Data Privacy Handbook

Utrecht University

19 december 2022

2

Contents

Storage, Sharing, Publication 59

Resources 81

Intro

Data Privacy Handbook

RDM Support

Last updated: 19 December 2022

The Data Privacy Handbook is a guide on handling personal data in scientific research, in line with European data protection and privacy regulations. It consists of:

- A **knowledge base** which explains how the EU General Data Protection Regulation (GDPR, Dutch: Algemene Verordening Gegevensbescherming) applies to scientific research, including guidelines and good practices in carrying out GDPR-compliant scientific research;
- An overview of privacy-enhancing **techniques & tools** and practical guidance on their implementation;
- **Use Cases** in the form of research projects with privacy-related issues, for which a reusable solution (e.g., tool, workflow) has been developed.

The Data Privacy Handbook synthesises information across various sources and presents it a *practical* format. This includes workflows, tools, and practical translations of the GDPR, which could be used by researchers and (data) support staff within Utrecht University and beyond.

The Data Privacy Handbook is an initiative of Research Data Management Support at the Utrecht University Library, in collaboration with privacy and data experts at Utrecht University. It is part of a larger project, the Data Privacy Project, which aims to develop knowledge, tools, and experience on how researchers can and should deal with personal data. This project is funded by the Utrecht University Research IT Program and an NWO Digital Competence Center grant.

This is an Utrecht University (UU) community-driven, open-source project. We welcome feedback and contributions of any type, please read our contributing guidelines for more information.

0.1 How to use this Handbook

The Data Privacy Handbook aims to make knowledge and solutions on handling personal data *Findable, Accessible, Interoperable, and Reusable* (FAIR) and present them in a practical format.

The Handbook need not be read like a textbook. You are invited to navigate to the topic you need based on the table of contents, or use the guide below.

0.1.1 What are you looking for?

I want to...:

Learn about the GDPR in the context of scientific research

Introduction to the GDPR

Definitions

Plan a GDPR-compliant research project

Designing your research project

Informing participants

Obtaining consent

Collaborating on personal data

Setting up agreements

Planning documents, such as the Privacy scan, Data Protection Impact Assessment, or Data classification

Work safely with personal data

Storing personal data

Using GDPR-compliant tools and services

Reducing the sensitivity of personal data

Sharing personal data during research

Share personal data with others

Sharing data legally

Sharing personal data during research

Reducing the sensitivity of personal data

Using GDPR-compliant tools and services

Sharing personal data for reuse

Sharing personal data case by case

Learn from other projects

Publishing metadata only

Pseudonymising different types of data

Minimising personal data processing in a survey

Get help or information

Getting help at Utrecht University

Definitions

References

0.2 License and Citation

The Data Privacy Handbook is licensed under a Creative Commons Attribution 4.0 International License. You can view the license here.

0.3 Disclaimer

The content presented in the Data Privacy Handbook has been carefully curated by Research Data Management Support, in collaboration with privacy officers and data experts of Utrecht University.

The Data Privacy Handbook is a 'living' book that is continually being written, updated and reviewed. Its contents can therefore change, or become outdated or redundant. Hence, the information presented is provided "as is", **without guarantees of accuracy or completeness**.

As scientific research may differ depending on the discipline, topic, and context, measures needed or taken to ensure GDPR-compliance will vary across research projects. The authors can therefore **not be held responsible, nor accountable** for any negative consequences arising from interpretation and use of the content of the Data Privacy Handbook.

The Handbook is not endorsed by the Board of Utrecht University and does not constitute a mandatory directive. For the most up-to-date and official/ authoritative information, please refer to the university website and intranet, to which this Handbook is a hands-on, practical supplement. Moreover, before

implementing the guidance laid out in this Handbook, always seek the advice of your privacy officer or RDM Support to confirm the suitability of any proposed solution to your project.

Throughout the Data Privacy Handbook, links to external webpages may be provided for additional information or assistance. The authors of the Data Privacy Handbook are **not responsible for the content of any such linked webpages**, nor is the content of external webpages necessarily endorsed by Utrecht University.

Utrecht University is committed to sharing knowledge in line with the principles of open science and therefore welcomes readers from outside of the organization. However, the contents of the Data Privacy Handbook may not be in line with readers' institutions' policies or views. For more authoritative information, these readers should refer to resources from their own institutions.

0.4 Contributions

The Data Privacy Handbook is a collaborative effort, made possible by a large number of contributors (also to be viewed in our GitHub repository):

Neha Moopen, Dorien Huijser, Jacques Flores, Mercedes Beltrán, Wies Cipido, Ruud Dielen, David Gecks, Joris de Graaf, Judith de Haan, Saskia van den Hout, Frans Huigen, Artan Jacquet, Rik Janssen, Sanne Kleerebezem, Annemiek van der Kuil, Danny de Koning-van Nieuwamerongen, Pieter Sebastiaan de Lange, Frans de Liagre Böhl, Francisco Romero Pastrana, Najoua Ryane, Johanneke Siljee, Maarten Schermer, Ron Scholten, Garrett Speed, Robert Steeman, Jacqueline Tenkink-de Jong, Liliana Vargas Meleza, Martine de Vos, and others.

Would you like to contribute to this Handbook yourself? Please read our Contributing Guidelines.

Chapter 1

Privacy FAQs

On this page you can find some often-heard questions about handling personal data in research. Click a question you have to read its answer.

This is a first FAQ. More will follow!

Am I a processor as employee of my university?

No. As an employee you are still determining your own why (research question) and how (methods) of personal data processing. This makes you a controller, acting as an "agent" of the legal controller (your university). Read more on the difference between processors and controllers on the definitions page.

Knowledge Base

The GDPR

This chapter will present the most important definitions, principles and rights of data subjects outlined in the GDPR and how it applies to your research. Most of the practical advice that we provide in this Handbook will be rooted in and builds on the concepts presented here.

1.0.1 Chapter summary

The GDPR is a EU-wide regulation that controls the processing of personal data. If you process personal data, you should:

- Make sure you have a legal basis to process the data. In research, this is often informed consent.
- Be transparent and fair towards data subjects.
- Be specific in which personal data you process and for what purposes. Limit the amount of data you process to what is necessary, and only store the data for that necessary amount of time.
- Protect the confidentiality of the data by incorporating privacy by design into your project from the start.
- Make sure your data subjects can exercise their data subjects' rights, and they know how to do so.

1.1 What is the GDPR?

The General Data Protection Regulation (GDPR, Dutch: *Algemene Verordening Gegevensbescherming* [AVG]) is an EU-wide regulation meant to protect the privacy of individuals within a rapidly growing technological society. The GDPR facilitates the free movement of personal data within the European Economic Area (EEA). Its data processing principles are meant to ensure a fair balance between competing interests – for example, the right to conduct research vs. the right to protect personal data (Articles 13 and 8, from the Charter of Fundamental right of the EU).

1.1.0.1   The GDPR in a nutshell

All articles and recitals of the GDPR can be found online via https://gdpr-info.eu/. The video below highlights some important aspects of the GDPR:

Click to read the English video transcript

The General Data Protection Regulation (GDPR) regulates what we can and cannot do with personal data such as a person's name, sexual orientation, home address and health. This also applies to personal data used in research and education. The regulation consists of 88 pages. Fortunately, the basics are easy to remember in 3 steps:

First, there must be a clear legal basis for processing personal data. This can include consent, a legal obligation, or public interest.

Second, appropriate technical and organisational measures must be taken while processing personal data to ensure maximum privacy.

Lastly, the persons whose data you have collected must always have the option of inspecting, changing, or removing their personal data.

That is the GDPR in a nutshell.

1.1.0.2   When does the GDPR apply?

The GDPR has been applicable from May 2018 onward and applies when:

- you are processing personal data (material scope, art. 2).
- the controller or processor of the data *resides* in the EEA (territorial scope, art. 3). This is independent of whether the actual processing takes place in the EEA. In some cases, the GDPR also applies when the controller or processor is not established in the EEA, but is processing data from EU citizens.

If you are collecting or using data that originated from individuals (or is related to individuals), it is very likely that the GDPR applies to your project. You can read more in the chapter What are personal data?.

1.1.0.3   Implementation

While the GDPR is a regulation for the entire EEA, each EEA country can additionally implement further restrictions and guidelines in national implementation laws. The Dutch implementation law is called "Uitvoeringswet AVG (UAVG)" (most recent version). The UAVG determines, for example, that it is forbidden to process Citizen Service Numbers (BSN), unless it is for purposes determined by a law or a General Administrative Order (AMvB).

1.2 Definitions in the GDPR

Below, you will find a selection of important terms in the GDPR that you should become familiar with when working with personal data (also included in the Definitions). Click a term to see the definition.

Data subject

A living individual who can be identified directly or indirectly through personal data. In a research setting, this would be the individual whose personal data is being processed (see below for the definition of processing).

Personal data

Any information related to an identified or identifiable (living) natural person. This can include identifiers (name, identification number, location data, online identifier or a combination of identifiers) or factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of the person. Moreover, IP addresses, opinions, tweets, answers to questionnaires, etc. may also be personal data, either by itself or through a combination of one another.

Of note: as soon as you collect data related to a person that is identifiable, you are processing personal data. Additionally, pseudonymised data is still considered personal data. Read more in What are personal data?.

Special categories of personal data

Any information pertaining to the data subject which reveals any of the below categories:

racial or ethnic origin

political opinions

religious or philosophical beliefs

trade union membership

genetic and biometric data when meant to uniquely identify someone

physical or mental health conditions

an individual's sex life or sexual orientation

The processing of these categories of data is prohibited, unless one of the exceptions of Article 9 applies. For example, an exception applies when:

the data subject has provided explicit consent to process these data for a specific purpose,

the data subject has made the data publicly available themselves,

processing is necessary for scientific research purposes.

Contact your privacy officer if you wish to process special categories of personal data.

Processing

Any operation performed on personal data. This includes collection, storage, organisation, alteration, analysis, transcription, sharing, publishing, deletion, etc.

Controller

The natural or legal entity that, alone or with others, determines or has an influence on why and how personal data are processed. On an organisational level, Utrecht University (UU) is the controller of personal data collected by UU researchers and will be held responsible in case of GDPR infringement. On a practical level, however, researchers (e.g., Principal Investigators) often determine why and how data are processed, and are thus fulfilling the role of controller themselves.

Processor

A natural or legal entity that processes personal data on behalf of the controller. For example, when using a cloud transcription service, you often need to send personal data (e.g., an audio recording) to the transcription service for the purpose of your research, which is then fulfilling the role of processor. When using such a third party, you must have a data processing agreement in place.

Legal basis

Any processing of personal data should have a valid legal basis. Without it, you are now allowed to process personal data at all. The GDPR provides 6 legal bases which are explained further in this chapter.

Anonymous data

Any data where an individual is irreversibly de-identified, both directly (e.g., through names and email addresses) and indirectly. The latter means that you cannot identify someone:

by combining variables or datasets (e.g., a combination of date of birth, gender and birthplace, or the combination of a dataset with its name-number key)

via inference, i.e., when you can deduce who the data are about (e.g., when profession is Dutch prime minister, it is clear who the data is about)

by singling out a single subject, such as through unique data points e.g., someone who is 210 cm tall is relatively easy to identify)

Anonymous data are no longer personal data and thus not subject to GDPR compliance. In practice, anonymous data may be difficult to attain and care must be given that the data legitimately cannot be traced to an individual in any way. The document Opinion 05/2014 on Anonymisation Techniques explains the criteria that must be met for data to be considered anonymous.

Pseudonymous data

Personal data that cannot lead to identification without additional information, such as a key file linking pseudonyms to names. This additional information should be kept separately and securely and makes for de-identification that is reversible. Data are sometimes pseudonymised by replacing direct identifiers (e.g., names) with a participant code (e.g., number). However, this may not always suffice, as sometimes it is still possible to identify participants indirectly (e.g., through linkage, inference or singling out). Importantly, pseudonymous data are still personal data and therefore must be handled in accordance with the GDPR.

1.3   Principles in the GDPR

The GDPR has a number of principles at its core which dictate the (method of) data processing. Every type of processing has to comply with these principles. Understanding these principles is the first step to determining what type of personal data can be collected and how they can processed.

The GDPR principles are explained further below the image. The next chapter will describe how to implement these principles in your research. You can also always contact your privacy officer for help.

1.3.0.1   1. Lawful, fair and transparent

When working with personal data, your processing should be:

Lawful

Make sure all your processing activities (e.g., data collection, storage, analysis, sharing) have a legal basis. Ideally, you should have determined your processing purposes (e.g., research questions) in advance.

Fair

Consider the broad effects of your processing on the rights and dignity of the data subject.

Give data subjects the possibility to exercise their rights.

Avoid deception in the communication with data subjects: processing of personal data should be in line with what they can expect.

The processing of personal data should not have a disproportionate negative, unlawful, discriminating or misleading effect on data subjects.

Transparent

Be transparent in the communication to your data subjects about who is processing the personal data (controllers, processors), which personal data are processed, as well as why and for how long, and how data subjects can exercise their rights. The information provided should be unambiguous, concise, easily accessible and relevant and shared with data subjects before the start of your research.

1.3.0.2   2. Purpose limitation

You can only process (i.e., collect, analyse, store, share, etc.) personal data for a specific purpose and only for as long as necessary to complete that purpose. For example, if you communicated to data subjects that you would use their personal data only to answer your specific research question, you cannot further share the personal data for new research questions, as these would be additional processing purposes. This means that you need to **plan what you will do with the (collected) personal data in advance and stick to that plan in order to be GDPR-compliant**.

1.3.0.3   3. Data minimisation

You can only process the personal data you need to for your predefined purpose(s), and not more just because they may "come in handy later". This principle makes sure that, for example, in the event of a data breach, the amount of data exposed is kept to a minimum.

1.3.0.4   4. Accuracy

The accuracy of personal data is integral to data protection. Inaccurate data can be a risk for data subjects, for example when they lead to a wrong treatment in a medical trial. You therefore need to take every reasonable step to remove or rectify data that is inaccurate or incomplete. Moreover, data subjects have the right to request that inaccurate or incomplete data be removed or rectified within 30 days.

1.3.0.5   5. Storage limitation

You can only store personal data for as long as is necessary to achieve your (research) purpose. Afterwards, they need to be removed. If the personal data are part of your research data (and not, for example, to simply contact data subjects), you are allowed to store (archive) them for a longer period of time, provided necessary safeguards are in place. This is an exemption that applies to data storage for scientific archiving purposes. You need to inform the data subjects on this storage duration beforehand.

If identification of the data subject is no longer needed for your (research) purposes, you do not need to keep storing the personal data just to comply with the GDPR, even if it means your data subjects cannot exercise their rights (art. 11).

1.3.0.6 6. Integrity and confidentiality

You have to process personal data securely and protect against unauthorised processing or access, loss or damage. To this end, you should put in place apropriate organisational and technical measures.

1.3.0.7 7. Accountability

The controller is ultimately responsible for demonstrating GDPR-compliance. As a researcher working with personal data, you are representing your institution (e.g., Utrecht University) and you should therefore be able to demonstrate that you process personal data in a compliant manner. Additionally, you should also have some knowledge of data protection so that you can implement the right measures into your research project.

1.4 Legal bases for working with personal data

You can only process personal data if you have a **legal basis** to do so, which should be registered, among other information, in the processing register and communicated to data subjects. There are 6 possible legal bases which are outlined below. In research, the legal bases 'informed consent', 'public interest' and to some extent 'legitimate interests of the controller' are most often used.

For different purposes in your research project, a different legal basis may apply. For example, you may contact data subjects before they start participating based on a legitimate interest and use informed consent for collecting, storing, analysing and publishing the data.

1.4.1 Legal bases suitable for research

Informed consent

Informed consent is the most frequently used legal basis in research and is often not only a legal (GDPR-consent), but also an ethical obligation (e.g., METC informed consent). When using informed consent, you should be able to demonstrate that the data subject was informed and has given consent, and for which purpose(s) they gave their consent. In all cases, consent has to be

freely given, specific, informed and unambiguous. Please refer to the Informed consent section for guidance on applying informed consent in your research.

Public interest

Public interest is sometimes used in research when the research is shown to clearly benefit the public good or fulfills a public task. In essence, public interest can be used for research that is conducted by employees of public institutions, when their research interest has been recognised by an official authority. For example, conducting research at Dutch universities has been officially recognised in the Higher Education and Scientific Research Act to be a public task. Public interest is often used when consent is not a good option. For example, it may be impossible or impractical to obtain consent when performing public observations or social media research. Or when participants actually do not have a free choice, such as in clinical trials when participants would experience significant disadvantages when not participating.

If you want to use public interest as a legal basis, you need to assess the necessity and proportionality of your processing. Additionally, you need to demonstrate that the interests of data subjects do not override your research interests. To do so, please contact your privacy officer to assess whether you can use this legal basis in your research.

Legitimate interest of the controller

Legitimate interest is often used by companies to process personal data necessary for the functioning of their own company, e.g., processing user data for fraud prevention, or keeping a registration system to provide better services. In research, legitimate interest is often used for processing activities that have no direct research purpose. For example, this can be the case when you need to collect contact information to approach data subjects to participate, and you can only obtain their consent for participating in your research after contacting them. Since contacting data subjects is a prerequisite to perform your research, it can be in the university's legitimate (research) interest to process their contact information.

To evaluate whether you can use legitimate interest as a legal basis, you always need to weigh the interests of the controller (e.g., Utrecht University) and the data subjects in a Legitimate interest assessment. Please contact your privacy officer to assess whether you can use this legal basis in your research.

1.4.2   Legal bases not suitable for reseach

Processing is necessary because of a legal obligation of the controller

This basis is not suitable for research. As an example, Utrecht University has to share tax data with the Dutch tax administration in order to comply with tax legislation.

Processing is necessary for the performance of a contract

This basis is not suitable for research. As an example, Utrecht University has contracts with its employees, which require it to manage the employees' financial data.

Processing is necessary to protect a person's vital interests

This basis is generally not suitable for research. If processing someone's personal data is crucial to their health or even life, that processing is allowed under the GDPR.

1.4.3   Further processing for research purposes

It may happen that you want to process personal data for other purposes than previously specified (e.g., because you formulated an additional research question), or you want to reuse previously collected personal data in your research. In these cases, it may be possible to make use of article 5(1)(b), which states that "further processing for [...] scientific purposes shall [...] not be considered to be incompatible with the initial purposes". Basically, this means that you can reuse personal data, that were previously collected for other purposes, for scientific research purposes. This is only allowed if you put in place sufficient safeguards to protect the personal data, inform data subjects, and allow them to exercise their rights (art. 89). "Further processing" is not strictly a legal basis. Instead, it functions as a way to legitimise *further* processing of personal data (which was previously collected for a different purpose, using one of the six legal bases) for research purposes.

Public interest, legitimate interest, and relying on further processing are ways to meet your *legal* requirements for processing personal data, but not necessarily your *ethical* requirements: you may still need consent if demanded so from an ethical perspective. Before you rely on any of these, you should first assess whether they are indeed suitable with your faculty privacy officer, and determine whether your research interests outweigh the privacy rights of the data subjects.

1.5   Data Subjects' Rights

The GDPR provides data subjects with several rights that gives them a relatively high degree of control over their own personal data. Below, we list these rights and how you can apply them in your research:

Right to be informed

Data subjects need to be clearly informed about what you are doing with their personal data (a.o. art. 12). This usually happens via a privacy notice or information letter. This right does not apply if your research will be seriously

harmed by meeting it and if you haven't obtained the personal data directly from the data subjects themselves.

Right of access

Data subjects have the right to access a copy of the personal data you have on them and to know what you are doing with that personal data and why (art. 15).

Right to rectification

Data subjects have the right to correct and complement the personal data that you have of them art. 16).

Right to erasure/Right to be forgotten

Data subjects have the right to have their personal data removed (i.e., equivalent to the right to withdraw consent, art. 17). This right does **not** need to be granted if:

the personal data are published and need to be archived for validation purposes.

it would seriously obstruct the research purpose(s).

it would hinder complying with a legal obligation or carrying out a task in the public interest.

If the personal data have already been made public or shared, you need to take reasonable measures to inform other users of the data of the erasure request. A privacy officer can help you with this.

Right to restriction of processing

Data subjects have the right to have you process less of their personal data (art. 18), for example if their personal data are inaccurate or your processing of it is unlawful or no longer needed.

Right to data portability

Data subjects have the right to have their personal data transferred to another party (art. 20).

Right to object

Data subjects have the right to object to what you are doing with their personal data. This right applies when the processing is based on legitimate or public interest (art. 21). In case of objection, you have to stop your processing activities, unless you can demonstrate grounds for overriding the data subject's rights.

1.5.0.1 How can data subjects exercise their rights?

Data subjects need to be **informed** about their rights and who to contact in order to exercise them, including when you use a legal basis other than informed

consent. In research, this is usually done via a privacy notice or information letter, which states a contact person responsible for handling questions and requests.

Incoming requests need to be **coordinated with a privacy officer**, so that they can be picked up in accordance with the GDPR. Additionally, at Utrecht University, data subjects can always contact privacy@uu.nl (the Data Protection Officer) for requests or complaints.

1.5.0.2 What to do when receiving a request concerning data subjects' rights?

You have to provide a substantive response to the data subject **within 30 days**, in the same way as you received the request. Depending on the complexity and number of requests, the response period may be extended by 2 months. In that case, you must inform the data subject about this extension (including the motivation) within one month. If needed, you can (and sometimes should) ask for additional information to confirm the data subject's identity.

For granting requests about data subjects' rights, there should be a procedure in place, in which you should at least consider:

- how you are going to retrieve the data (e.g., using a name-number key)
- who is responsible for granting the request and informing the data subject about it (e.g., a data manager)
- how the request is going to be granted, for example how they will be sent securely (access, portability), removed (forgotten, object, restriction) or corrected (rectification)

For larger projects, it may be wise to put a Standard Operating Procedure (SOP) in place.

1.5.0.3 What if the data have already been anonymised?

The principles of data minimisation and storage limitation are considered more important than keeping personal data just for the sake of identification (art. 11). Therefore, when receiving a request about anonymised data, you can make it clear that you cannot retrieve the data subject's personal data, because they have been anonymised. In this case, **the data subject cannot exercise their rights anymore**. If you can still retrieve the data subject's personal data in some way (i.e., when data are pseudonymised), you are **obliged to retrieve them**. In order to do so, you can (and sometimes should) ask for additional information that can confirm the data subject's identity.

Chapter 2

What are personal data?

In order to know whether you should comply with the GDPR in your research project, the first question to answer is: do you process personal data? To answer this question, we need to know: (1) What exactly are personal data, and (2) how do you know if you are working with personal data in your research?

2.1 Definition of personal data

According to the GDPR, personal data are "any information relating to an identified or identifiable natural person" (art.4(1)):

- **Natural person**: Data by themselves (numbers, text, pictures, audio, etc.) are not inherently personal. They only become personal when they refer to or relate to a **living individual**. When data refer to an organisation, deceased person, or group of individuals, they are not considered personal data under the GDPR.
- Data are personal if they **relate** to an individual. This means practically anything that someone is, has said or done, owns, may think, etc.
- The person should be **identified or identifiable**. This is the case not only through **directly** identifying information, such as names and contact information, but also through **indirectly** identifying information, for example if you can single someone out or identify them by combining datasets (see the next page).

2.2   How to assess whether data contain personal data?

Whether your data contain personal data depends on which data you are collecting (nature) and under which circumstances (context). A date like "12 December 1980", is not personal data – it is just a date. However, that date becomes personal data if it refers to someone's birthday.

In assessing whether data are personal, you should take into account all the means that you and others may **reasonably likely** use to identify your data subjects, such as the required money, time, or (future) developments in technology (rec. 26).

Data can be identifiable when:

- They contain **directly identifying information**.
  Examples (click to expand)
  name, image, video recording, audio recording, patient number, IP address, email address, phone number, location data, social media data
- It is possible to **single out** an individual, for example when there are unique data points or unique behavioural patterns which can only apply to one person.
  Examples
  You have a data subject who is 2.10 meters tall. If this is a unique value in your dataset, this distinguishes this person from others and thus can make them identifiable.
  You have a data subject who only follows far-right accounts on Twitter. If they are the only one in your dataset who do so, this distinguishes this person from others and can make them identifiable.
- It is possible to **infer information** about an individual based on information in your dataset.
  Examples
  Inferring a medical condition based on registered medications.
  Guessing that someone lives in a certain neighborhood based on where they go to school.
- It is possible to **link records** relating to an individual. This can happen when combining multiple variables within your dataset (e.g., demographic information, indirect identifiers). However, it can also happen when combining your dataset with other datasets (the "Mosaic effect"). In that case, your data still contain personal data, even if the data in your dataset are not identifiable by themselves.
  Examples
  Linkage is often possible with demographic information (age, gender, country of origin, education, workplace information, etc.) and indirect identifiers (pseudonyms, device ID, etc.), for example:

> In the year 2000, 87% of the United States population was found to be identifiable using a combination of their ZIP code, gender and date of birth.
> An agricultural company's Uniek Bedrijfsnummer (UBN) can be used to search for the address of the company in the I&R mobile app. Often, this address is also the owner's home address.

- De-identification is still **reversible**. This often happens when data are pseudonymised, but there is still a way to link the pseudonymised data with identifiable data, for example when a name-pseudonym key still exists.

You can assume that you are processing personal data when you collect data **directly** from people, even if the results of that collection are anonymous. But also when you use data that are **observed or derived from people**, even if those data were previously collected, made public or used for non-research purposes.

In short, even if you cannot find out someone's real identity (name, address), the data you process can still contain personal data under the GDPR. Besides the examples mentioned here, there are many other examples of personal data. If you need help assessing whether or not your data contain personal data, please contact your privacy officer or data manager.

2.3   Special types of personal data

There are a few special types of personal data that are worth taking note of: special categories of personal data, and otherwise sensitive personal data. These types of personal data have additional requirements. If you want to process them, please contact your privacy officer first.

2.3.1   Special categories of personal data

The GDPR explicitly defines seven 'special categories of personal data'. It is information that reveals:

- racial or ethnic origin
- political opinions
- religious or philosophical beliefs
- trade union membership
- genetic or biometric data when meant to uniquely identify someone
- physical or mental health conditions
- sex life or sexual orientation

It is in principle **prohibited** to process these types of personal data, unless an exception applies (art.9). For example, it is allowed to process these if:

- Data subjects have provided explicit consent to process these data for a specific purpose
- Data subjects have made the data publicly available themselves
- Processing is necessary for scientific research purposes (incl. historical and statistical purposes)

Even if you can make use of one of these exemptions, special categories of personal data warrant additional security measures to make sure they are protected. Always contact your privacy officer if you intend on processing these types of data.

2.3.2   Data that are otherwise sensitive

Other types of data can also be sensitive, because they can carry higher risks for the data subjects. These types of data can either not be processed at all, or only under certain circumstances. Either way, they require additional security measures. Always contact your privacy officer if you intend on using these types of data.

Examples are:

- Financial data
- Data about relationship problems
- Data that can be misused for identity fraud, such as the Dutch Citizen Service Number (BSN). In principle, the BSN cannot be used in research at all.
- Criminal or justice-related data: they can only be processed under governmental supervision or when a derogation exists in national legislation (art. 10).

Chapter 3

Designing your project

Research projects typically go through a number of stages: conception, proposal, planning, execution, publishing, preservation, etc. If you work with personal data, you should think about how you will protect those data throughout all those stages. To do so, the concepts of Privacy by Design and Privacy by Default (art. 25) are important:

- **Privacy by Design** in research means that your project integrates personal data protection right from the beginning, all the way throughout the project, and even afterwards. It should not be an afterthought: Privacy by Design is a key feature of the project, permeating all phases of a research project.
- **Privacy by Default** in research means that any questions, tools, or methods you use in your research should process as little personal data as necessary by default, and that you share the personal data only with those who really need access.

To get proper support in designing your project, it is important to contact your privacy officer early on, preferably already in the conception or design phase. The privacy officer will help you go through the different stages smoothly, and eventually save you time and effort. They can help you review and possibly adjust your plans, determine the appropriate protection measures, and determine whether you need to perform a more elaborate Data Protection Impact Assessment.

3.1 Privacy by Design strategies

To incorporate the concepts of Privacy by Design and Privacy by Default into your project, the approach of **privacy design strategies** (Hoepman, 2022) of-

fers a way to make the GDPR principles more concrete. Hoepman distinguishes 8 strategies that you can apply to protect the personal data in your research: minimise, separate, abstract, hide, inform, control, enforce, and demonstrate. In the next sections, we explain what these mean and how you can apply them.

The GDPR does not prescribe *which* specific measures you should apply in your project, only that they should protect the personal data *effectively*. Which measures will be effective, will depend on your specific project, the risks for data subjects, and the current progress in technology (i.e. will the data be protected on the long haul?).

3.2   Data-oriented strategies

3.2.1   Minimise

Limit as much as possible the processing of personal data, for example by:

- Collecting as little data as possible to reach your research purpose.
- Preferably not using tools that automatically collect unnecessary personal data. If possible, prevent tools you do use from doing so (Privacy by Default). For example, the survey tool Qualtrics can automatically register location data, which can be turned off by using the "Anonymize Responses" option.
- Removing personal data when you no longer need them. Remove them from repositories, data collection tools, sent emails, back-ups, etc. (see also the Storage chapter).

3.2.2   Separate

Separate the processing of different types of personal data as much as possible, for example by:

- Pseudonymisation: Store directly identifying personal data (e.g., contact information) separately from the research data. Use identification keys to link both datasets, and store these keys also separately from the research data.
- Using different databases to separate datasets.
- Applying secure computing techniques, where the data remain at a central location and do not have to be moved for the analysis.

3.2.3   Abstract

Limit as much and as early on as possible the detail in which personal data are processed, for example by:

- Anonymising or pseudonymising the data.
- Adding noise to the data, e.g., voice alteration in audio data.
- Summarising the data to simply describe general trends instead of individual data points.
- Synthesising the data, e.g., for sharing trends in the data without revealing individual data points.

3.2.4 Hide

Protect personal data, or make them unlinkable or unobservable. Make sure they do not become public or known. You can for example do so using a combination of:

- Using encryption, hashing or strong passwords to protect data. Keep in mind that you should store passwords securely as well!
- Using encrypted transport protocols (such as TLS, SFTP, HTTPS, etc.).
- Applying privacy models like Differential Privacy, where noise is added to individual data points to hide their true identity.
- Only providing access to people who really need it, and only for the necessary amount of time and with the necessary authorisations (e.g., read vs. write access; only the relevant selection of personal data, etc.). Remove authorisations when access is no longer required.

3.3 Process-oriented strategies

3.3.1 Inform

Inform data subjects about the processing of their personal data in a timely and adequate manner, for example by:

- Providing information via an information letter or privacy notice on a project website.
- Providing verbal explanation before an interview.
- Obtaining explicit consent via an informed consent procedure.

3.3.2 Control

Give data subjects adequate control over the processing of their personal data, for example by:

- Specifying a procedure and responsible person in case data subjects want to exercise their data subject rights.

- Providing data subjects with a contact point (e.g., email address) for questions and exercising their data subject rights.

3.3.3 Enforce

Commit to processing personal data in a privacy-friendly way, and adequately enforce this, for example by:

- Using only Utrecht University-approved tools to collect, store, analyse and share data.
- Entering into agreements with third parties if they are working with UU-controlled personal data. Such agreements will make sure everyone will treat the data up to UU-standards.
- Always keeping your software up-to-date.
- Appointing someone responsible for regulating access to the data.
- Always reporting data breaches. At UU, contact the Computer Emergency Response Team.
- If needed, drawing up a privacy and/or security policy that specify roles and responsibilities and best practices on how personal data are handled throughout a project.

3.3.4 Demonstrate

Demonstrate you are processing personal data in a privacy-friendly way, for example by:

- Registering your research project in the UU processing register (once available).
- Performing a Privacy Scan and storing it alongside the personal data.
- Performing a Data Protection Impact Assessment (DPIA) for projects that have a high privacy risk for the data subjects.
- Keeping information for data subjects and (signed) informed consent forms on file. This is not needed if you can fully anonymise the data: then you should delete the (signed) consent forms as well.

Chapter 4

Privacy & Security Assessment

Chapter 5

Documents & Agreements

This chapter addresses privacy-related documentation that is most commonly used within research. For each type of documentation, it explains what it entails, when to use it, and how.

The following documents are addressed:

- Privacy notices
- Informed consent forms
- Privacy scan
- Data Protection Impact Assessment
- Data classification
- Legitimate interest assessment
- Data Transfer Impact Assessment
- Several agreements, such as the non-disclosure agreement, data processing agreement, data transfer agreement, joint controller agreement, and data use agreement.

This chapter is a work in progress: sub-chapters are added continuously.

5.1   Privacy notices

A privacy notice is any information given to data subjects about what is happening with their personal data. In research, a privacy notice is usually combined with general information about the research project and often with an informed consent form, to satisfy both privacy and ethical concerns. Generally, the aim of a privacy notice is to inform data subjects on how and why their data are being processed. Providing that information is the "cornerstone of data subjects' rights", as without it, data subjects cannot exercise their other privacy rights.

5.1.1   When to use a privacy notice?

Informing data subjects is always required, for **all legal bases** (so not only when you use informed consent). Being properly informed is a data subject's right in itself (art. 12): it is necessary so that data subjects can exercise their other rights (e.g., right to be forgotten, right to object, etc.).

You need to inform data subjects **before** you start collecting or otherwise processing their personal data (so before the start of your research project).

When you use personal data from another source, you have to inform your data subjects within a month after obtaining their data (art. 14), except if this would involve a disproportionate effort, or would seriously impair your processing purposes (e.g., if you cannot answer your research question anymore, art. 14(5)(b)). If you share personal data with an external party, you should inform data subjects at the latest when first sharing those data with that external party.

5.1.2   Content and examples of privacy notices

Below you can find a list of items to include in your information to data subjects (Template) and some example sentences to (not) include in your privacy notice (click to expand).

Template privacy notices

Example sentences

Bad promise

Alternative

"After the project ends, we will delete all of your data, so that you will not be identifiable anymore."

"After the end of the study, we will delete the code linking your data to your name. We will store your de-identified data for 10 years for integrity purposes."

"Your data will be fully anonymised before they are shared with others."

"We will remove personal information that could reasonably identify you before we share any files with other researchers."

"All data that you will provide will be kept strictly confidential and will not be shared further."

"The main researcher will keep a link that identifies you to your coded information. They will keep this link secure and available only to the selected members of the research team."

"Your data will only be accessible by the research team, and no one else."

"We will only share your de-identified data with other researchers if they agree to treat your data confidentially and only after approval from the original research team."

"You can withdraw your consent from this study at any time up until the end of the research project. If you withdraw your consent, we will delete all your data from our dataset immediately."

"You can withdraw your consent from this study at any time, without stating a reason why and without any repercussions. Please inform the researcher about your decision. We will then delete any personal data referring to you that we still have, where this is still possible."

5.1.3   Form of a privacy notice

The format of the privacy notice is also crucial. Even if you include all necessary components in your privacy notice, it will **not be GDPR-compliant** if you fail to provide the information in an appropriate form, shape and time.

The information you provide to data subjects should be:

- **Clear and understandable** A privacy notice is not a legal document, so do not write it like you would write a legal contract. The information should be understandable for data subjects and it should have a clear and concrete meaning. For example, avoid using words like "may", "some", "often" and write active and short sentences.

- **Easily accessible** Data subjects should be able to find the information easily. For example, publish the privacy notice on your project website, give participants a copy, or provide a QR-code or short URL.

- **Via multiple channels (when appropriate)** Textual information sheets are by no means the only way to inform data subjects. If appropriate, you can provide the information via other channels too, e.g., oral statements, images, audio, video (example), etc. For some data subjects, such other channels of informing can lead to a better understanding of your processing activities.

- **Layered (when appropriate)** To balance being complete with being understandable, you can layer the information you provide. For example, provide concise information up front and provide more detailed information elsewhere (e.g., via a link or dropdown menu).

If you are uncertain about the level of intelligibility and transparency of the information, you can test these, for example through user panels, readability testing, or by interactions with data subjects themselves (or their representatives).

5.2   Informed consent

Of the 6 possible legal bases to process personal data, informed consent is currently the one most often used in research. With the term consent, we mean the process of data subjects deciding whether or not to agree to specific statements, such as a statement to participate in a research project.

5.2.1   Consent step-by-step

1. Determine if consent if the legal basis you need

   Determine if consent is the legal basis you need for your research: there are other legal bases besides consent which can sometimes be more suitable in a research context.

   In some situations, consent is likely the only way to process data, for example, if you want to process special categories of personal data, or if you process personal data from children under 16 years old. In the latter case, the GDPR requires to obtain additional obtain consent from a legal representative (e.g., parent), and there are additional requirements when your research falls under the Dutch Medical Research Involving Human Subjects Act.

2. Consider if you meet all requirements for consent

   If you need to use consent as a legal basis, consider if you meet all requirements listed below. If you do not, consent is not a valid legal basis, and you should consider another one.

3. Determine what you will ask consent for

   Determine what specifically you are asking consent for. If you cannot determine a specific purpose, for instance because your research question is not yet entirely clear, contact your privacy officer to consider obtaining broad consent.

4. Prepare information for data subjects

   Prepare a privacy notice or information letter for data subjects to inform them before asking for their consent.

5. Obtain demonstrable consent

   Different forms of consent are valid. Note that often a signature is not required.

6. Keep the consent forms available

   Treat the consent declarations as personal data: store them separately and securely from the research data, and for as long as your research data contain personal data.

Note that the term "consent" is used both in the GDPR as well as in an ethical context. As a legal basis, data subjects give consent to process their personal data (e.g., "I consent to my data a, b, c be used for purpose x, y, z"). In an ethical context, consent is a safeguard to give data subjects more control over their personal data, and makes sure they participate voluntarily in the research project (e.g., "I have read the information and agree to participate under the conditions described"). Thus, it can happen that consent is not be the best legal basis to use, but should still be used as an ethical requirement.

5.2.2 Requirements for valid consent

Under the GDPR, consent is only valid when it is **all** of the below (art. 4, art. 7, rec. 32, rec. 42, rec. 43; click to expand):

Freely given

Data subjects should have an actual voluntary choice and should not experience negative consequences if they don't consent or withdraw their consent. Moreover, they should not be pressured to provide consent, and so there cannot be a power imbalance between the controller (e.g., researcher) and data subjects (rec. 43). Some examples:

Consent is not valid when the researcher is also a teacher and asks their students to participate, who depend on the teacher for a good grade. In this case, another legal basis such as public interest may be more appropriate.

Consent can still be used for children and persons legally incapable to provide consent when their legal representative(s) provide the consent.

Specific

Data subjects should know as specifically as possible what they are asked to consent to. Separate processing purposes therefore require explicitly separate consent (rec. 32, rec. 43), and accompanying specific information that will allow the data subjects to decide if they consent or not. If consents for multiple purposes are necessary for your research, you can combine those. Some examples:

Combined consent may be possible to collect, store, analyse, and share personal data with your collaborators – all actions are needed to answer your research question.

Separate consent is needed for conducting a survey vs. for conducting a subsequent interview, if participation in that interview is not required for your research project.

Separate consent is needed for the current research project vs. for contacting data subjects for future research projects.

Informed

Data subjects need to be clearly and accessibly informed about which personal data are processed and why, and about their rights (see Privacy notices). Data subjects should be able to access this information easily (also after they have provided consent).

Unambiguous and affirmative

It should be clear what data subjects are providing consent for, using a clear, affirmative statement. Importantly, "silence, pre-ticked boxes or inactivity" do not constitute valid consent (rec. 32): consent should be active.

5.2.3   What forms of consent are valid?

The way you obtain consent may differ per research project and can depend on how you interact with your data subjects. The only requirement is that it should be demonstrable. Some examples:

- Ticking a box (**not** pre-ticked!)
- Writing or replying to an email ("I agree to be interviewed")
- Filling in an electronic form
- Audio- or video-recorded consent (separate it from the research data!)
- Signing a paper document (not usually necessary)

5.2.3.1   To sign or not to sign?

Signatures in consent forms are rarely needed. In fact, if you are only processing *pseudonymised* research data, you will only collect unnecessary personal data by obtaining a signature (art. 11), and a checkbox should be sufficient. In order to link the consent form with the data subject, you should include the pseudonym on the consent form (the identifier you will use for the participant, e.g., "part-001"). Inform your participants of this pseudonym; they can use it to exercise their rights under the GDPR, such as for withdrawing their consent.

Only when the identity of the data subjects will be used in the process (e.g., clinical trials), a signature may make sense or be required. For example, if your research is subject to the Dutch Medical Research Involving Human Subjects Act (WMO), different requirements apply, e.g., consent needs to be provided in writing using a signature and date.

5.2.4   Demonstrating (valid) consent

As long as you process personal data, you should be able to demonstrate that the data subjects consented to that processing (rec. 42). So as long as you analyse, use, store, archive, etc. the personal data, the proof of consent needs to be retained. It is preferable to store the proofs separately from the research data.

If you collected consent on paper, it is best practice to scan the consent forms and securely delete the paper version after having made sure the scanning went well. Only after there is no personal data anymore (e.g., after fully anonymising the dataset), you can remove the proof of consent.

5.2.5 Broad consent in research

In research, it can sometimes be difficult to formulate very specific research questions in advance. In this case, you may be able to formulate the research purposes on a more general level and obtain consent for these more general purposes (EDPS, 2020; Deutsche Datenschutzkonferenz, 2019). However, you can do this only as long as:

- data subjects can give consent to only part of the research and easily withdraw consent (rec. 33).
- data subjects are kept informed as specifically as possible about what will happen to their personal data. As soon as you know more, you should also inform data subjects in more detail.
- you use additional protection measures, for example:
    - obtain ethical approval for using the data for new research questions.
    - offer a consent withdrawal possibility before using the data for new research questions. This is especially relevant when it is still possible to reliably identify data subjects in the dataset.
    - make sure the data are not transferred to countries outside of the EEA, unless one of the derogations from GDPR Chapter V applies (e.g., adequacy decision, standard contractual clauses, explicit consent for transfer).
    - enforce specific requirements for access the data, e.g., "research in general" is not a sufficiently specific purpose for reuse of the personal data.
- you document your considerations and ask for help from a privacy officer.

5.2.6 Examples and templates

Note that all examples below assume that they are preceded by sufficiently specified information.

Template in Qualtrics Examples CESSDA

Example sentences

Good example sentences:

"I consent to the collection and use of my personal data to answer the research question described in the information letter."

"I consent to linking the new research data to data previously collected about me in this research project."

"I agree that research data gathered for the study may be published or made available provided my name or other identifying information is not used."

"I understand that the research data, without any personal information that could identify me (not linked to me) may be shared with other researchers."

Bad example sentences:

"Any information I give will be used for this research project only and will not be used for any other purpose": this restricts all future uses of the data, including sharing the data with your collaborators, performing analyses for new research questions, and sharing the data for reuse. It's preferred to tell data subjects how their data can be safely used in different ways.

"I do not give consent to share my data": this sentence is ambiguous and may confuse data subjects.

"I acknowledge that the personal data collected by the researcher belongs to the university and that I have no rights in the research performed on it": it is not allowed to deny data subjects their data subjects' rights.

5.3   Privacy scan

5.4   Data Protection Impact Assessment

5.5   Data classification

5.6   Legitimate interest assessment

5.7   Data Transfer Impact Assessment

5.8   Agreements

There are many types of agreements that can – and sometimes should – be set up during a research project. Which one you need depends on the purpose of transfer, where the data are transferred to, and the external party's role in the research project. The flowchart below gives an indication which agreement should be used in which situation. Under the figure, you can also find a short explanation about each individual type of agreement.

Quick links to: NDA DPA DTA Joint controllers agreement DUA SCCs

5.8.1 How to set up an agreement?

In order to set up an agreement, you should always get in touch with your Research Support Office (RSO) or privacy officer to get the ball rolling!

- Any agreement should be signed by someone who is authorised/mandated to do so. Usually this is a research director or faculty dean, but rarely you yourself. The RSO can tell you who in your case is mandated to sign the agreement.
- An agreement is not a replacement for consent or any other legal basis. It is a safeguard to make sure all parties involved treat the data safely and in accordance with the GDPR.

5.8.2 Non-disclosure agreement

A Non-Disclosure Agreement (NDA), or Confidentiality agreement, is an agreement that makes sure that either the receiver of the data or both parties handle data with care. Often, an NDA is meant to make sure that the receiving party keeps the data they get access to safe and processes the data according to specific guidelines. In research, it is often used between university researchers and students who perform research on their behalf. In this case, it is sometimes necessary to use an NDA, because students are not (always) bound to confidentiality through a contract with the university, whereas the researchers are.

Model NDA Utrecht University (two-sided) Example NDA (one-sided)

5.8.3 Data processing agreement

A data processing agreement (DPA) is mandatory when you transfer personal data to a third party who fulfills the role of a processor (art. 28(3)). In other words: a person or organisation that processes personal data on your behalf, without having a say as to why or how the data are processed. Important components of a DPA are a description of the data that are being shared, why they are being shared, what the third party can or cannot do with them, for how long, and what happens in case of a personal data breach. For example, the third party cannot use the data for their own purposes and is required to keep the data safe and report any potential data breaches to you.

A DPA is most often used when you use an application or tool that processes personal data. Examples of these are survey tools, analysis tools, transcription tools, documentation tools and data repositories. With many parties that offer such tools, there is already a processing agreement in place at the UU-level, and so they are already safe to use, see the Tooladvisor.

Utrecht University template (UU only) SURF template

5.8.4   Data Transfer Agreement

A data transfer agreement (DTA) is advisable when you transfer data to a third party who will (re)use the data for their own purposes, without having an active role in your research project. It is used often when this third party is an external institution, but is also recommended when the third party is someone from within your own institution. A DTA is used to ensure that both parties are aware of their responsibilities and are bound to do what the agreement says. It contains a description of the personal data that are being shared, why they are being shared, under which legal basis, and how the data should be protected by each party.

A DTA can be used when (for example):

- you want to share personal data for reuse purposes with other researchers. This also requires adding statements on the terms of use, although these terms of use can also be separately registered in a Data Use Agreement.
- you are using a software tool, and the software provider wants to run analytics on the personal data you process in their tool. This makes the software provider a controller, with a separate purpose from your own (e.g., answering your research question). Note that you likely already have a Data Processing Agreement with the software provider, and thus a clause in that DPA can also suffice in such cases.

Health-RI template Example Utrecht University

5.8.5   Joint controllers agreement

A joint controllers agreement is mandatory when you work together with another controller on the same personal data, and you have common purposes (why) and means (how) of processing. In a joint controllers agreement, the respective responsibilities of both (all) parties are formalised, e.g., who informs data subjects, who is the contact point for data subjects, and how are data kept secure (art. 26). In research, this happens most often in a consortium, where multiple institutions participate in a research project. Therefore, a joint controllers agreement is often part of the consortium agreement, in which also topics other than the processing of personal data are formalised (e.g., intellectual property rights, how data are shared, etc.).

Examples Erasmus University Rotterdam template SURF template

5.8.6   Data Use Agreement

A data use agreement (DUA), or user agreement, is basically a custom license for your dataset. It specifies the terms and conditions under which the receiver

of the data can (re)use the data and is therefore in many ways similar to a Data Transfer Agreement. For example, it may contain statements on what the receiver can do with the data, and how the original data owners should be attributed. It can also state that the receiver must comply with the GDPR and cannot try to reidentify data subjects. DUAs are often used in data repositories (e.g., custom terms of use in DataverseNL) or as part of a Data Transfer Agreement, and often (but not always) when an open license is not suitable (e.g., with personal data).

Example Donders Institute

5.8.7 Standard Contractual Clauses for international transfers

Standard Contractual Clauses (SCCs, art. 46) are model clauses that have been "pre-approved" by the European Commission to include in agreements. They are specifically meant as a (sometimes necessary) safeguard when personal data are transferred to processors and controllers outside of the EEA. This is because the SCCs contain (among others) a list of minimal necessary safeguards that the receiving party should implement to ensure that the personal data are properly protected. SCCs only have to be used in specific situations, and they should ideally be preceded by a Data Transfer Impact Assessment that identifies SCCs as an appropriate measure. Please contact your privacy officer if you have questions about them.

SCCs for international transfers

There are also SCCs for transfers to processors *within* the EEA (art. 28(7)). These can be included in a Data Processing Agreement. Because they are standardised clauses, including these can make it easier to finalise a DPA.

Techniques & Tools

Chapter 6

Privacy-Enhancing Techniques

6.1 Pseudonymisation & Anonymisation

6.2 Secure computing

6.3 Encryption

6.4 Synthetic Data

6.5 Differential privacy

6.6 Data donation

Chapter 7

Tools & Services

When you are using a tool that processes personal data, that tool should do so in compliance with the GDPR. If you work at Utrecht University (UU), you can use https://tools.uu.nl to find not only suitable storage media and data repositories, but also an overview of which tools are safe to use. It contains for example tools for audio transcription, file sharing, and performing surveys.

Most of the tools listed are safe to use either because no (personal) data are being used by the tool, data are processed at UU premises, or because of a Data Processing Agreement between UU and the supplier of the tool, in which the supplier agreed to sufficiently protect the data entered into their tool.

Are you developing a website or application yourself that uses user data? Check out the CNIL GDPR Guide for Developers for step-by-step guidance on how to develop your software in compliance with the GDPR.

7.1 Requirements for a third-party tool

If your tool of choice is not listed in https://tools.uu.nl, please contact IT servicedesk. They will help you assess whether a tool is safe to use. If a tool is processing personal data, two aspects are important to consider: 1) who is controlling the personal data, and 2) the security level of the tool.

7.1.1 1. Who is processing the personal data: arrange an agreement

When you use a third-party tool that processes personal data, the data are not under the (full) control of your employer. In this case, it is required to ensure GDPR compliance of the tool provider using (art. 46):

- A Data Processing Agreement - when the provider processes (e.g., stores, analyses, collects) personal data within the European Economic Area (EEA) or a country with an adequate level of data protection.
- Standard contractual clauses (SCCs) - when personal data are processed by a supplier outside of the EEA. These make sure the provider will use sufficient measures to protect the personal data and enable data subjects to exercise their rights.
- Explicit consent of data subjects who have been informed on the risks involved - in the absence of an agreement. Please contact your privacy officer if you are considering this option.

You can assume agreements are in place for the tools recommended by UU. If there is no agreement in place between UU and the tool provider, using this tool is **not allowed**, even if the provider is located within the EEA, has an adequate level of data protection, or has high security standards. The only exception is when data are always end-to-end encrypted, because then the tool provider cannot learn anything from the data.

7.1.2   2. Security level

The tool provider should employ good security practices, such as regular backups in distinct geographical areas (preferably in replication rather than on tape), regular integrity checks, encryption at rest, multi-factor authentication, etc. Most of these aspects will likely be covered in the agreement, and sometimes a data classification will need to be performed. Information security can help you determine all necessary security requirements.

Storage, Sharing, Publication

Chapter 8

Storing personal data

In research, storage of personal data is one of the most common processing activities. Assuming you have a legal basis to store personal data, you then need to:

- Choose a storage medium that is GDPR-compliant and that provides a sufficient level of data protection;
- Take into account procedural and legal aspects, e.g., how will you handle the data once they are stored, and for how long will you store the data?

These aspects of storing personal data are discussed in this chapter.

8.0.1 Chapter summary

Where should I store personal data?

Use a medium that has been approved by your institution. If you work at Utrecht University, and your preferred storage medium is not included in the Storage Finder, then please contact RDM Support or your privacy officer to find an alternative solution.

How to store personal data?

- Apply organisational and technical safeguards, e.g., restrict access, encrypt data, pseudonymise data, specify responsibilities, etc.
- Store (personal) data preferably in a structured, commonly used, machine-readable and interoperable format: others should be able to open, understand and work with your data.

For how long should I store personal data?

- Delete or fully anonymise personal data when they are no longer necessary, and preferably determine when you will do this in advance.
- In research, you can archive personal data that are necessary for validation purposes for a longer period of time, e.g., 10 years or longer.

8.1 Where should I store personal data?

8.1.1 Storage media at UU

If you work at Utrecht University (UU), you can find a suitable storage medium for digital research data via the Storage Finder (see below). For personal data, select Basic, Sensitive or Critical (depending on the sensitivity of your data) under question 4 about Confidentiality.

Most storage media in this overview are suitable for storing personal data, either because they are controlled by UU (e.g., U- and O-drive, Beta File System) or because UU has a Data Processing Agreement in place with the storage supplier (e.g., Microsoft Office 365, Yoda, SURFdrive).

Is your preferred storage medium not included in the storage finder? Contact RDM Support, your data steward, or your privacy officer to find an alternative solution.

- Consider encrypting your data, especially when using **portable devices** (e.g, memory sticks, phones). Portable devices are also not suitable as back-up, due to bit rot and being easily lost.
- Physical personal data (e.g., paper questionnaires, informed consent forms) should be stored securely too, e.g. in a locked room, cabinet or drawer. You should also avoid leaving unsecured copies lying around (e.g., on a desk or printer).

Do not store research data containing personal data on public cloud services, e.g., Google Drive, Dropbox, OneDrive, Box, Mega, iDrive, iCloud, NextCloud, etc. These services are not (always) GDPR-compliant and/or may not offer sufficient data security. Moreover, UU does not have any formal agreements with these services, enabling them to use the data stored on their platforms for their own purposes.

8.2 How should I store personal data?

Once you have chosen a suitable storage medium, you should act in accordance with the nature of your data as well, for example through:

- Controlling access: make sure that only the necessary people have the right kind of access (e.g. read/write) to the personal data, and remove their access when they do not longer need it (e.g. when someone leaves the research project).

- Specifying responsibilities, e.g. who is responsible for guarding access to the data on both the short and the long term? Make people aware of the confidential nature of the data. Tell them what to do in case of a data breach.
- Procedural arrangements, e.g. capture access conditions in agreements like the consortium agreement, data processing agreement or non-disclosure agreement.
- Storing different types of personal data in different places, e.g., research data should be stored separately from data subjects' contact details.
- Applying other safeguards where appropriate, e.g., encryption, anonymisation, pseudonymisation, etc. (see chapter Privacy-enhancing techniques).

See Designing a GDPR-compliant research project for more tips.

Personal data should be stored in a **"structured, commonly used, machine-readable and interoperable format"** (rec. 68). In practice, this means that you should consider whether your files are structured and named in a logical way, use sustainable file formats, and provide understandable metadata so that others can interpret the data. You can read more about this in the RDM guide "Storing and preserving data".

8.3 For how long should I store personal data?

As per the GDPR, anyone processing personal data can only store those for as long as is necessary for prespecified purposes (art. 5(e)). Afterwards, the personal data have to either be fully anonymised or deleted. However, there is an exemption for research data, as described below.

In research, we often see a division in different types of retention periods:

- If the personal data underpin a scientific publication, it is usually necessary to archive some personal data for **integrity and validation purposes** (art. 5(e)), because they are part of the research data. At UU, any research data necessary for validation should be archived for at least 10 years (UU research data policy). If this includes personal data, they too should be archived. Importantly, this still means that you need to **protect** the personal data, and **limit** the personal data stored to the amount necessary for validation (art. 89)! This also implies that you should keep

the documentation about the legal basis used (e.g., consent forms) during that time, so that you can demonstrate GDPR compliance.
- Specific retention periods may apply additionally to specific types of data. For example, in the Netherlands there are specific retention periods for medical data.
- Personal data that were used for purposes other than answering your research question (e.g. contact information) should have their own retention policy: they should be removed or anonymised after the retention period (e.g. the research project) has ended.

For all types of data in your project (incl. to be archived research data), we recommend to formulate which data you will retain and for how long (for example in your Data Management Plan), and communicate the (possibly different) retention period(s) to data subjects. If you want to change the storage term you initially set and communicated for your personal data, please contact your privacy officer.

8.3.1   Deleting personal data

If you do not need personal data anymore, you must delete it, except when the data should be archived for validation purposes. When deleting data, it is important to make sure that there are no visible or hidden copies being left behind and that files cannot be recovered. The Storage Finder will soon include information on how to fully delete data on the UU-approved storage media. For your own file system, you can use software like BleachBit, BCWipe, DeleteOnClick, and Eraser to delete data.

Chapter 9

Sharing data with collaborators

This chapter addresses guidelines to take into account when you want to share personal data with collaborators outside of your own institution **during** your research project. For guidelines to share personal data after a research project, please refer to the chapter on Data sharing for reuse.

To be able to share personal data with external collaborators, you should:

1. Make sure you have a legal basis and inform data subjects

Make sure data subjects are well-informed about your intentions to share the data with collaborators. Include information in your privacy notice on the identity of your collaborators, which data are shared with them and why, how, and for how long. Avoid using statements that preclude sharing such as "Your data will not be shared with anyone else".

Make sure you have a legal basis to share the data, e.g., informed consent or public interest. If you use consent, make sure that data subjects are aware that they are also providing consent to share their data with your collaborators.

Inform data subjects timely - before you start processing their data - and proactively - directly if possible.

2. Protect the personal data appropriately

Assess the risks of sharing the data and the measures you will take to mitigate those in your Data Management Plan, Privacy scan, or if applicable, Data Protection Impact Assessment. This is especially important if you will share your data with collaborators outside of the European Economic Area.

Share only the data that the collaborator needs (data minimisation), for example by deleting unnecessary data, pseudonymising the data, and sharing only with those who need access to the data.

Make sure data subjects can still exercise their data subjects' rights. For example, if a data subject withdraws their consent, not only you, but also your collaborators will have to stop processing the data subject's personal data. It is important to make clear how you and them will do so.

3. Come to agreements with collaborators

In order to protect the personal data effectively, it is important to determine which role every collaborator has: controller or processor? And if there are multiple controllers, are they separate or joint controllers? For example, in many collaborative research projects (e.g., in consortia), there are multiple controllers that collectively determine why (e.g., research question) and how (e.g., methods) to process personal data. These parties are then joint controllers, and agreements need to be made in a joint controllers agreement.

In any collaboration in which data are shared, you need to (art. 26):

Come to a formal agreement on:

The role of each party in the research project

Respective responsibilities in terms of data protection, such as informing data subjects and handling requests relating to data subjects' rights

Who is the main point of contact for data subjects

Communicate (the essence of) the agreement to data subjects.

Your privacy officer can help you draw up a valid agreement.

4. Pay special attention to third-country transfers

If you share personal data with international collaborators (for example, with countries that have no adequacy decision), you may need to take additional measures. Usually, these measures include drawing up an agreement to make sure the other party is GDPR-compliant and uses the necessary security measures (if you haven't already done so). The exact type of agreement will depend on your specific situation: your privacy officer can help you choose and set up the right one.

The flowchart below indicates conditions under which you can share data internationally. Note that they assume that you have taken sufficient safeguards to protect the personal data. To determine the possibilities of sharing data internationally in your project, we strongly advise you to consult with your privacy officer. In some cases a Data Transfer Impact Assessment may be required, which can take some effort.

5. Use a secure way to share the data

Granting access: It is preferable to grant a user access to an existing and safe infrastructure (e.g., add someone to a Yoda group or OneDrive folder), rather than physically sending the data elsewhere. This allows you to keep the data in one place, define specific access rights (read/write), have users authenticate,

and easily revoke access to the data after your collaboration has ended. It is also a good idea to take measures to prevent the data from being copied elsewhere.

Transferring data: When it is absolutely necessary to transfer the files to a different location, you must do so securely. Researchers at Utrecht University can use SURF Filesender with encryption.

Chapter 10

Sharing data for reuse

Chapter 11

(PART*) Use Cases

Chapter 12

Data minimisation in a survey

For a course, a teacher at the faculty of Veterinary Medicine collected data on the health of pets and the pets' owners. The initial purpose of the survey was to create simply datasets for students to learn about statistics. However, besides for the course, the teacher also wanted to use the collected data for research purposes and share the data with others. In order to do so, the teacher created a new version of the survey that asked for less identifiable information and could be more easily anonymised. Additionally, the new version of the survey informed participants about the legal basis used to process their personal data.

Here, you can find the survey before and after data minimisation:

Before minimisation After minimisation

Note that the new version of the survey:

- minimises the amount of personal data collected:
    - Student number and pet names are not asked in the new version of the survey.
    - Instead of Age, the new version asks the Age category of the owner/caretaker.
    - The survey includes questions on Weight and Height. For data publication, they are used to calculate the Body Mass Index (BMI) and deleted after this calculation.
- contains information about the legal basis used to be able to use (legitimate interest) and publish (consent) the data for purposes other than education.

Chapter 13

Data pseudonymisation

Coming soon!

Chapter 14

Publishing metadata

In 2020, the Open Science Programme of Utrecht University sent out the first
Open Science monitor. The aim was to gain insights into the awareness, atti-
tudes, practices, opportunities and barriers of employees of Utrecht University
and Utrecht University Medical Center regarding several Open Science prac-
tices. As the dataset contained a lot of demographic information (e.g., gender,
age, nationality, position, type of contract, etc.), and all of those variables com-
bined could lead to identification, it could not be shared publicly. For this
particular dataset, full anonymisation was not desirable, as that would greatly
decrease its scientific value. Therefore, the Open Science Programme chose to
publish only the metadata and documentation, without sharing the data, in or-
der to protect participants' data while still complying with the FAIR principles.

Here's the strategy they took:

- They published the dataset under restricted access on Yoda, so that the
  dataset was at least Findable and Accessible.
- They shared other relevant documentation publicly:
    - The preregistration.
    - The questionnaire itself, including the information provided to par-
      ticipants.
    - The final report written about the dataset.

Note that in the metadata of all these publications, cross-references to the other
publications are included to allow for maximum findability of the project's out-
puts.

Chapter 15

Reusing education data for research

A research group at the Science faculty wanted to investigate the effects of the Covid-19 pandemic on students' motivation and study success in a specific course. To do so, they wanted to analyse:

- Students' evaluations of the course from both before and during the pandemic.
- Students' test and final grades in the course from both before and during the pandemic.

The primary researchers already had access to the data for their educational activities, and so they wanted to use the data for research purposes. They went to their faculty privacy officer to find out how they could reuse these data in a responsible way.

The following privacy issues are relevant in this use case:

- **The raw data were identifiable** The student grades were linked to names, and both the grades and the evaluations were linked to student IDs. Moreover, the evaluations could potentially contain names of teachers and other personal information, as they consisted of partly open-ended questions. To decrease identifiability, the principal investigator and a second examiner, who already had access to the students' data, first removed or replaced all names with pseudonyms (both names of student and teachers), and went through the open-ended questions to remove potentially directly identifiable information. Only after deidentification were the data shared with research assistants who performed the main data and content analyses.

- **Data subjects' rights** Most students had already finished the course, and were not informed about the use of their evaluations and grades for this research project. The researchers argued that the majority of the students could not be traced anymore to provide this information or to enable them to exercise their data subjects' rights (art. 14(5)(b)). Moreover, in case a student did want to exercise their rights, it would prove difficult to retrieve the correct data, as the data were deidentified as soon as possible.
- **Legal basis** Students did not provide explicit consent to process their grades and evaluations for this research project. Moreover, if they had provided consent, it could be argued that the consent was not freely given, as the primary researchers were also involved as teachers, and therefore there was a hierarchical relationship between the students and the teachers. For these reasons, consent was not a suitable legal basis in this case. Instead, the researchers relied on:

  - **Public interest**: processing students' data for the course itself is a public task, namely that of providing education. It was the legal basis for the initial data collection.
  - **Further processing for scientific research purposes**: processing data to answer the research question can be considered as secondary use of the students' personal data. The GDPR does not consider secondary use of personal data for scientific research purposes incompatible with the original purpose (i.e., the original purpose being to provide education and improving the course, art. 5(1)(b)). Thus, it was not necessary to rely on a new legal basis for this research project, provided that the data were protected sufficiently: The researchers made sure that the data were well-protected (i.e., minimised, pseudonymised, and access controlled, art. 89).

Resources

Chapter 16

Seeking help at Utrecht University

The glossary consists of frequently used jargon concerning the GDPR and research data.

16.0.0.1 A

Anonymous data

Any data where an individual is irreversibly de-identified, both directly (e.g., through names and email addresses) and indirectly. The latter means that you cannot identify someone:

by combining variables or datasets (e.g., a combination of date of birth, gender and birthplace, or the combination of a dataset with its name-number key)

via inference, i.e., when you can deduce who the data are about (e.g., when profession is Dutch prime minister, it is clear who the data is about)

by singling out a single subject, such as through unique data points (e.g., someone who is 210 cm tall is relatively easy to identify)

Anonymous data are no longer personal data and thus not subject to GDPR compliance. In practice, anonymous data may be difficult to attain and care must be given that the data legitimately cannot be traced to an individual in any way. The document Opinion 05/2014 on Anonymisation Techniques explains the criteria that must be met for data to be considered anonymous.

16.0.0.2 C

Controller

The natural or legal entity that, alone or with others, determines or has an influence on why and how personal data are processed. On an organisational level, Utrecht University (UU) is the controller of personal data collected by UU researchers and will be held responsible in case of GDPR infringement. On a practical level, however, researchers (e.g., Principal Investigators) often determine why and how data are processed, and are thus fulfilling the role of controller themselves.

16.0.0.3   D

Data subject

A living individual who can be identified directly or indirectly through personal data. In a research setting, this would be the individual whose personal data is being processed (see below for the definition of processing).

16.0.0.4   E

European Economic Area (EEA)

The member states of the European Union and Iceland, Liechtenstein, and Norway. In total, the EEA now consists of 30 countries. The aim of the EEA is to enable the "free movement of goods, people, services and capital" between countries, and this includes (personal) data (source: Eurostat).

16.0.0.5   H

Hashing

Hashing is a way of replacing one or multiple variables with a string of random characters with a fixed length. It can be used to create a "hashed" pseudonym, or to replace multiple variables with one unique value. It is usually quite difficult to reverse the hashing process, except if an attacker has knowledge about the type of information that was masked through hashing. To prevent reversal, cryptographic hashing techniques add a "salt", i.e., a random number or string, to the hash (the result is called a "digest"). If the "salt" is kept confidential or is removed (similar to a keyfile), it is almost impossible to reverse the hashing process.

16.0.0.6   L

Legal basis

Any processing of personal data should have a valid legal basis. Without it, you are now allowed to process personal data at all. The GDPR provides 6 legal

bases: consent, legitimate interest, public interest, legal obligation, performance of a contract, vital interest. Consent and public and legitimate interest are most often used in a research context.

16.0.0.7   P

Personal data

Any information related to an identified or identifiable (living) natural person, for example identifiers (name, identification number, location data, online identifier or a combination of identifiers) or factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of the person. Moreover, IP address, opinions, tweets, answers to questionnaires, etc. may also be personal data, either by itself or through a combination of one another.

Of note: as soon as you process data related to a person that is identifiable, you are processing personal data. Additionally, pseudonymised data is still considered personal data. Read more in Designing a GDPR-compliant research project.

Processing

Any operation performed on personal data, including collection, storage, organisation, alteration, analysis, transcription, sharing, publishing, deletion, etc.

Processor

A natural or legal entity that processes personal data on behalf of the controller. For example, when using a cloud transcription service, you often need to send personal data (e.g., an audio recording) to the transcription service for the purpose of your research, which is then fulfilling the role of processor. When using such a third party, you must have a data processing agreement in place.

Pseudonymous data

Personal data that cannot lead to identification without additional information, such as a key file linking pseudonyms to names. This additional information should be kept separately and securely and makes for deidentification that is reversible. Data are sometimes pseudonymised by replacing direct identifiers (e.g., names) with a participant code (e.g., number). However, this may not always suffice, as sometimes it is still possible to identify participants indirectly (e.g., through linkage, inference or singling out). Importantly, pseudonymous data are still personal data and therefore must be handled in accordance with the GDPR.

16.0.0.8   S

Special categories of personal data

Any information pertaining to the data subject which reveals any of the below categories:

racial or ethnic origin

political opinions

religious or philosophical beliefs

trade union membership

genetic and biometric data when meant to uniquely identify someone

physical or mental health conditions

an individual's sex life or sexual orientation

The processing of these categories of data is prohibited, unless one of the exceptions of art. 9 applies. For example, an exception applies when:

the data subject has provided explicit consent to process these data for a specific purpose

the data subject has made the data publicly available themselves

processing is necessary for scientific research purposes

Contact your privacy officer if you wish to process special categories of personal data.

Chapter 17

References

For further reading, we prepared a Zotero library with additional resources, some of which are specific to Utrecht University, others more general. Click on the image below to see the most recent version of the reference library online.