Bundesministerium
des Innern, für Bau
und Heimat

Digital Gipfel
Dortmund 2019

Rolf Schwartmann / Steffen Weiß (Ed.)

Draft for a Code of Conduct on the use of GDPR compliant pseudonymisation

A working paper of the Data Protection Focus Group of the Platform Security, Protection and Trust for Society and Business at the Digital Summit 2019

Rolf Schwartmann / Steffen Weiß (Ed.)

Draft for a Code of Conduct on the use of GDPR compliant pseudonymisation of the Data Protection Focus Group of the Platform Security, Protection and Trust for Society and Business at the Digital Summit 2019

Management:
Prof. Dr. Rolf Schwartman
Cologne Research Centre for Media Law - TH Köln

Sherpa:
Steffen Weiß, LL.M.
German Association for Data Protection and Data Security (GDD e.V.)

Members:
Patrick von Braunmühl
Bundesdruckerei GmbH

Susanne Dehmel
German Association for Information Technology, Telecommunications and New Media (BITKOM e.V.)

Philipp Ehmann
Association of the German Internet Industry (eco e.V.)

Maximilian Hermann
Cologne Research Centre for Media Law - TH Köln

Dr. Detlef Houdeau
Infineon Technologies AG

Angelika Hüsch-Schneider
Deutsche Telekom AG (DTAG)

Frank Ingenrieth, LL.M.
Selbstregulierung Informations-wirtschaft e.V. (SRIW)

Clemens John
United Internet AG

John Landvogt
Federal Commissioner for Data Protection and Freedom of Information (BfDI)

Prof. Dr. Michael Meier
University of Bonn/German Society for Computer Science (GI e.V.)

Robin L. Mühlenbeck
Cologne Research Centre for Media Law - TH Köln

Michael Neuber
Federal Association of Digital Business (BVDW e.V.)

Dr. Frank Niedermeyer
Federal Office for Information Security (BSI)

Jonas Postneek
Federal Office for Information Security (BSI)

Frederick Richter, LL.M.
Data Protection Foundation (Stiftung Datenschutz)

Dr. Sachiko Scheuing
Acxiom Deutschland GmbH

Achim Schlosser
European netID Foundation

Irene Schlünder
Technology and Method Platform (TMF e.V.)

Sebastian Schulz
HÄRTING Attorneys at Law

Dr. Tobias Stadler
Federal Commissioner for Data Protection and Freedom of Information (BfDI)

Dr. Claus D. Ulmer
Deutsche Telekom AG (DTAG)

Dr. Martina Vomhof
German Insurance Association (GDV e.V.)

Benjamin Walczak
Independent Centre for Data Protection Schleswig-Holstein (ULD)

* The Focus Group wants to thank Mr Gary LaFever (Anonos) for contributing to this translation.

Version 1.0, 2019

Author: Data Protection Focus Group of the Digital Summit

Management:
Prof. Dr. Rolf Schwartmann
(TH Köln/GDD)

Kölner Forschungsstelle für Medienrecht

Technology Arts Sciences TH Köln

Sherpa:
Steffen Weiß
GDD e.V.

Heinrich-Böll-Ring 10
53119 Bonn, Germany
Tel.: +49 228 96 96 75 00
info@gdd.de

Gesellschaft für Datenschutz und Datensicherheit e.V.

PREFACE

On the Digital Summit 2019, the Data Protection Focus Group of Platform 9 "Security, Protection and Trust for Society and Business" set itself the task of drafting a Code of Conduct for the pseudonymisation of personal data. The pseudonymisation has various references to this year's summit theme in the form of digital platforms and the platform economy. Platforms have enormous amounts of data that can be used, for example, to develop and implement AI applications. At the same time, data can be used to create individual profiles of users. Pseudonymisation can make a functional contribution to ensuring that the personal rights of users are protected when operating digital platforms and that they are protected from individualised profiling.

A Code of Conduct for pseudonymisation gives platform operators the opportunity to carry out pseudonymisation based on transparent guidelines. Users benefit from the application of uniform standards. The referenced application examples provide an insight into further areas in which pseudonymisation can play a role. This document does not constitute a final Code of Conduct. In addition to approval by a Data Protection Authority, this requires the definition of processes for monitoring compliance with the Code. The existing application examples will also be expanded to include sector-specific good practices. This is because, in practice, we need a deeper understanding of pseudonymisation in order to be able to determine a suitable pseudonymisation method and to understand its implementation. This will be done in a later version of the Code.

All contributors deserve our heartfelt thanks for their continuous work in the Focus Group. I especially thank Mr. Steffen Weiß from the Gesellschaft für Datenschutz und Datensicherheit (German Association for Data Protection and Data Security) for the coordination of the work.

Cologne, October 2019

Professor Dr. Rolf Schwartmann

Head of the Data Protection Focus Group of the Security, Protection and Trust for Society and Business Platform at the Digital Summit 2019 and member of the Federal Government's Data Ethics Commission

Content

Preface. 5

1. Introduction 8
1.1. Scope of application 8
1.2. Definitions of CoC terms 9

2. Process specifications for the use and operation of a pseudonymisation. 9
2.1. Organisational questions 9
2.1.1. Designation of the person responsible for the entire process 9
2.1.2. Assemssment and documentation of the criteria necessary to determine the pseudonymisation method 10
2.1.3. Risk-adequate concept for rights and roles 15
2.1.4. Definition of guidelines for re-identification 17
2.1.5. Fulfilment of information and notification obligations towards data subjects 17
2.1.6. Unintentional/unlawful reversal of a pseudonymisation 18
2.1.7. Definition of a regular review process concerning the necessity of processing 18
2.1.8. Notification obligations to supervisory authorities in special cases 19
2.1.9. Documentation and regular evaluation of the process of the considerations made and the measures actually taken 19
2.2. Technical questions 20
2.2.1. General requirements for pseudonymisation 20
2.2.2. General requirements for identifiers (IDs). 21
2.2.3. Calculation method 21

3. Application examples of pseudonymisation 23

1. Introduction

The aim of this Code of Conduct (CoC) is to describe specific rules of conduct for pseudonymisation in conformity with data protection requirements in accordance with Art. 40 para. 2 lit. d of the General Data Protection Regulation (GDPR).

Pseudonymisation protects data subjects from unwanted identification and is an implementation of the principle of data minimisation from Art. 5 para. 1 lit. b GDPR. It constitutes a technical and organisational protection measure in accordance with Art. 25, 32 GDPR. Nevertheless, it also influences the lawfulness of the processing of personal data, as Art. 6 para. 4 lit. e GDPR shows. It thus fulfils both a protective and an enabling function. According to its legal definition, pseudonymisation is characterised by the fact that personal data are processed in such a way that these data can no longer be attributed to a specific person without additional information (cf. Art. 4 No. 7 GDPR).

Even though a direct personal reference is possible within the scope of a pseudonymisation but must be prevented by means of technical or organisational measures apart from a desired disclosure. The GDPR does not contain any technical or organisational information on how a pseudonym can be created, nor does it provide information on possible protective measures regarding the created pseudonym.
For this purpose, this Code of Conduct defines both procedural as well as organisational and technical requirements, which enable both controllers and processors to implement the pseudonymisation in a practical way.

1.1. Scope of application

This CoC applies to controllers or processors regardless of their industry or sector if they pseudonymise personal data themselves in accordance with the requirements of the GDPR or are responsible for the use of pseudonymisation of personal data. The CoC's statements apply independently of the internal organisational and task distribution of the controller or processor.

Controllers or processors who use pseudonymised data in their services or products may join this CoC in order to prove that the pseudonyms used were created in accordance with the rules defined herein.

As a rule, controllers and processors will carry out data processing that relates to pseudonymisation as well as data processing that is in no way related to pseudonymisation. Even if data processing takes place in connection with pseudonymisation, it is to be assumed that not all data processing is subject to the GDPR or is to be subject to this CoC, especially in the case of internationally active controllers or processors. In this respect, controllers and processors can decide for themselves which pseudonymisation processes are to be subjected to this CoC. In the case of those products, services or other data processing that fall back on pseudonyms that originate from pseudonymisation processes that were subject to this CoC, this fact must be pointed out transparently.

1.2. Definitions of CoC terms

- **Pseudonymisation** means pseudonymisation in the sense of Art. 4 No. 5 GDPR: 'pseudonymisation' [means] the processing of personal data in such a manner that the personal data can no longer be attributed to a specific data subject without the use of additional information, provided that such additional information is kept separately and is subject to technical and organisational measures to ensure that the personal data are not attributed to an identified or identifiable natural person.

- According to Art. 4 No. 5 GDPR, the **additional information** is the only information with which the connection of a pseudonym to the person represented can be established. Depending on the pseudonymisation method, the additional information can be a direct assignment or an assignment rule.

- A **pseudonym** is a string of characters that replaces a person's identity data and thus represents that person.

- The **pseudonymisation method** describes the technical-organisational process by which a pseudonym is generated.

- **Specialist managers** are all persons or departments within a company or a public body who are not responsible for the organisation of the entire processing activity, but who only design individual sub-areas in compliance with data protection regulations (such as the proper pseudonymisation of personal data).

- The **Specialist Responsible for Pseudonymisation** (SRP) are all persons or departments within a company or a public body who are responsible for the design of the pseudonymisation process in accordance with data protection regulations, at least in the form of a supervisory and advisory function.

2. Process specifications for the use and operation of pseudonymisation

2.1. Organisational questions

2.1.1. Designate the person responsible for the entire process

From an organisational point of view, the controller or the processor must appoint a Specialist Responsible for Pseudonymisation (SRP). The responsibilities and duties of the controller laid down in the GDPR are not transferred. This SRP coordinates the individual organisational responsibilities before, during and after the pseudonymisation process.

Explanation: Here, the term SRP does not mean the controller in the sense of the GDPR, but (untechnically) the person internally responsible for the organisation and the proper process of pseudonymisation. The pseudonymisation of personal data is usually part of a more general processing activity (according to the record of processing activities).

The SRP can also take on other specialist responsibilities or take overall responsibility for the respective data processing. In any case, the responsibility of this person or department is to be documented as SRP regardless of other responsibilities. The appointment of a data protection officer as SRP is not permitted.

Explanation: The SRP is not to be equated with the data protection officer of the controller or processor. In contrast to the SRP, the data protection officer is not responsible for the lawfulness of data processing. His/her legal duties are defined in Art. 39 GDPR and are characterised by giving advice and to monitor. In the area of pseudonymisation, the data protection officer can advise on the planning and implementation of the pseudonymisation and also monitor compliance with the legal requirements for pseudonymisation and this CoC. Due to the allocation of organisational responsibility for the SRP, an identity of the data protection officer and the SRP would not be compatible with the legal requirements.

The SRP must possess the technical and organisational expertise required for pseudonymisation. If a department has been designated as SRP, the department must have the necessary specialist knowledge in (partial) totality if and to the extent that it is ensured from an organisational point of view that this department always exercises responsibility in an appropriate (partial) totality.

2.1.2. Assemssment and documentation of the criteria necessary to determine the pseudonymisation method

For the legally compliant use of pseudonymisation, the following criteria must be considered in documented form.

2.1.2.1. Type and risk class of personal data processed

The type and risk class of the data processed must be specified in order to ensure pseudonymisation in conformity with data protection regulations. Based on this risk assessment, the selection of the adequate, GDPR compliant pseudonymisation procedure must take place.

Explanation: In principle, there can be different categories of data:

- Personal data pursuant to Art. 4 No. 1 GDPR
- Special categories of personal data pursuant to Art. 9 para. 1 GDPR

The categories of data processed can be found in the record of processing activities.

Within the framework of the risk assessment of the processed data, assessments from risk analyses or a data protection impact assessment can also be applied.

The data category used does not represent a suitable criterion for a risk assessment in itself and can at best be used as an indication. Rather, other aspects must also be taken into account within the framework of risk assessment. This is for example

- the purpose and context of the processing (see below 2.1.2.2. and 2.1.2.3.); for example, identical personal data may be used in the context of contract performance or to track user activities;
- the category of data subjects; e.g. children or members of certain population groups without immediately triggering the scope of application of Art. 9 (1) GDPR;
- the number of persons concerned (see 2.1.2.4 below) or the combination of the different data categories.

2.1.2.2. Intended processing purposes

The purposes for which the data are to be processed must be specified.

Explanation: There may be more than one purpose of processing. Purposes cannot easily be changed in the aftermath of data collection, so that these should be documented as comprehensively as possible. However, purposes must also be sufficiently precise to allow the purpose limitation principle to be respected. Examples for a processing purpose may include data processing for billing purposes, for checking the network utilisation of a mobile phone provider, for product development purposes or for the processing of data for research purposes. Research purposes should be specified in the documentation to the extent that the research context or the research objective can actually be comprehended in terms of whether an actual, future processing is subject to the intended research purpose and therefore a risk assessment can also be adequately derived. The description of the purpose also has an influence on the assessment under data protection law as to whether data processing for the intended purposes still falls within the scope of the relevant statutory provision and, on the other hand, it must be examined whether pseudonymisation changes this assessment in some way.

2.1.2.3. Context of pseudonymisation

The context of pseudonymisation shall be documented.

Explanation: The context of processing refers to the legal context for pseudonymisation. Pseudonymisation can be used, for example, in the course of its enabling function within the framework of Art. 6 para. 1 lit. f and Art. 6 para. 4 GDPR or as purely technical and organisational measures pursuant to Art. 32 GDPR or within the framework of Art. 25 GDPR.

Documentation is necessary because this context also influences the choice of the appropriate pseudonymisation procedure.

2.1.2.4. Expected number of processed records

It shall be checked and documented how many records will be processed.

Explanation: There must be an overview over whether only a few data sets or a large number of data are pseudonymised. When checking the number of data records to be processed, it is relevant whether the data records are static or dynamic, i.e. whether they are a fixed number of data that is pseudonymised or whether the data record is continuously enriched with further data. Classical list procedures for pseudonymisation using tables are for example not suitable for a large amount of data.

2.1.2.5. Suitable pseudonymisation types

The different types of pseudonyms required shall be documented.

Explanation: Different types of pseudonyms, for example, are particularly suitable for certain purposes, although they may be completely unsuitable for other purposes.

A distinction can be made between the following types of pseudonyms:

- Personal pseudonyms that replace identity data such as name, ID number or mobile phone number
- Role pseudonyms where one or more persons are assigned to a pseudonym (e.g. IP number)
- Relationship pseudonyms where a person uses a different pseudonym for each (communication) relationship, e.g. different nicknames, role relationship
- Role-relationship pseudonyms that are a combination of the two pseudonym types
- Changing pseudonyms where, for example, a new pseudonym is used for each transaction or each entry. Used, for example, in online banking

Considering the purpose and context of the processing, those types of pseudonyms are to be preferred which are suitable for the respective purpose and at the same time protect the persons concerned as far as possible against unwanted identification. The SRP supports the selection of the appropriate type of pseudonymisation. The weighing carried out for the decision for or against a relevant type of pseudonymisation must be documented.

Explanation: In general, the risk of reversal of personal pseudonyms is higher than that of role or relation pseudonyms. This is related to the connection of a pseudonym with a person standing behind it. Depending on the purpose and context of the processing, the use of personal pseudonyms may be necessary. On the other hand, there is a lower risk of reversal of persons with role-relationship pseudonyms and changing pseudonyms than with the abovementioned person pseudonyms.

2.1.2.6. Determination of the appropriate pseudonymisation method and the time of pseudonymisation

Different methods are available for pseudonymisation[1].

The strength of the applied method must be examined, determined and documented taking into account all objective factors, risks to the rights and freedoms of the parties concerned as well as the costs of identification and the time required for this when using the technologies available at the time of processing as well as foreseeable technological developments. When using calculation methods, a state-of-the-art transformation procedure must be used (for technical requirements, see 2.2.1.).

Furthermore, pseudonymisation procedures shall be designed in such a way that simple and efficient selection and deletion of the data is possible, insofar as the processing purpose no longer exists or the legal basis for the processing is no longer applicable.

Explanation: The principle of data minimisation must always be observed. The principles of privacy by design must also be taken into account. As a result, the technical design must provide the appropriate framework conditions from the beginning. Compliance with these principles thus avoids the per se inadmissible continuous storage of pseudonymised data that is difficult to reidentify. In addition, such pseudonymisation methods are to be preferred which simply enable the subsequent anonymisation of data.

The pseudonymisation has to be made in the processing process as early as possible.

Explanation: Personal data must be limited to what is necessary for the purposes of processing (cf. Art. 5 para. 1 lit. c GDPR). If the pseudonymisation has been identified as suitable processing by the controller or processor, its technical implementation should be carried out promptly. Likewise, pseudonymisation should also be carried out as early as possible in multi-stage data processing, especially if non-pseudonymised data are not required at the upstream processing stages.

2.1.2.7. Planned disclosure of pseudonymised data

It must be documented whether pseudonymised data are to be transmitted to third parties. The data controller or processor must take appropriate measures to ensure that the data passed on is only processed by the recipient(s) for the purposes specified beforehand. The controller or processor shall ensure that

[1] Schwartmann/Weiß (Ed.), Requirements for the use of pseudonymisation solutions in compliance with data protection regulations - a working paper of the Data Protection Focus Group of the Platform Security, Protection and Trust for Society and Economy in the context of the Digital Summit 2018, D.2.2 ff.

the transfer of the pseudonymised data to the recipient is covered by a legal basis. In addition, the controller or processor must take appropriate measures to prevent the recipient from inadmissibly reidentifying data subjects.

Explanation: Since pseudonymised data also have a personal reference, the general data protection requirements apply to the processing, including the purpose limitation pursuant to Art. 5 para. 1 lit. b GDPR. The data provider as well as the recipient must agree on a purpose before passing on pseudonymised data. The purpose of the processing can be confirmed by the recipient in text form or in writing (e.g. as part of a contract) as an appropriate measure. Since only the pseudonymised transfer of data falls within the scope of this CoC, the identification of data subjects within the scope of data transfer must be omitted. The CoC therefore formulates an obligation for the data provider to take appropriate security measures in this regard before passing on the data to the recipient. This includes, for example, an obligation to check on the part of the receiver with regard to obvious identification possibilities, since a detailed knowledge of the possibilities of linking the data on the receiver side cannot be assumed. The controller or processor should in any case obtain supplementary confirmation in text form or in writing (e.g. as part of a contract) that an identification does not take place by the recipient.

Insofar as an audit has shown that the recipient could obviously carry out a re-identification, appropriate supplementary protective measures should be implemented as far as this appears necessary due to the expected risks for the data subjects.

Regarding the legal basis for disclosure, in particular in cases where consent has been obtained for the collection of personal data, the fact of disclosure to another body in pseudonymised form must be covered. If not, another legal basis would be required.

2.1.2.8. Planned processing of pseudonymised data in the third country

It shall be documented whether pseudonymised data are to be processed outside the EEA. In the event of the transfer of personal data to a third country, the data controller or processor must ensure that the requirements of Chapter V of the GDPR regarding the guarantee of an appropriate level of data protection are met. When collecting personal data, the data subject must be made aware of the transfer of such data to a third country, even if pseudonymised data are transferred.

Explanation: The GDPR places special requirements on the processing of personal data in a third country outside the EU or the EEA. These requirements are regulated in Chapter V of the GDPR and include, for example,

the transfer of personal data on the basis of an adequacy decision by the EU Commission or other suitable guarantees according to Art. 46 GDPR. The fact that the data to be transfersed are pseudonymised shall not exempt the data exporting body from complying with the requirements of Chapter V. After all, a data subject can also be re-identified in a third country using the key to pseudonymisation.

2.1.2.9. Planned/ foreseeable frequency of re-identification?

The planned or foreseeable frequency of re-identification of data subjects shall be specified in documented form.

Explanation: The chosen processing purposes have an influence on the question of whether a re-identification of the persons concerned must be carried out promptly and in the short term. In the area of network monitoring, for example, it may be necessary to identify a workstation infected with malicious code on the basis of pseudonyms at short notice.

The planned or expected frequency of re-identification of data sets shall be defined. It must also be documented for which planned or expected reasons or for which purposes such re-identification will take place (e.g. to safeguard the rights of the data subjects). In addition to the reasons and purposes, it must also be documented what delay of tolerance exists in the event of re-identification, i.e. what the maximum delay may be until sufficient re-identification of a data set.

Explanation: Pseudonymisation methods differ, among other things, in their efficiency and manageability with regard to the re-identifications that have to be carried out. Similarly, the frequency of expected re-identifications also interacts with an appropriate allocation of functions as defined in Section 2.1.3. The documentation to be prepared here should enable the SRP to create a binding basis for consideration. On the other hand, it should enable the SRP to evaluate the hypothesis and planning presented here over time. Such an evaluation would have to take into account, for example, whether a possibly very high, expected re-identification rate actually occurs in practice. It would also be necessary to consider, for example, whether the delay tolerance could be adhered to or whether other pseudonymisation methods are now also able to adhere to these tolerances due to new technical developments.

2.1.3. Risk-adequate concept for rights and roles

Regarding access to pseudonymised data and the combinations thereof required for the respective activity, possible existing translation tables and keys for the re-identification of a person and other information, an appropriate rights and role concept shall be provided. The more sensitive the processed data or the higher the expected risks for the rights and freedoms of the data subjects, the more effective such a separation should be.

Explanation: According to the legal definition of pseudonymisation, additional information that enables the identification of data subjects must be kept separately and identification must be prevented by technical and organisational measures. An existing rights and role concept can represent such a technical-organisational measure. Depending on the risk of the data and the context of the processing, different models are suitable within such a rights and roles concept:

"All-in-one-hand"-model: Here, the controller or processor has both the pseudonymised data and the possibility at any time to reverse the processed pseudonyms or to re-identify the data subject. A possibility of re-identification can be asigned to a person, a department or a legal entity. In these cases, at least internal requirements should exist, which would result in permissible and impermissible circumstances for carrying out a re-identification as well as possible documentation obligations regarding re-identifications. As the expected risks increase, these internal requirements should also be supplemented by an appropriate internal rights and role concept based on the need-to-know principle (cf. mixed models).

Trustee model: In the classic trustee model, the trustee is a legal entity outside the controller or processor acting as a "third party". It is therefore a trust center that is independent of data collection and usage in terms of location and organisation. A trustee can, for example, be entrusted with the storage of keys for the re-identification of data subjects. The processing of pseudonyms by the third party is also a possibility, while any keys and raw data remain with the controller or processor.

Key management is the most common scenario in which a trustee can be involved in various ways. The method chosen within the trustee model should always be based on the documented risks for the data subjects:

- Ex ante: The trustee shall re-identify the data subjects for purposes or circumstances defined prior to the commencement of processing.
- Ad hoc: The trustee re-identifies the data subjects on the basis of previously defined consideration criteria, but not according to previously defined purposes and circumstances.
- Ex post: The trustee is informed of any re-identifications that have taken place, together with the reason (e.g. as an individual case or via statistics). The trustee can evaluate this information and take appropriate measures based on it, e.g. training or disciplinary measures.

Mixed models: Mixed models are also conceivable. Here, for example, the separation of the information necessary for re-identification can also take place within the organisation of the controller or processor, in which the information is subjected to a rights and role concept.

This can also include, for example, distributing information across several hierarchical levels or independent departments. The departments usually responsible for such issues (internal audit or compliance or legal department, (IT) security or data protection officer) could also be suitable for this purpose anyway. Particularly in large organisations, the establishment of a trustworthy "third party" of its own, which offers the separate administration of data and/or secrets or keys internally, is also an option.

Such mixed models are conceivable, for example, particularly in cases where the processing comprises several processing steps and several pseudonymisation stages, for each of which different risks to the data subjects have been documented.

2.1.4. Definition of guidelines for re-identification

In the event that a re-identification of data subjects on the basis of the pseudonymised data is planned, the following requirements must be observed and documented. The SRP supports:

1 In the case of pseudonymisation as a simple protective measure, no permission to trace pseudonyms back to individual is required beyond the original legitimation for data processing. The reversaö is covered by the original purpose of use.

2 In the case of pseudonymisation to enable the further processing of data in accordance with Art. 6 para. 4 GDPR, the following applies:

- In cases where the data subject has an overriding interest in being re-identified (e.g. for the purpose of information or an opportunity to object), the admissibility must be examined in relation to the data processed (Art. 6 or Art. 9 GDPR).
- In cases where it cannot be established whether the data subject has an interest in being re-identified, consent to re-identification must be obtained. This does not apply to re-identification based on a legal permit.
- In cases where the person responsible has an overriding interest in being re-identified (e.g. for the purpose of providing information), the admissibility data must be asessed (Art. 6 para. 4 GDPR)

3 In cases where a dynamic data set (cf. 2.1.2.4) is pseudonymised, it must be checked at regular intervals whether this dynamic makes it possible to re-identify data subjects. In the event of the possibility of re-identification, the provisions of paragraphs 1 and 2 shall apply.

2.1.5. Fulfilment of information and notification obligations towards data subjects

If the pseudonymisation is only used as a technical-organisational measure with protective function, no separate information beyond the general information about the data processing is required. If further processing is to be

carried out for compatible purposes, the following two purposes have to be distinguished:

1 The compatible further processing in accordance with Art. 6 (4) GDPR is intended from the outset - in which case the information should be provided directly in the data protection information.
2 Compatible further processing will only be decided at a later point in time - then information of the data subjects in accordance with Art. 13 para. 3 GDPR is required at this point in time.
The information and notification obligations towards the data subjects also relate to the right to object or in a consent scenario.

2.1.6. Unintentional/unlawful reversal of a pseudonymisation

In the event of an unintentional or unlawful reversal of a pseudonymisation, a response plan must be drawn up. The SRP supports. The response plan shall include the following elements:

- Risk assessment for data subjects
- Measures to prevent/control the risk
- Evaluation of a notification obligation according to Art. 33/Art. 34 GDPR
- Notification to the supervisory authority and the data subjects in case of the existence of an obligation to notify.

The response plan can be integrated into an existing process (for example, Incident Response Plan) at the controller or processor.

Explanation: According to recital 85 sentence 1, the reversal of a pseudonymisation can constitute a data breach which, in the event of a risk associated with the breach for the data subjects concerned, must be reported to a supervisory authority or, in the event of a probable high risk, also to the data subjects. Controllers and processors should therefore document any necessary steps in a response plan in the event that a pseudonymisation is reversed. The response plan does not have to be created separately for the pseudonymisation, but can generally exist for data protection incidents at the controller or processor, but must explicitly address the reversal of a pseudonymisation.

2.1.7. Definition of a process for the regular review of the requirement of processing

The intervals shall be defined and documented at which the necessity of processing of pseudonymised data has to be assessed. The SRP provides advice and support in this regard. Such a review should in general take place at least every two years. The assessment shall be documented. If, in the course of this review, it is determined that processing is no longer necessary, the pseudonymised data must be deleted or made anonymous in accordance with data protection regulations.

Explanation: Since pseudonymised data make it possible to re-identify data subjects, such processing activity is also subject to the principle of storage limitation under Art. 5 para. 1 lit. e GDPR. If pseudonymised data are no longer required for the specified purpose of processing, they must be deleted. Consequently, it is necessary to establish a regular cycle for an assessment of necessity by the controller or processor in order to determine the necessity of the processing.

2.1.8. Notification obligations to supervisory authorities in special cases

If, despite a pseudonymisation, a high risk for rights and freedoms of data subjects[2] can still be identified within the scope of a processing activity and if pseudonymisation is the only protective measure, the competent supervisory authority pursuant to Art. 36 GDPR must be consulted. In such circumstances, the SRP must be consulted.

Explanation: Controllers must consult the supervisory authority in advance of any processing if it emerges from a data protection impact assessment pursuant to Art. 35 GDPR and when the processing would pose a high risk to the data subjects, unless measures are taken to contain it. If there is a high risk for those data subjects and pseudonymisation is the only protective measure, there is a legal obligation to consult the competent supervisory authority.

2.1.9. Documentation and regular evaluation of the process, the considerations made, and the measures actually taken

For each section of Chapter 2.1, the measures taken as well as the influencing factors for determining an appropriate pseudonymisation method (Section 2.1.2) are to be documented.

Insofar as the determination of the measures taken is to be preceded by an assessment, such assessments shall also be documented.

The documentation shall be prepared by the SRP. The SRP can, however, fall back on documentation from other technical experts and third parties. Here it must be ensured that modifications of the documentation are exclusively transparent; in particular regarding the aspects "what", "by whom" and "when".

2.1.9.1. Documentation of processes and other measures taken

Processes and measures taken shall be documented in such a way that

1. the SRP is capable
   - to evaluate the process or measure in terms of effectiveness;
   - to verify the implementation of the processes or the measures taken;
   - to evaluate compliance with the processes or measures taken, as well as,
2. the SRP and all persons entrusted with implementation are able to
   - understand the process or the measure and to implement it according to the defined specifications.

2 Information on risk determination can be found, for example, in short paper No. 18 of the Conference of Independent Data Protection Commissioners of the Federal Government and the Länder (of Germany) or in Working Paper 248 of the Article 29 Working Party.

2.1.9.2. Documentation of considerations

Considerations must be documented, including a statement of reasons. It must be ensured that the conclusions reached within the framework of the consideration - e.g. determination of the appropriate pseudonymisation method or an applied risk classification - can also be easily understood by third parties. These considerations shall be reviewed regularly, in particular regarding the state of the art and conformity with the intended purpose, and these reviews shall also be documented. References to further documentation are permissible, as far as it concerns referenced methods or correspondingly documented consideration results of this section and the reference shows the concrete title, storage or storage location and version of the referenced document.

Explanation: The documentation to be prepared in accordance with this section fulfils a number of objectives. The documentation forces the controller or processor to systematically process the requirements of this Code. Insofar as the SRP makes use of the services of other specialist managers, the SRP shall have an information base which is always comprehensible also for him/herself. This documentation also enables the SRP to review the original assumptions on a regular basis and adjust them if necessary. Such an evaluation is necessary to the extent that the GDPR requires processing in accordance with the current state of technology. It is therefore likely that measures taken or considerations made on the basis of documented information will have to be modified as the technical status quo progresses. The documentation also enables both the SRP and any compliance departments to carry out conformity checks.

2.2. Technical questions

221. General requirements for pseudonymisation

The technical implementation takes place only in consultation with the SRP. The SRP shall consult the specialist managers when selecting and evaluating the appropriate pseudonymisation method. The specialist managers must also consult the SRP on planned changes to the technical implementation.

For the implementation of a pseudonymisation different procedures can be used. For example, an allocation table can be used in which one or more pseudonyms are allocated to each date in plain text. Alternatively, various cryptographic methods can be used for pseudonymisation, each of which converts a plain text date into one or more pseudonyms. The reversibility of pseudonymisation can be controlled/restricted here by establishing access controls concerning used cryptographic keys and, if necessary, other parameters.

When selecting the pseudonymisation to be used, the test steps of the inventory (in particular, subsections 2.1.2.1 to 2.1.2.6) must be followed.

222. General requirements for Identifiers (IDs)

Regardless of the other requirements, an ID must be used as a pseudonym that does not allow any conclusions to be drawn about the input data or the natural person concerned.

Application scenarios and challenges:

- When pseudonymising data, it must be ensured that the ID used cannot be re-identified if individual information in the data set is viewed in context with other data.

Explanation: The postal code is used as the ID; the data also contains individual information about the date of birth. With a sufficiently small number of data sets, the natural person can be re-identified by comparing all data sets with identical birth data.

- If IDs are generated on the basis of the combination of individual information in the data records under consideration, it must be ensured that a direct comparison of the output data with the input data or knowledge of the scheme used does not lead to re-identification by simple means. This can be achieved, for example, by adding a secret key ("salt") to the calculation of pseudonyms.
- Preference should be given to methods which do not allow any conclusions to be drawn about the sorting of the data or the sorting of the data before the methods are applied must be sufficiently random.

Explanation: Pseudonymised data could be re-identified by an easily understandable chronological or alphabetical sequence.

With regard to the technical procedures used, the relevant current technical guidelines of a general nature must also be taken into account, in particular the relevant guidelines of the BSI ("TR-02102 Cryptographic procedures: Recommendations and key lengths") if, for example, procedures are used that use hash functions as a basis. Transformation procedures used for pseudonymisation must also be replaced by current procedures - especially for pseudonymised data used over a long period of time - in order to guarantee a maximum of security.

223. Calculation method

The choice of the specific pseudonymisation method must be based on the inventory and coordinated with the SRP; accordingly, the technical implementation is also subject to regular evaluation, cf. 2.1.9.2.

When using calculation methods to determine pseudonyms (in particular for pseudonymous users), it must be ensured that these have the following properties:

1. They must be based on state-of-the-art secure cryptographic methods.

Explanation: Software to create pseudonyms should use available crypto libraries instead of reimplementing the algorithms. This is why Open Source implementations are useful.

2. For the given plain text space (e.g. the set of all user IDs or names or telephone numbers) the function pseudonym = f(plaintext ID) must be unique, i.e. different pseudonyms must result for different plain text keys in order to avoid homonym errors.

Explanation: A homonym error occurs if identity data of different persons falsely lead to the same pseudonyms.

3. The inverse function plaintext ID = g(pseudonym) must not be calculable with reasonable effort.

Explanation: The reasonable should also be determined on the basis of the specific circumstances. In particular, the value of the re-identified data for unauthorised parties should be considered. The risk analysis carried out can be used for this purpose. This information is important for the determination of the reasonable effort, as it allows conclusions to be drawn about the expected technical and professional resources of unauthorised parties: The higher the value of the data, the greater the effort that can be justified from the point of view of unauthorised parties.

4. Similar, especially consecutive plaintext IDs must not lead to similar pseudonyms, small changes to plaintext IDs must lead to completely different pseudonyms in order to make it more difficult to "guess" plaintext IDs.

5. The security of pseudonymisation must not be achieved by keeping the algorithm secret, but by using a secret key.

6. From the knowledge of a pair (plaintext ID/pseudonym) it must not be possible to deduce the secret with reasonable effort.

7. The recommendation to carry out the pseudonymisation with the aid of a cryptographic hash function or a symmetrical block cipher procedure in which, in addition to the plaintext IDs, a secret, consistent key is used whose entropy is at least 100 bits results from points 1.-6. Entropy is a measure of the indeterminacy of a character string (e.g. ten independent coin tosses (head/tails) provide ten bits of entropy). If a hash function is used, the minimum length of the hash value shall result from the requirement in point 3.

3. Application examples of pseudonymisation

3.1. Pseudonymisation Magenta TV (DTAG)

3.1.1. Introduction

Deutsche Telekom generates anonymous statistics based on the use of the Magenta TV product. Personal data is first pseudonymised in order to convert it into anonymous statistics. Certain usage data, so-called events, which are provided with an identifier (ID), are used in particular for pseudonymisation. This makes it possible, for example, to carry out a different count. This means that the question can be answered, as to how many households or how many set-top boxes have watched a particular channel at a certain time. Every user has the possibility to object to this processing (opt-out) at any time.

The abovementioned IDs are ultimately no longer present in anonymous statistics, making it impossible to trace back from the pure numbers to the encrypted IDs.

3.1.2. Description of responsibilities

Telekom Deutschland GmbH is responsible for the personal data generated when using the Magenta TV product. The pseudonymisation is provided by T-Systems GmbH as an IT service provider. T-Systems will be integrated by Telekom Deutschland in this process via a controller-processor agreement. Another legal unit of T-systems, the Tel-IT, provides an automatically generated key for pseudonymisation. Tel-IT is also involved in development and operation.

The assignment of pseudonymisation is carried out by the "Private Customers Germany" segment. In other words, this division commissions the pseudonymisation of the IT service provider, after consultation with Deutsche Telekom's Corporate Data Protection Department. The Data Protection Department is also responsible for the legal conformity of the pseudonymisation process as such.

3.1.3. Criteria for determining the appropriate pseudonymisation method

The data types to be pseudonymised are usage data from Magenta TV, cf. Art. 4 No. 1 GDPR. In addition, there is also metadata, which also flows into the pseudonymisation. These are pseudonymised for the creation of user profiles, cf. Art. 6 para. 1 lit. f) in connection with Art. 32 para. 1 lit. a) GDPR. This involves more than 10 million data records per day, which are pseudonymised. Personal and device pseudonyms are created for pseudonymisation.

Data field NAME	IDENTIFICATION	Risk Class	Remarks
Subscriber_ID	ACCOUNT_ID	1	Pseudonym Subscriber
Physical_Device_ID	DEVICE_ID	2	Pseudonym Device identifier

Fig.: Data fields and risk classes

3.1.4. Rights and role concept as well as key management

The authorisations are clearly distributed both by role assignment and technical purpose assignment, which is laid down in the organisation and authorisation concept. The division of Telekom Deutschland GmbH (TDG), which is responsible for the product Magenta TV, has no influence on the pseudonymisation. It only has access to the generated anonymous statistics, which are generated at the end. T-Systems performs pseudonymisation by automatically encrypting the usage data via the AcL (Acquisition Layer).

An independent technical instance (Tel-IT) supplies the key. This system can only be accessed by the Tel-IT and the administrators of T-Systems. The crypto material (keys/salts) required for pseudonymisation is separately encapsulated in a so-called Trust Center (Tel-IT). During configuration, the employee has no way of gaining knowledge of it. Only technical users and a small group of persons (3-4 persons) have access to the cryptomaterial. However, they have no admin rights. This is the organisational separation.

In an additional agreement, TDG also waives its authority to issue instructions regarding the crypto material which it would have according to the controller-processor data processing, i.e. TDG may not request this information. Tel-IT is not allowed to hand them over, not even to third parties. The data is only transferred from the AcL to the BDMP (Big Data Management Platform), where it is available to the TDG for analyses, when the pseudonymisation has been completed. The pseudonymised usage profiles are aggregated on the BDMP. Access to the information in the AcL and to the technical instance is excluded.

3.1.5. Data generation

When using a Magenta TV Set-Top-Box (STB) - i.e. when the user presses the remote control - different events are generated depending on which keys have been pressed and in which context the user is. These STB events form the basis of the evaluations.
Examples for these events are e.g. the switching on/off processes, channel switching, information about the watched channels or information about activities around recording or watching recordings.

These event data records contain, for example, information about the set-top box (=DeviceID), date/time, and other specific data fields. The personally identifiable information of these events is collected by means of an AES128[3] material cipher.

3 Advanced Encryption Standard with a key length of 128 Bit.

3.1.6. Pseudonymisation

The underlying pseudonymisation process leads to linkable but not detectable pseudonyms. These are generated using so-called deterministic, cryptographically strong ciphers. Since deterministic processes map identical plain texts to identical result values (pseudonyms), linkability is ensured. Through the secure administration of the key material and the organisational separation of access to the keys, the inadmissible reversal of pseudonymisation, i.e. the disclosure of the plain date, is prevented.

The pseudonyms created for the AccountID (ID for the customer) and the DeviceID (ID for the respective set-top box) are used for further evaluations. The event files necessary for the evaluation and the references ACCOUNT_PS and DEVICE_PS do not contain any attributes that directly contain personal data. These references (ACCOUNT_PS and DEVICE_PS) are the person and device pseudonyms.

The pseudonyms are used to record the usage information of Magenta TV in order to generate anonymous statistics. Here, it is important to be able to recognise which event is occurring from the same device or user. Pseudonymisation ensures that employees cannot draw any conclusions about the actual devices or users. The resulting statistics are completely exempt from pseudonymised identifiers and are therefore anonymous.

3.2. Pseudonymisation for the optimisation of online platform advertising

3.2.1. Introduction

Targeting advertising at a desired audience via online platforms such as social media, e-commerce shops or online publishers enables the minimisation of advertising dispersion loss. At the same time, targeted advertising saves platform users unnecessary irritation caused by irrelevant video adverts. The use of commercially available consumer information such as sociodemographic or lifestyle data helps to reach relevant audiences. Acxiom licenses target audience formed with selection criteria and uses multiple pseudonymisation methods, so that the data can be linked for the purpose of presenting individualised online advertising, on the one hand, and to protect data subjects from direct identification, on the other.

3.2.2. Preparation: Creation of a pseudonym-to-pseudonym reference table with the platform partner

In order to reach the desired audience online on one platform online, a two-stage process is required. Firstly, and as an independent process from carrying out campaigns for a customer, a data comparison of the databases of Acxiom and the platform operator takes place. In the first step, the plain text name-and-address-database is loaded onto Acxiom's proprietary Privacy Enhancement Tool (PET). There, each data record of the name-and-address-database

receives a pseudonymous personal key. This personal key is again hashed with a Salt. In addition, Acxiom pseudonymises the plaintext data of the name-and-address-database by hashing it. The result is a file with two fields: the hashed personal key and the hashed name-and-address-data (Acxiom's match file). The platform operator on the other side pseudonymises his user data in a similar way and saves the user contact data with the platform's own user ID in a file (match file platform). After comparing the two match files using the pseudonyms or the hash values, a cross reference table is created between the platform user ID and the hashed Acxiom personal key. The platform operator only stores the mapping of the platform user ID to the hashed Acxiom personal key in the cross reference table. All other information is deleted immediately after the comparison.

3.2.3. Audience selection

For the selection of the relevant audience on a platform, Acxiom creates its own data product which contains the pseudonym (an Acxiom personal key associated with a certain salt) as a key variable for matching, but with no names nor addresses.

Audiences can be selected on the basis of sociodemographic data, calculated affinities for certain products or services, but also on the basis of purely geographical information (e.g. advertising for high-speed Internet only in regions where it is available).

Acxiom has a wide range of microgeographic variables that are calculated on a fine-spatial neighborhood level using official data, surveys, market research studies, etc. (i.e. all households in a geographical cell or neighborhood are assigned the same values). For example, the assumption "has a cat" is assigned to all households of this microgeographic cell, regardless of the individual situation of the different families in the neighborhood. The neighborhood must always comprise of at least 4 households[4]. By using these characteristics, the identification or detectability of a natural person is prevented by means of these pseudonymous data sets.

The resulting audience that is selected (e.g. "has a cat" and "lives in an apartment") is always a list of hashed Acxiom personal keys. This is uploaded by Acxiom onto the advertising account of Acxiom at the platform operator, and can then be shared with the advertiser, or its agency, so that they can use the audience.

3.2.4. Placing advertisements

Based on the reference created through data comparison between the platform user ID and the hashed Acxiom personal key, the platform operator displays the advertisement to the audience uploaded to the account.

4 In accordance with the recommendations of the 3rd Geo Progress Report of the Federal Government from October 2012.

4.2.1. Technical and organisational measures

The platform operator has contractually committed to Acxiom to keep the reference data between the platform user ID and the hashed Acxiom personal key separated, and physically set apart from its CRM system. The advertisement is displayed through a separate advertising delivery system. The corresponding contractual obligations and the processing of the data, physically separated from its own user database, ensure that there is no possibility for the platform operator to identify a person on the basis of these pseudonymous IDs.

At Acxiom, access authorisation to the salt, used to encrypt the personal key, is only granted to a few selected employees. The selected ID numbers cannot be detected by the Acxiom employees who select the audience, since the process described above does not allow them to assign the hashed Acxiom personal keys to a person.

In addition, the twice pseudonymised and hashed Acxiom personal keys are anonymous for both the advertisers who license the audience of Acxiom, and for their agency, as well as for any other third party, since they have no means of identifying a person from the twice pseudonymised personal data, nor the ability to assigning this data to an individual. In addition, advertisers have no access whatsoever to the hashed Acxiom personal keys in the selected audience, because the selection of audience and uploading to the platform takes place exclusively at Acxiom. Viewing of the uploaded audience group is technically not possible for the advertiser.

On the Digital Summit 2019, the Data Protection Focus Group of Platform 9 "Security, Protection and Trust for Society and Business" set itself the task of drafting a Code of Conduct for the pseudonymisation of personal data. In times of enormous amounts of data with which applications of artificial intelligence or machine learning can be fed, pseudonymisation can make an important contribution to balancing technological progress and the personality rights of users.