | X | Deleting/categorisation |
| Province |  |  | X | Categorisation |
| Municipality |  |  | X | Categorisation |
| Region  |  |  | X |  |
| Audio file (of a voice) | X |  |   | Deleting/distorting |
| Video (of a person) | X |  |  | Deleting/distorting |
| Photo (of a person) | X |  |  | Deleting/distorting |
| Year of birth |  |  | X | Categorisation |
| Age |  |  | X | Categorisation |
| Gender |  |  | X |  |
| Marital status |  |  | X |  |
| Household composition |  |  | X | Categorisation |
| Profession |  |  | X | Categorisation |
| Training |  |  | X | Categorisation |
| Work situation |  | (X) | X |  |
| Native language |  |  | X | Categorisation |
| Nationality |  |  | X | Categorisation |
| Workplace |  | (X) | X | Categorisation |
| Car registration number |  | X |  | Deleting |
| Website |  | (X) | X | Deleting |
| Student no.  |  | X |  | Deleting |
| Account no.  |  | X |  | Deleting |
| Health information |  | (X) |  | Categorisation/deleting |
| IP address |  | X |  | Deleting |
| Ethnicity |  | (X) |  | Categorisation/deleting |
| *Criminal offences* |  |  | X | Categorisation/deleting |
| Political associations |  |  | X | Categorisation |
| Religious beliefs |  |  | X | Categorisation |
| Social condition |  |  | X | Categorisation/deleting |
| Sexual orientation |  |  | X | Deleting |
| Biometric data |  | X |  | Deleting |