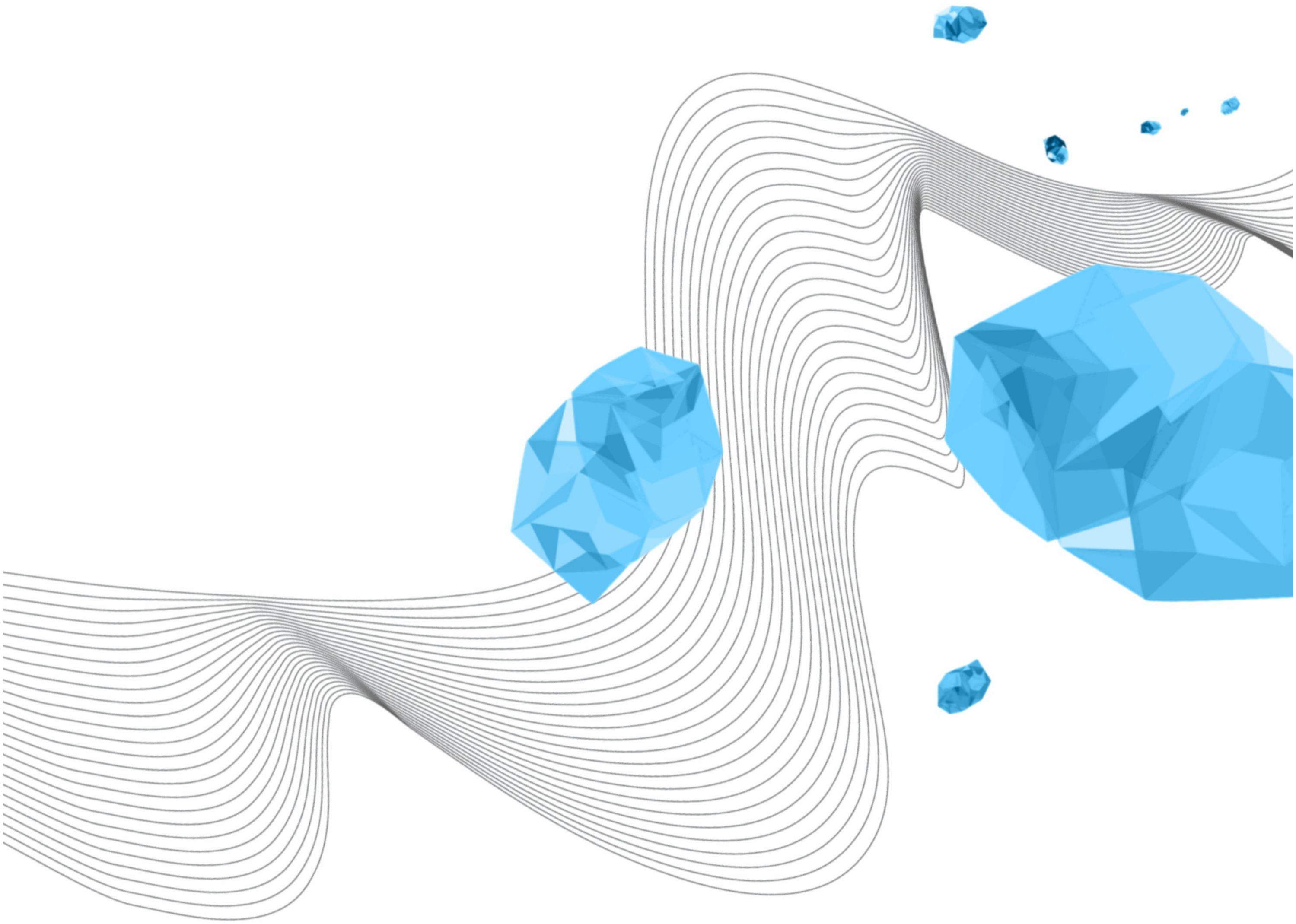
LIBRARY, ICT SERVICES & ARCHIVE

GUIDELINE: TESTING WITH PERSONAL DATA UNDER THE GENERAL DATA PROTECTION REGULATION (GDPR)

Floris Aanstoot

3.0

29-08-2023



UNIVERSITY OF TWENTE.

PUBLISHING DETAILS

ORGANISATION

Library, ICT Services & Archive

TITLE

Guideline: testing with personal data under the General Data Protection Regulation (GDPR)

VERSION (STATUS)

3.0

DATE

19-12-2023

AUTHOR(S)

Floris Aanstoot

DOCUMEN THISTORY

0.1	10-07-2018	Floris Aanstoot	Initial version.
0.2	12-07-2018	Floris Aanstoot	Internal review processed. Extensively detailed, examples of personal data added.
0.3	20-07-2018	Floris Aanstoot, Erik van den Bosch	Review comments processed: Marc Berenschot, Rianne te Brake, Joyce Pasman and Erik van den Bosch. Rewritten from advice to guideline.
0.4	12-09-2018	Floris Aanstoot	Guideline completed.
0.5	21-09-2018	Floris Aanstoot, Erik van den Bosch	Review comments Erik van den Bosch processed.
0.6	08-10-2018	Floris Aanstoot	Review comments and approval of the members of the I-Beraad (custodians of university systems) processed.
1.0	08-11-2018	Floris Aanstoot	Approval of the members of the CDO (Central Director's Meeting) processed.
1.1	04-02-2020	Meike Davids	Advice from attorney Mirjam Elferink processed in the guideline.
1.2	17-02-2020	Meike Davids	Review comments Henk Swaters processed; procedure added.
2.1	07-07-2022	Floris Aanstoot, Meike van de Ven	The guideline has been updated.
2.2	29-08-2023	Floris Aanstoot, Annika van der Putten	Changes to chapter 2 due to new ruling oct '22
3.0	19-12-2023	Annika van der Putten	Review comments Rene van Arnhem processed

DISTRIBUTION LIST

0.2	12-07-2018	Floris Aanstoot	Erik van den Bosch, Henk Swaters, Joyce Pasman, Arno Holterman, Marc Berenschot, Rianne te Brake and Daisy Oolbekkink
0.3	20-07-2018	Floris Aanstoot, Erik van den Bosch	Erik van den Bosch
0.4	19-09-2018	Floris Aanstoot	Erik van den Bosch
0.5	21-09-2018	Floris Aanstoot, Erik van den Bosch	Jan Evers, Erik van den Bosch, members of the I-Beraad
0.6	08-10-2018	Floris Aanstoot	Jan Evers, Erik van den Bosch, members of the I-Beraad
1.0	08-11-2018	Floris Aanstoot	Made publicly available on the cyber safety website
2.1	07-07-2022	Floris Aanstoot, Meike van de Ven	The updated guideline is published on the cyber safety website.
3.0	19-12-2023	Floris Aanstoot, Annika van der Putten	The updated guideline is published on the cyber safety website.

TABLE OF CONTENTS

1 INTRODUCTION

1.1 Motivation

On 25 May 2018, the General Data Protection Regulation (GDPR) became effective. The GDPR is effective across the European Union (EU). As per 25 May 2018, the Dutch Personal Data Protection Act ('Wet bescherming Persoonsgegevens') is no longer effective.

The University of Twente (UT) may use personal data[1] when testing its information systems. The GDPR's explanation of test data is ambiguous and subject to many interpretations. In order to ensure consistency and to create a common approach, an analysis was carried out that resulted in this guideline for handling test data within the UT.

1.2 Scope

This guideline limits itself to the UT's institutional systems. The highly integrated nature of these systems is reflected by the dependency of the data collections of these systems.
This guideline has been written with this dependency in mind: how can we best achieve a comprehensive solution where test data is concerned?

This guideline, or parts thereof, may be suited to non-institutional systems as well, but these systems were not included in the analysis.

1.3 Purpose of the document

This guideline describes how the UT handles personal data when testing its institutional systems.

[1] See '5 Appendix 1: personal data' to find the definition of personal data.

2 PRINCIPLES OF TESTING WITH PERSONAL DATA

Testing with personal data is considered a way of processing personal data. Processing personal data has to meet the requirements of the GDPR. For example, there must be a legitimate purpose and a legal basis for the processing, as well as compliance with the principles of the GDPR (amongst others data minimisation and integrity and confidentiality of the personal data). And lastly, the processing operation must be included in the register of processing operations.

The GDPR also stipulates that personal data may not be further processed ('further processing') in a manner that is incompatible with the purposes for which the personal data were initially collected. When testing with personal data, the personal data were initially collected for a certain purpose and are subsequently processed for a different purpose during the testing of personal data. The question is whether this purpose (testing the system(s)) **is incompatible** with the original purpose. If this is not incompatible, you must still adhere to the other conditions of the GDPR, but a separate legal basis is not required.

The Dutch Data Protection Authority (Dutch DPA) is the independent supervisor in the Netherlands that supervises and monitors the protection of personal data.[2] The Dutch DPA has indicated in the past that they do not recommend testing with personal data. This is because testing is a complex process that requires care and multiple separate environments. After all, testing with personal data entails risks.

The European Court [3]hears requests from national courts for a preliminary ruling, certain annulments and appeals. In October 2022 they ruled in a case related to testing with personal data[4]. They indicated that if the personal data used for testing is consistent with the purpose for which the personal data was collected, then using the same data for testing would not be in conflict with the GDPR. For example: for improving an app, for where the personal data was primarily collected, the purpose could be compatible.

Since test data is sometimes handled with less care than production data, the risk of data breaching is lurking. That is why there always should be considered whether there is not a better alternative, like fictitious data.

Appendix 2 to this document contains a step-by-step plan that should be followed before working in a test environment. Further substantiation of the various steps follows in section 4 of this guideline.

[2] Taken en bevoegdheden van de AP | Autoriteit Persoonsgegevens

[3] CURIA - Algemene presentatie - Hof van Justitie van de Europese Unie (europa.eu)

[4] EUR-Lex - 62021CJ0077 - EN - EUR-Lex (europa.eu)

3 THE IDEAL SITUATION

... And why this is not entirely feasible within the UT ...

In an ideal situation, the UT would not use personal data for testing its institutional systems. In that situation, data cannot be traced back to an individual. This can be done by:

- Testing with anonymised data
    - To create a test set, a copy is made of the data from the production environment. The personal data in this test set are anonymised using special software.
- Testing with fictional data
    - The complete test set is designed. None of the data from the production environment is used.

The holders of institutional systems have carried out an impact analysis for both measures. This impact analysis has shown that it is not feasible for all institutional systems to carry out tests without any personal data.

- In some institutional systems it is technically not possible to anonymise all personal data, as some personal data are part of the technical key in the database.
- Institutional systems work together in chains. Many links are created between data. For each data set that needs to be anonymised, an assessment has to be made as to whether it is possible to fully anonymise those data. In some situations, it might be possible to combine data sets and still identify individuals.
- Institutional systems use structured information (e.g. tables in a database) and unstructured information (e.g., email, documents, log files). Information on anonymisation and also anonymisation software primarily focus on structured information. Few articles have been written about the anonymisation of unstructured information, nor do there seem to be any fitting solutions available on the market.
- The UT does not have the required testing expertise to create a test set with only fictitious data.
- The UT does not have the required production data expertise to create a representative test set with fictitious data.

Due to the reasons mentioned above, it cannot be guaranteed that no personal data is included in a chain.

4 GUIDELINE

The GDPR states, amongst others, that you may not process more personal data than necessary. In practice, this means that it must be determined how the purpose can be reached in the most privacy-friendly way. In case there are no alternatives possible that are more privacy-friendly, this must be explained. Therefore, the basic principle of this guideline is: in principle, use no personal data, and if that is not possible, use as little personal data as possible for testing systems and applications.

The holder of an institutional system is responsible for complying with the GDPR, and therefore also for complying with the GDPR when it comes to testing with personal data. A precondition for this is that integrated testing must remain possible. After all, integrated testing guarantees the correct operation of the chain of information systems that support the UT's primary processes and is also required by the accountant in certain cases; this operation cannot be guaranteed or demonstrated without integrated testing.

Internal research has shown that it is not possible to define a set of measures that can be universally applied to all UT's institutional systems. Therefore, this guideline contains two sets of measures for testing with personal data.

- The first set contains measures that must be applied to each institutional system.
- The second set contains measures that must be analysed for impact one by one for each institutional system, in collaboration with LISA.

The second set of measures is based on the 'apply or explain' principle. The controller must examine how each measure can be applied. If it is not possible to apply a measure, the holder must explain why the measure cannot be applied. As a final step, the measures taken by the holder must be checked against the basic privacy principles (as described in section 4.5).

4.1 Basic principles for this Guideline

- Even if all proposed measures are carried out, personal data may still be used. By implementing as many measures as possible and subsequently testing the measures taken against the basic privacy principles, the UT believes that it is handling the privacy of data subjects with great care (no extra risks are created for the privacy of the data subjects).
- The UT considers an acceptance environment to be part of a production environment. The acceptance environment serves to verify changes that have to be implemented in the production environment. Therefore, the UT is of the opinion that personal data may be used in an acceptance environment.
- The UT does not consider the analysis of disruptions to production as the testing of an information system. In many cases, fictitious or anonymised data cannot be used for the analysis of disruptions to production. This guideline therefore does not apply to the analysis of production disruptions.
    a. As a measure, a copy of the production environment can be used for the analysis of production disruptions that is only made available during the analysis of the production disruption (e.g. by using virtualisation techniques).
- The UT does not regard the validation of datasets (e.g., the delivery of a salary test file to ADP and checking management reports) as testing an information system. This guideline therefore does not apply to the validation of data sets.

- The UT distinguishes between information systems that work together in chains and stand-alone information systems (e.g. the Internship Monitoring System used by Technical Medicine). For stand-alone systems, it is easier not to use personal data for testing, because, for example, when anonymising personal data from the production environment, dependencies with other information systems do not have to be taken into account.
- The aim of the proposed measures is to be transparent about the way in which the UT handles personal data during testing and to reduce the risk to the privacy of data subjects (Risk = Chance * Impact).
- It is not necessary to set up a development, testing, acceptance and production (OTAP) environment for each institutional system.

4.2 Mandatory measures

The following measures must be applied by the holders of institutional systems for each institutional system.

A. Before carrying out a test, a well thought-out and documented test plan has to be drawn up.
   a. The aim of this measure is to reduce the chance of data breaches due to test errors by thinking carefully about the testing and the test cases in advance.
   b. When designing the test cases, thought should already be given to the risks involved in carrying out a specific test case (privacy by design). This may mean, for instance, that emails sent by an application are captured so they are not sent to the end-user, or that certain functions are removed or changed.

   c. In case you will test with personal data, the data subjects must be informed about this prior to that processing, for example with a privacy statement[5]. It must also be registered in the register of processing activities. You can contact the data protection officer or the privacy contact person[6] of the faculty or service department for this registration.

B. Acceptance environments are included in the register of processing activities and data subjects are informed prior to the processing, for example in a privacy statement. You can contact the data protection officer for assistance.
   a. The aim of this measure is transparency towards data subjects, so they are aware of the use of personal data in acceptance environments.
   b. On the Cyber Safety website you can see what information must be included in a privacy statement. Please contact the data protection officer or the privacy contact person for registration in the register of processing activities.

4.3 Measures that should be applied as much as possible

The measures in this section are based on the 'apply or explain' principle.

[5] General Data Protection Regulation (GDPR) | GDPR definitions | Cyber Safety (utwente.nl)

[6] Contact | Cyber Safety (utwente.nl)

C. Do not use personal data from production environments in development and test environments

   a. The aim of this measure is to reduce the probability and impact of a data breach occurring in one of these environments by not using personal data.
   b. If a development and/or test environment is available for an institutional system, no personal data is used in these environments. This can be achieved by using fictitious or anonymised data.

If measure C is not possible, it should be examined whether the other measures in this chapter can be applied.

D. Make development and/or test environment only available at the time of a test ('on demand')

   a. The aim of this measure is to reduce the risk of a data breach occurring in the development and/or test environment by making this environment only available during the period that tests are carried out.
   b. By using virtualisation techniques, it is possible to make a copy of the production environment available during the testing period only, as a development and/or test environment.

E. Testing in the development and/or test environment with a subset of the personal data of the production environment.

   a. The aim of this measure is to reduce the impact if a data breach occurs in the development and/or test environment by working in this environment with a subset of the personal data of the production environment. As a result, there are therefore fewer data subjects whose personal data is leaked in case of a data breach.
   b. When using a subset in a chain of information systems, the fact that the same subset is used in the entire chain must be taken into account. This is because performing integrated tests is a precondition for the UT.

F. Limit authorisations in the development, test and acceptance environment

   a. The aim of this measure is to reduce the chances of a data breach occurring in an environment by limiting the number of people with access to the environment to those who perform the tests.
   b. It is often not necessary for e.g. developers to have access to an acceptance environment.

4.4 Optional measures that may be examined

G. No testing by the UT (cannot be applied to custom-made software)

   a. The aim of this measure is to reduce the chance and impact of a data breach by not making a development, testing and acceptance environment with personal data available.
   b. The purpose of testing is to be able to make a risk assessment in advance, which will

make testing a risk management measure. If testing is no longer carried out by the UT, other risk measures have to be assessed, for instance, contractual agreements with suppliers. For some information systems, such agreements have already been made with suppliers. In some agreements it is included that the supplier tests the system and that the supplier will resolve any disruptions during upgrades.

i. Pure Portal[7]: The supplier hosts the Pure Portal and ensures that it is tested in a test environment of the supplier. The UT does not have a test or acceptance environment. The supplier installs the upgrades on the production environment and is available to fix errors that occur during upgrades in the production environment.

ii. Microsoft Exchange (email): the supplier will test upgrades in its own test environment. The supplier is available to repair errors that occur during upgrades to the production environment.

4.5 Assessing the measures against the privacy principles

After the measures have been determined, the holders must perform an assessment against the basic privacy principles for each institutional system.[8] The purpose of this assessment is to demonstrate that the holders act in accordance with the GDPR and treat the privacy of data subjects with care.

- Lawfulness, fairness and transparency
    - *The person whose personal data will be processed, must be informed about that processing and his/her rights.*
- Purpose limitation
    - *The personal data may only be processed for a well-defined legitimate purpose and may only be used for other purposes when those purposes are compatible with the initial purpose.*
- Data restriction
    - *Only personal data that is necessary for the intended purpose may be processed.*
- Correctness
    - *Personal data must be up-to-date and correct.*
- Data retention
    - *Personal data may not be kept longer than necessary for the intended purpose.*
- Integrity and confidentiality
    - *Personal data must be protected from unauthorised access and from loss or destruction.*
- Accountability
    - *The Controller must be able to demonstrate compliance with the GDPR.*

4.6 Recording of choices made

To ensure that holder of institutional systems can demonstrate that they take the privacy of data subjects into consideration, they must record what choices are made in the register of processing activities:

- Which measures of this guideline have been applied and how?
- Which measures of this guideline have not been applied, and why have they not been

[7] University of Twente Research Information (utwente.nl)

[8] Loosely translated into: https://autoriteitpersoonsgegevens.nl/sites/default/files/atoms/files/verordening 2016 - 679 final.pd' (GDPR). Chapter 2, Article 5 'Principles relating to the processing of personal data’

applied?

- If you will process personal data, it must be explained how the privacy principles have been applied.

4.7 Implementation of this guideline

Since an impact analysis must be carried out by each holder for each institutional system, and the holders themselves probably do not have all the knowledge and skills to carry out this impact analysis independently, an impact analysis workshop can be requested via the department head of LISA-PD.

During an impact analysis workshop, the measures described in sections '4.2 Mandatory measures', '4.3 Measures that should be applied as much as possible' and ‘4.4 Optional measures that can be investigated', are analysed for impact one by one by all stakeholders. The results of the impact analysis are recorded in the relevant data processing description during this workshop.

The enumeration below is a proposal for the stakeholders who should be involved in an impact analysis workshop:

- System owner or delegated system owner
- Functional manager
- Privacy Contact Person
- Application manager
- Technical administrator
- Contact person or project leader of LISA-PD

After the impact analysis workshop, the holder is responsible for implementing (or having implemented) the measures that were agreed, and it may therefore be necessary to include these as projects in the UT IT project portfolio.

5 APPENDIX 1: PROCEDURE FOR TESTING WITH PERSONAL DATA

5.1 Step 1: Setting up a test plan

Objective

The objective of drawing up a test plan is to:

1. minimise the use of personal data in test environments;
2. reduce the chances of data breaches due to testing errors;
3. think about the testing and the test cases prior to testing.

Who

The holder of the system concerned is responsible for drawing up the test plan.

What

Before carrying out a test, a well thought-out and documented test plan must be drawn up. Part of the test plan is an impact analysis to see if it is possible to test without any personal data.

Impact analysis

The impact analysis examines whether it is possible to test with:

- anonymised data: to create a test set, a copy of the data from the production environment is made. The personal data in this test set is anonymised by special software; or
- fictitious data: the complete test set is designed, no use is made of a copy of the data from the production environment.

It is important that it is properly recorded if and why it might not be possible to implement the measures.

Testing without personal data

If it becomes apparent that testing without personal data is possible, this process must be continued.

Testing with personal data

If it proves impossible to test without personal data, the risks involved in carrying out a specific test case should already be considered when designing test cases (privacy by design). This may mean, for instance, that emails sent by an application are captured so they are not sent to the end-user, or that certain functions are removed or changed.

The test plan should also discuss which of the following measures are to be adopted. The 'apply or explain' principle applies here:

A. Development and/or test environment only available at the time of a test ('on demand’)

   a. The purpose of this measure is to reduce the risk of a data breach in the development and/or test environment by making this environment available only during the period that tests are performed.

   b. By using virtualisation techniques, it is possible to make a copy of the production environment available as development and/or test environment only during the test period.

B. Testing in the development and/or test environment with a subset of the personal data of the production environment

   a. The purpose of this measure is to reduce the impact if a data breach occurs in the development and/or test environment by working in this environment with a subset of the personal data of the production environment. As a result, there are therefore fewer data subjects whose personal data is leaked in the event of a data breach.

   b. When using a subset in a chain of information systems, it must be taken into account that the same subset is used in the entire chain. This is because performing integrated tests is a precondition for the UT.

C. Limit authorisations in the development, testing and acceptance environment.

   a. The purpose of this measure is to reduce the risk of a data breach in an environment by limiting the number of people who have access to the environment to those who are performing the tests.

   b. It is often not necessary for e.g. developers to have access to an acceptance environment.

When testing with personal data, step 2 follows.

5.2 Step 2: Register data processing in the UT register of processing activities and draft/amend privacy statement

Objective

The purpose of drawing up or adapting a privacy statement is transparency towards data subjects so they are aware of the use of personal data in acceptance environments. Furthermore, the processing activity must be registered in the register of processing activities. You can contact the data protection officer or privacy contact person for assistance.

Who

The holder of the system in question is responsible for drawing up/amending the privacy statement and registration in the register of processing activities. The data protection officer can provide you with a template privacy statement and a document for the register of processing activities.

The holder can ask the privacy contact person (PCP) of their service department/faculty for assistance.

What

Acceptance environments must be included in the register of processing activities. Data subjects must also be informed about the processing activity with a privacy statement. The data protection officer can provide a template privacy statement and a document for the register of processing activities.