Bundesministerium
des Innern

Digital Gipfel
Ludwigshafen 2017
Rhein-Neckar

Schwartmann / Weiß (Eds.)

White Paper on Pseudonymization Drafted by the Data Protection Focus Group for the Safety, Protection, and Trust Platform for Society and Businesses in Connection with the 2017 Digital Summit

– Guidelines for the legally secure deployment of pseudonymization solutions in compliance with the General Data Protection Regulation –

Schwartmann / Weiß (Eds.)

White Paper on Pseudonymization
Drafted by the Data Protection Focus Group
for the Safety, Protection, and Trust Platform
for Society and Businesses
in Connection with the 2017 Digital Summit

– Guidelines for the legally secure deployment of pseudonymization solutions in compliance with the General Data Protection Regulation –

Digital Gipfel
Ludwigshafen 2017
Rhein-Neckar

Headed by: **Prof. Dr. Rolf Schwartmann**
Research Center for Media Law – Cologne University of Applied Sciences

Sherpa: **Steffen Weiß, LL.M.**
German Association for Data Protection and Data Security

Members: **Prof. Dr. Christoph Bauer**
ePrivacy GmbH

Patrick von Braunmühl
Bundesdruckerei GmbH

Susanne Dehmel
German Association for Information Technology, Telecommunications and New Media

Members: **Walter Ernestus**
German Federal Commissioner for Data Protection and Freedom of Information

Nicolas Goß
Association of the Internet Industry

Michael Herfert
Fraunhofer Society for the Advancement of Applied Research

Serena Holm
SCHUFA Holding AG

Dr. Detlef Houdeau
Infineon Technologies AG

Version 1.0., 2017

Issued by the Digital Summit's data protection focus group

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

Kölner Forschungsstelle für Medienrecht

Technology
Arts Sciences
TH Köln

Sherpa contact:

Steffen Weiß

German Association for
Data Protection and Data Security
Heinrich-Böll-Ring-10
53119 Bonn
Germany +49 228 96 96 75 0
info@gdd.de

GDD
Gesellschaft für Datenschutz
und Datensicherheit e.V.

Foreword

Data is the raw material of the future. It can consist of information about how we live our lives, which can be deployed for medical products, or of information about our driving behavior which can be based on weather forecasts or our mistakes at the wheel. Data is generated when we use interconnected appliances, from television sets to toasters.

Because this raw material is made up of human characteristics, it may not be mined like metal ores that are brought to a forge for smelting into steel. As a form of "digitized personality", this raw material requires special protection before it may be used.

European data protection law in the General Data Protection Regulation makes companies responsible for protecting personal data by requiring them to ensure that data is protected by means of technological features and data protection-friendly default settings. In line with the ideas underlying the new law, pseudonymizing personal data represents a very important technique for protecting such information. It envisions the use of suitable processes to break the link between data and a given person, thereby making it possible to use the data in a manner that conforms to data protection requirements.

Due to the particular importance of pseudonymization, the Data Protection Focus Group at the German Government's Digital Summit assumed responsibility for this economically significant issue. An interdisciplinary group of representatives from the business community, ministry administration, research, and watchdogs worked with the Focus Group to produce this White Paper.

We would like to express our thanks for their constructive, efficient, and sustained cooperation as part of this project. We would like to express our particular thanks to assessor Steffen Weiß for his expert and painstaking coordination of the Focus Group's work.

Cologne, June 2017

Professor Dr. Rolf Schwartmann

(Head of the data protection focus group for the safety, protection, and trust platform for society and businesses in connection with the German government's digital agenda)

Contents

1. 1. Introduction

The global phenomena of digitization and networking continue apace and are linked with a significant increase in data volumes and information requirements. This also pertains to our personal data. Protecting this data should be a core ambition if we want to ensure that nobody's data can be subjected to processing that he/she can no longer control or even clearly understand. At the same time, technical advances also provide us with new opportunities that arise when we divulge our personal data. There is therefore a need to carefully weigh our desire to have our data excluded from processing against the extent of essential data processing, i.e. by a data controller. As its starting point, this kind of balancing act should be based on the question of whether it is or is not necessary to process personal data.

Pseudonymizing personal data provides one way to negotiate a route midway between the conflicting interests of individuals and data processors, and to develop usage scenarios where plaindata is no longer required. Pseudonymizing is when information pertaining to a person's identity is removed during the course of processing.

Pseudonymization can benefit data protection in myriad ways. For example, even if a controller loses data by accident, pseudonymization makes it very difficult to connect data to the individuals whose information was lost. This ensures their personal rights are not infringed. Data pseudonymization can also win people's trust if highly transparent processes (involving technology) are used to "encode" personal information for the data subject's benefit, and if this approach makes it evident that the relevant controller is interested in safeguarding this information.

This White Paper is intended to provide an overview of the relationship between pseudonymization and data protection, and what functions it can serve. The paper also looks at technical and organizational options for performing pseudonymization. Finally, it investigates specific application scenarios that are already in use when handling pseudonymized data.

The paper's contents are aligned with the legal provisions of the EU's General Data Protection Regulation (GDPR), which will enter into force on May 25, 2018. The GDPR explicitly deals with the topic of pseudonymization in a number of different sections. The Digital Summit can play an important role by contributing, within a European context, information derived from many years of experience with pseudonymization as a result of Germany's still-applicable Federal Data Protection Act, in the hope of assisting all those involved to arrive at a shared understanding of and approach to the matter.

2. Mission for the Digital Summit / goals of the Focus Group

This White Paper was developed within the context of the Digital Summit, the nine-platform response of the German government to the digital transformation. Platform 8, "Safety, protection, and trust for society and businesses," is overseen by Germany's federal interior minister, Dr. Thomas de Maizière, and its mission is to establish online safeguards and security in such a way that digitization can realize its full potential for society and the economy in Germany. A modern, highly effective data protection system can safeguard the freedom and privacy rights of individuals while at the same time making it possible to seize the opportunities associated with digitization. The data protection focus group was established as part of the preparations for Platform 8, and it is headed up by Prof. Dr. Schwartmann.

In addition to investigating the issues of pseudonymization and its application, the Focus Group would also like to see this paper generate guidelines for the legally compliant handling of pseudonymization solutions for deployment by private- and public-sector organizations and bodies alike. Guidelines can play a major role in ensuring that pseudonymization is deployed in a uniform manner across the board.

However, the Focus Group's objectives go further than simply developing guidelines. In the name of Europe-wide harmonization, it would be beneficial to use these guidelines to draft a code of conduct on pseudonymization and thereby create an acknowledged and binding set of standards. The GDPR expressly encourages associations and organizations to draft rules and so add to the detail and scope of GDPR data processing activities, pseudonymization among them. The relevant data protection supervisory authorities will then approve these rules. The European Commission may decide that the approved code of conduct submitted to it has general validity within the European Union.

3. Framework conditions of pseudonymization

3.1. Definition

Pseudonymization has long played a role in Germany's data protection law, and it is enshrined in Section 3.6a of the country's Federal Data Protection Act. Its application takes the form of technical and organizational measures for secure data processing as well as of data-minimizing measures. Section 15.3 of Germany's Telemedia Act even allows the use of data on the condition that such data is pseudonymized. The GDPR now introduces a harmonized definition of pseudonymization for all of Europe, and refers to this solution in several places.

Article 4.5 of the GDPR defines pseudonymization as "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 organizational measures to ensure that the personal data are not attributed to an identified or identifiable natural person".

This gives rise to three requirements that pseudonymization must meet:

a) Data cannot be attributed to a specific person without the use of additional information

If it is possible to attribute data to an identifiable person **without further ado** (i.e. because the data contains a name, an address, or a staff number), this data is not pseudonymized. In this situation, the additional pseudonymization requirements do not apply.

b) Separate storage of additional information

The data which would make it possible to identify a person must be stored separately in such a manner that it cannot be combined without further effort. One option is the logical separation of data by means of different access authorizations. A technical or organizational separation is insufficient if it does not prevent access to the data which facilitates attribution. The protection level of the data in question can be used as a gauge for how thorough this separation should be.

Before applying pseudonymization processes, it is always necessary to clarify who has access to the allocation tables or encryption processes, who generates the pseudonym, whether the risk of depseud-

onymization can be ruled out, and under what conditions the combination of identification data is permitted.[1] If other data is to be added to the pseudonymized data, the new data must be checked to assess whether it could undo pseudonymization because the merger of the two sets of data makes it possible to unambiguously identify a person.

c) Securing technical and organizational measures for non-attribution

Recital 26 of the GDPR makes it clear that personal data that has undergone pseudonymization can be considered information about an identifiable natural person if it is possible to attribute it to a natural person by including additional information. As **data which can be attributed to an individual**, it is subject to the GDPR.
The regulation understands pseudonymization first and foremost as a **risk-reducing technical or organizational measure** (see Recital 28). This recital also clarifies that the explicit introduction of pseudonymization in the Regulation is not intended as an impediment to other data protection measures.

To incentivize the deployment of pseudonymization, Recital 29 clarifies that pseudonymization measures, even if allowing general analysis, should be possible within the same controller if that controller has taken technical and organizational measures necessary to prevent the unauthorized re-identification of the person in question. In other words, it is not necessary to involve a third party, i.e. a data custodian. Individual cases can be checked to assess which variant should be prioritized in light of data protection requirements.

Pseudonymization should also be considered as a criterion when assessing the question of whether processing data for a purpose other than the one for which the data was initially collected is compatible with this original purpose (Article 6.4.e).[2] As pseudonymized data should be treated as personal data, it must be deleted when such data is no longer necessary in relation to the purposes for which it was collected or otherwise processed.[3]

1 Paal/Pauly, General Data Protection Regulation, GDPR, Section 4, Lines 40-47.
2 See point 3.3.2.
3 See point 4.4.

3.2. Differentiating between pseudonymization and anonymization

During the course of the discussion about the GDPR, one question that resurfaced time and again is whether the same difference was made between pseudonymization and anonymization everywhere in Europe. The EU's current Data Protection Directive (95/46/EC) does not mention pseudonymization, and only Recital 26 clarifies that the GDPR's protection principles "should therefore not apply [...] to personal data rendered anonymous in such a manner that the data subject is not or no longer identifiable". Concerning this issue, the Regulation establishes clarity by detailing a definition of pseudonymization in Article 4.5 and distinguishes between pseudonymization and anonymization in the recitals (above all in Recital 26). Article 4 does not itself define anonymization, but such a definition arises from the definition of "personal data" in Article 4.1 of the GDPR, and it is laid out in Recital 26:

"The principles of data protection should therefore not apply to anonymous information, namely information which does not relate to an identified or identifiable natural person ***or to personal data rendered anonymous in such a manner that the data subject is not or no longer identifiable****. This Regulation does not therefore concern the processing of such anonymous information, including for statistical or research purposes."*

The difference between pseudonymized and anonymized data can therefore be described as follows:

Pseudonymized data	⟷	Anonymized data
Individuals can be identified following the addition of information that is stored separately or publicly accessible.	⟷	Identifying individuals is impossible or requires inordinate effort.

At what point exactly it is no longer possible to identify the individual in question is a topic that was repeatedly disputed in the past, and it is relevant for distinguishing between pseudonymization and anonymization. Section 3.6 of Germany's Federal Data Protection Act explains when this is the case, stating that anonymization existed only if disproportionate effort (time, expense, and labor) was required to attribute information to a specific person.

Recital 26 of the GDPR also supplies additional details regarding when information is anonymized in such a manner that it is no longer possible to identify the person in question:

"To determine whether a natural person is identifiable, account should be taken of ***all the means reasonably likely to be used, such as singling out****, either* ***by the controller or by another person*** *to identify the natural person directly or indirectly. To ascertain whether means are reasonably likely to be used to identify the natural person, account should be taken* ***of all objective factors, such as the costs of and the amount of time required for identification****, taking into consideration the* ***available technology at the time of the processing and technological developments****."*

In other words, the following considerations are of key importance:

- The costs of identifying someone
- How long it takes to identify someone
- Technology available at the time of processing
- Technological development (i.e. what future developments can be predicted)
- Other objective factors

Like the Federal Data Protection Act, the GDPR adopts a relative view of how to identify someone. It does not require the absolute irreversibility of anonymization for all time, instead focusing on a **situation** in which nobody can or would, in all probability, undertake de-anonymization because it would require too much effort and be too complex, if not impossible. This view should be based on conditions at the time when the data is processed, but attention should also be paid to technological developments for the future.[4] Another issue where this plays an important role concerns data that a controller pseudonymized and for which only it possesses the information to link the data to the relevant data subject: If this information is passed to another controller, can it be considered as anonymized or not for this second controller?[5]

4 For information on the legal means that a website provider has for linking people to the data of its visitors' dynamic IP addresses, see the European Court of Justice C-582/14, dated October 19, 2016.

5 See point 5.2.3. for additional information

The GDPR criteria named here can therefore be used to distinguish pseudonymization and anonymization in individual cases.

3.3. Functions

3.3.1. Protective function

Pseudonymization fulfills a range of purposes in the GDPR. Of major importance is the protective function it extends to someone whose data is being processed. This protective function is expressed at several points in the law. Pseudonymization protects someone from direct identification. To this end, Article 4.5 of the GDPR states that it is not possible to attribute pseudonyms to a specific person without making use of additional information. It would potentially require inordinate effort to identify the relevant person without additional information over which the processing party has control. The pseudonym functions as a "mask". Further details are provided regarding the prevention of direct identification: This additional information must be stored separately and be subject to technical and organizational protective measures.

The principle of data minimization as per Article 5.1.c (i.e. personal data must be adequate and suitable for the given purpose, and its scope must also be limited to what is required for the purpose of processing) does not directly address the pseudonymization of personal data, but it becomes relevant, and beneficial for the individual in question, when attribution to a specific person is no longer necessary to achieve the purpose of processing. Here, pseudonymization is a measure for implementing what is stipulated by law and an expression of the parsimonious approach to personal data. Continuing in this vein, "privacy by design", detailed in Article 25[6], is a data protection principle that calls for data minimization to be put into practice when the means for processing personal data are chosen and also when technical and organizational measures are used to perform the actual processing. Pseudonymization is one of the options at these two junctures. Following the "privacy by design" principle, pseudonymization ensures that it is possible to uncouple personal information from other data at an early stage. This can facilitate effective and comprehensive protection for the individuals concerned.

According to the GDPR, the security of personal data processing, expressed in Recital 83 by measures for protecting data from its unintentional or unlawful destruction, loss, modification, unauthorized disclosure, or unauthorized access, can also involve pseudonymization. In other words, pseudonymization would be part of a data

6 See Art. 25.

7 Sentence 2 of Recital 28 states: "The explicit introduction of "pseudonymization" in this Regulation is not intended to preclude any other measures of data protection."

protection strategy[7] which can be put into practice via a catalog of technical and organizational measures. In this case, the personal data should be subject to protection that is commensurate with the risk (see Article 32, first sentence, first half-sentence).

Pseudonymization also affords protection according to Recital 28 by reducing risks for people whose data is being processed. Details of these "risks" are provided by Recital 75's information on infringements of the personal data's protection. Such infringements can be described as "data breaches". For example, risks can take the form of physical, material, or non-material damage, such as identity theft or fraud, financial loss, or reputational damage. Pseudonymized data can reduce these risks by making it impossible or difficult to identify the relevant people in the event of data being lost or stolen. EU lawmakers have therefore identified particular information obligations, as outlined in Recital 85, should pseudonymization be breached. The controller for the data must therefore act immediately if the protective function is compromised.

3.3.2. Enabling and easing function

If the purpose of processing does not require directly identifying data subjects but data anonymization is not an option, pseudonymization can be used to protect the data subjects. For anyone involved in processing who does not have access to the key, Identifying the sources is just as impossible is in the case of anonymized data, because the additional information necessary for attributing personal data to specific sources is stored separately.

Given the GDPR's risk-based approach, it is therefore justified that pseudonymization also benefits the controller. Pseudonymization can permit it to process data in ways that would otherwise not be permitted, something that is particularly important now, in the era of **big data** and the **internet of things**.

Article 6.4 describes an important example of this and applies to **processing for a purpose other than that for which the personal data have been collected**. Whether a new purpose is compatible with the original purpose and whether further processing may therefore use the same legal basis require careful consideration. Article 6.4) lists several criteria that must be taken into account. The purposes are compatible if there are suitable guarantees[8] , and pseudonymization can be one of these guaran-

8 Article 6.4.e, along with Article 29 of the data protection group's WP203, pp. 25 and Monreal ZD 2016, 507, 511 put forward the claim that suitable guarantees can make up for deficiencies arising from the application of other consideration criteria.

tees. Pseudonymization is normally not the sole factor, but it can play the most important role in the decision to permit further processing.

A special case regarding the compatibility of processing with the original purpose is outlined in Article 5.1.b: Processing for archiving purposes in the public interest, scientific or historical research purposes, or **statistical purposes**. These activities are compatible with the original purpose if the requirements outlined in Article 89.1 are adhered to. This necessitates suitable measures to protect the rights and freedoms of the data subject. Pseudonymization can be one of these measures provided it makes fulfilling the above-mentioned goals possible (Art. 89.1), second sentence). Pseudonymization can therefore make statistics and research undertakings possible.

In addition to the cases explicitly named in the Regulation, pseudonymizing data can play a role in connection with provisions requiring the **consideration of interests**. For example, Article 6.1.f permits data processing necessary to safeguard the controller's legitimate interests, provided these are not overridden by the interests or fundamental rights and liberties of the data subject which require personal data to be protected. When weighing up different interests, it can be of benefit to the controller if it has pseudonymized the data.[9]

Pseudonymizing data can **free** the controller from **certain data protection obligations or minimize** these obligations. For example, Article 11.1 limits the obligation to **uphold the rights of the data subject**. If identifying the data subject is not or no longer necessary when a controller processes data, the controller is not obligated to maintain, acquire, or process further information for identifying the data subject simply to comply with the Regulation. This satisfies the principle of data minimization as per Article 5.1.c. This changes the requirements regarding fulfillment of the data subject's rights as per Article 11.2.

Just how much pseudonymization can contribute to the enabling and easing function is something that must be assessed differently in different sectors and processing situations. It can therefore make sense to define pseudonymization rules for different situations in sector-specific codes of conduct. Article 40.2.d expressly addresses this.

9 For example, see Buchner/Petri in Kühling/Buchner, GDPR, Article 6, Rn. 154.

4. Techniques, and technical and organizational requirements

Different techniques can be used to implement pseudonymization. For example, it is possible to use an **allocation table** that links every plaintext item of data to one or more pseudonyms. If people have access to the allocation table, they can link a pseudonym to the associated plaintext item of data by scanning the relevant entries. Access to the allocation table should therefore be restricted. Alternatively, pseudonymization can entail the use of different **cryptographic** techniques which transform a plaintext item of data into one or more pseudonyms. Access to the cryptographic keys and, if necessary, other parameters can be used to manage/restrict the irreversibility of the pseudonymization process. The following sections take a particularly close look at cryptographic processes and their uses in connection with pseudonymization.

4.1. Cryptographic basics and processes

This section provides an overview of the main terms associated with cryptographic processes and their uses in connection with pseudonymization.

A **one-way function** is an "easily" computed mathematical function that is "difficult" to invert. This means that it is "easy" to compute the functional value of an input, but inversion is difficult, i.e. it is hard to take the functional value and identify the input from it. Here, "easy" and "hard" are to be understood in the sense of computational complexity theory. Simplifying greatly, "hard" can be rephrased as "practically impossible within a suitable length of time".

A **cryptographic hash function** is a one-way function that is collision-resistant. It attributes a hash value with a fixed length to an input of any length. Collisions exist as the input is larger than the output. However, "collision resistance" makes it practically impossible to calculate these collisions, i.e. different inputs with the same hash value. An example of a cryptographic hash function is SHA-256, which computes hash values with a length of 32 bytes.

An **encryption process** uses a key to transform plaintext into ciphertext. Decryption is the inverse process of transforming ciphertext into the original plaintext. Symmetric encryption processes see the use of the same key for encryption and decryption. Asymmetric encryption processes use a pair of keys, one public and one private. The public key is used for encryption, while the private key is used only for decryption.

Encryption processes are normally deterministic, i.e. the same plaintext is transformed into the same ciphertext (using the same key). If an encryption process gene-

rates different ciphertexts each time while using the same key, this process is described as being probabilistic. Every deterministic encryption process can be converted into a probabilistic process by attaching (for example) a new random value to the plaintext before deterministic encryption and removing the random value following deterministic decryption.

4.2. Requirements regarding pseudonyms

In order to process pseudonymized data, it may be necessary for the generated pseudonyms to contain certain features of the underlying plaintext. The features are determined in advance and then made available via the pseudonyms, while further information about the plaintexts remains hidden. These are described as **availability requirements** for pseudonyms. Pseudonyms which satisfy specific availability requirements are described below as pseudonyms with availability options. Pseudonymization safeguards the confidentiality of the protected data and so safeguards the privacy of the data's owner. Pseudonyms' availability options make part of the information in the underlying data available, thereby ensuring a degree of usability (utility) despite confidentiality. These options therefore provide "utility despite privacy".

4.2.1. Availability requirements

One availability option is the **disclosability** of the plaintext underlying the pseudonym in certain situations. This can be achieved by generating the pseudonym by encrypting the plaintext. Therefore, if someone knows the technique and the key used, he/she can decrypt the pseudonym and so disclose the underlying plaintext. This disclosure may be bound to fulfilling a specific purpose or usage by someone with a specific role.

Another availability option is **linkability** regarding a relation r. If linkability exists regarding a specific relation, someone can take two pseudonyms and assess if the underlying plaintexts align in that relation context. Linkability in terms of equality is a simple example. Pseudonyms that meet this availability option make it possible to check if the plaintexts underlying two pseudonyms are identical. Linkability can also be bound to specific roles or purposes.

4.2.2. Role binding

Availability options can be bound to specific roles. If different roles are defined in a system, different availability options can be assigned to the various roles. This makes it possible to ensure that only people with a given role have access to the information provided by a given availability option.

4.2.3. Purpose binding

Availability options can be restricted to usage for specific purposes. For example, a system can grant access to a pseudonym associated with a specific availability option under the condition that evidence of a spe-

cific situation pertaining to a purpose has been provided. It is possible to use technical or organizational methods to enforce the purpose binding of availability options.

In the case of **technical** purpose binding, purely technical methods are used to ensure such purpose limitation of an availability option without the need for a person to perform any action. These technical methods can include checking system values (which characterize the purpose) such as system time before approval of an availability option. Using threshold schemes is another example. If specific events documented in data sets occur very frequently, the relevant availability option is approved only when a fixed frequency threshold is exceeded.

In the case of **organizational** purpose binding, one or more natural persons are given the power to approve certain availability options. This means that a natural person interacts with the system to assess if the purpose applies. To establish an organizational purpose limitation, the relevant pseudonyms can be additionally encrypted using a key known only to that person.

The advantages of technical over organizational purpose binding deserve emphasizing. In contrast to organizational purpose binding, no user interaction with a system is necessary. This limits the exploitation of power. The underlying trust model of a technical purpose limitation can be derived from the security of the implemented technology. Purely technical implementation can ensure the real-time automatic enforcement of purpose limitation.

4.3. Examples of pseudonymization techniques for implementing availability requirements

This section describes examples for creating pseudonyms with specific availability options. The equality of the plaintexts underlying pseudonymization is considered for the purpose of data linkability.

4.3.1. Linkable, disclosable pseudonyms

Deterministic encryption algorithms can be used to create linkable, disclosable pseudonyms. Linkability (even without knowledge of the encryption or decryption key) is established because these algorithms assign identical plaintexts to identical ciphertexts (pseudonyms). Disclosability is established when the decryption key is known.

4.3.2. Non-linkable disclosable pseudonyms

Non-linkable but disclosable pseudonyms can be created by means of probabilistic encryption techniques. Pseudonym linkability is not established because probabilistic encryption algorithms map identical plaintexts onto non-identical ciphertexts (pseudonyms). Disclosability is established when the decryption key is known.

4.3.3. Linkable non-disclosable pseudonyms

Deterministic one-way functions (e.g. cryptographic hash functions) can be used to create linkable but non-disclosable pseudonyms. Linkability is established because deterministic algorithms map identical plaintexts onto identical result values (pseudonyms). A one-way function is used for preventing the reversal of pseudonymization, i.e. disclosure.

4.3.4. Role binding

Availability options can be linked to roles by means of an additional probabilistic pseudonym encryption with one of the decryption keys assigned only to the respective role. This ensures that only the role in question has access to the linkable or disclosable pseudonym. It is possible to restrict disclosure of pseudonyms by ensuring that the decryption key necessary for disclosure is assigned solely to a particular role.

4.3.5. Organizational purpose binding

Linking a role to the person assessing the purpose of a data processing can present an organizational purpose limitation regarding availability options of pseudonyms. In the case of disclosability of pseudonyms bound to a specific purpose, an exceptional situation can be recognized due to existing pseudonym linkability and corrective measures based on plaintext can be initiated.

Correspondingly, a person assessing the purpose can ensure a disclosability of pseudonyms restricted to a predefined exceptional situation by using the decryption key only he/she knows for decrypting the pseudonym.

4.3.6. Technical purpose binding

Technical purpose binding can also restrict data usage to exceptional situations, for example the frequent occurrence of data sets with pseudonyms representing the same plaintext. Such occurrence would be connected to a frequency threshold. In this context, a cryptographic secret sharing scheme[10] can be used which manages a secret key for every plaintext value and adds a unique partial key from the associated secret key to the pseudonym every time a plaintext value is pseudonymized. If the number of pseudonyms related to a plaintext exceeds the defined threshold, the partial keys can be used to determine the secret key, and decryption can be performed.

4.4. Technical and organizational requirements

The technical and organizational requirements called for in the GDPR can be elaborated as follows.

10 See A. Shamir. How to share a secret. In: Communications of the ACM Bd. 22, ACM, 1979, p. 612–613.

- Pseudonymization requires the use of **state-of-the-art** techniques (see BSI [German Federal Office for Security in Information Technology] or ENISA guidelines on cryptoprocedures for examples). These techniques must be replaced from time to time by current techniques (particularly in the case of pseudonymized data intended for long-term use). This provides a very high level of security.
- Pseudonymization must be performed as early as possible when processing data.
- In the case of plaintext data from small value ranges or with limited diversification within a range, pseudonymization processes are prone to pseudonym disclosure due to brute force attacks, for example using rainbow tables. In these attacks, the attacker calculates a plaintext-pseudonym allocation table for all (few) possible plaintexts or uses a previously created table. The table can be used to determine the corresponding plaintexts for given pseudonyms. **Salt values** can be used to make things difficult for the attacker and so reduce the risk. Before running the pseudonymization process, a salt value is selected according to the context (i.e. for each data set) and combined with the plaintext value. This enlarges the value range and value diversification, and it is no longer possible to use allocation tables calculated or created in advance as they cannot factor in the salt value combined with the plaintext. If the same pseudonyms need to be created for the same plaintexts, the same salt values must be used and stored in a suitable manner.
- **State-of-the-art** technical and organizational measures must be used when creating and managing (incl. distributing, storing, using, deleting) secret parameters (keys and salt values).
- Depending on the use case, suitable intervals (depending on time or data volume) must be defined for changing the secret parameters (salt values and keys) used.
- Access to salt values and keys must be restricted to an absolute minimum of trustworthy users (need-to-know principle).
- Integration of the pseudonymization concept in an IT security management system (i.e. as per ISO/IEC 27001) to prevent unauthorized access to pseudonymized data.
- Pseudonymized data must be deleted in line with data protection regulations when the purpose for processing no longer exists.

5. Transparency and data subjects' rights regarding pseudonymized data

5.1. General transparency requirements

Transparency must be established concerning data processing, the processed data, the purpose and procedure of data processing, and many other issues concerning data processing (see in particular Articles 13.1 and 13.2, 14.1. and 14.2, 15.1 and 15.2).

The goal of establishing transparency is to resolve the uncertainties of the data subjects regarding data processing, give them the tools to protect their data themselves, and to enable them to exercise their rights in the context of personal data processing. This must be done in a concise, **transparent, intelligible, and easily accessible form**, using clear and plain language (Article 12.1.1.) Suitable, easily understood (by the data subjects) icons can be used in the context of information obligations as per Articles 13 and 14 (Article 12.7).

5.2. Special conditions regarding pseudonymized data

5.2.1. Information obligations (GDPR Articles 13 and 14)

5.2.1.1. Initial situation

Basic information obligations as per Articles 13 and 14 entail virtually no special conditions concerning the processing of pseudonymized data. Information as per Articles 13.1 and 13.2 must be provided to the data subject "**when personal data relating to the data subject are collected from the data subject**". Information as per Articles 14.1 and 14.2 shall be provided "within a **reasonable period** after obtaining the personal data" (see Article 14.3.a). Normally, data may not be available in pseudonymized form when collected, permitting easy notification of the data subject. Similarly, the situation in Article 14.3.b, in which the personal data is to be used to communicate with the data subject (i.e. in the case of direct marketing) does not give rise to any special conditions as communicating "with" the data subject obviously precludes eliminating the data's assignment to the data subject.

5.2.1.2. Special situation: Disclosing pseudonymized data to a third party

In contrast, problems could arise in connection with the situations described in Article 14.3.c and (in particular) in the event of further processing for another purpose as per Articles 13.3 and 14.4. If the controller performed pseudonymization before disclosure to another recipient or before intended further processing, the recipient can no longer easily inform the data subject as it cannot identify this person without the addition of further information. In such situations, it is recommended that the third party processing pseudonymized data provide **general information** via its own website stating that it processes pseudonymized data.

Information as per Articles 13 and 14 of the GDPR does not need to be provided on an individual-by-individual basis: Instead, this can be handled in advance via a website, basic contractual details, or notices, as arises in particular from Recital 58, sentences 1-3. Another factor that rules against sending individual information is that such notification (e.g. via e-mail) would overburden the data subject with information. Most data subjects would therefore not take any notice of the information they receive. Again in connection with Articles 13.3 and 14.4, if the decision is taken at a later date to make further use of the data and the data subject therefore did not receive notification when his/her data was collected or initially used, the data subject can be informed via publicly accessible sources (i.e. campaigns on a website or in other media).

5.2.2. Data subjects' rights as per Articles 15-22 and 34

The following conditions apply when satisfying the information obligation in Article 15 and data subjects' rights as per Articles 16-22 after pseudonymization of data.

Special rules apply if the controller is not or no longer required (Article 11) to **identify the data subject**, something that is frequently the case following pseudonymization.

a) As per Article 11.1, the controller is not obligated to maintain, acquire, or process additional information in order to identify the data subject for the sole purpose of complying with the GDPR.

b) If the controller can prove that it is not able to identify the data subject, it must **inform** the data subject of this (Article 11.2) if possible (i.e. normally in connection with data subject rights requiring an action to be taken from the data subject).

In situations **a)** and **b)**, data subject rights as per Articles 15-20 do not apply. This pertains to rights concerning information, corrections, completion, deletion

(except when the purpose of processing no longer exists, in which case the data must be deleted), restrictions on processing, notification, and data portability. This does not apply only if the data subject successfully facilitates identification by **providing additional information** (Article 11.2, second sentence, last half-sentence).

Data subject rights as per Article 21 and 22 apply instead.[11] Accordingly, the data subject can continue to file an objection to the following types of data processing[12] (Article 21):

- Processing pseudonymized data on the basis of legitimate interest (as per Article 6.1.f), which also comprises profiling based on legitimate interest,
- Processing required to perform a task in the public interest, or in the exercise of official authority vested in the controller (Article 6.1.e),
- Processing for research or historical research purposes, or for statistical reasons as per Article 89.1 (Article 21.6),
- Processing for direct marketing purposes (Article 21.2).

It is unclear what role the **right to object to direct marketing** plays in the context of pseudonymization. If direct marketing is understood as an activity whereby a specific data subject is contacted in person (letter or e-mail), such a situation does not arise in the case of pseudonymized data. Therefore, advertising facilitated by tracking technology, i.e. cookies, mobile identifiers, fingerprinting, etc. is not subject to the right to object laid out in Article 21.2 as such technology does not enable an individual to be identified. A data subject would still be able to object as per Article 21.1 to advertising using pseudonymized data. In such a situation, the difference to Article 21.2 is that the latter features no further prerequisites, whereas the data subject has particular reasons when making an objection as per Article 21.1. The various interests must be weighed up.

Furthermore, the data subject has the right to not be subject to a decision based only on automated processing (incl. profiling) which produces legal effect concerning him or her or similarly significantly affects him or her (Article 22).

If the data subject's objection is successful, his/her original data set is **blocked**: It cannot be processed for the purpose to which he/she has objected. This blocking is unproblematic if the objection takes place before pseudonymization. If the objection takes place after pseudonymization, the controller must first reidentify the data subject before it can perform the necessary blocking process. However, it is normally not possible to remove the data set from data that has already been pseudonymized as it is normally a case of large-scale automated data processing.

To sum up: It is recommended that controllers planning to further process pseudonymized personal data to make this processing transparent to data subjects in advance. They should allow for a right to object which permits the future exclusion of a data subject's data from pseudonymized further processing after the data subject has exercised his/her right to object.

Article 34 deals with another data subject right: The controller must immediately inform a data subject if the personal data breach is likely to result in a high risk to the rights and freedoms of natural persons. Subject to a review of the specific case, it is to a degree likely that the breach will not result in a high risk for the data subject if the controller has already performed pseudonymization. For example, in the event of data theft, the culprit will certainly be unable to identify the data subjects in the majority of cases. It is recommended that the controller notifies the data subjects via specified channels. An individual information is not considered necessary.

5.2.3. Transferring pseudonymized data to third parties

If the controller transfers pseudonymized data to a third party, the recipient must check if the data can be considered to be unidentifiable data as per Article 11. More than the initial controller, the data recipient has the problem of being unable to comply with data subjects' rights because it cannot establish a link between the data and a person who wants to exercise his/her rights.

5.3. Tracing results to a person

Tracing pseudonyms based on more complex pseudonymization processes (i.e. using a recognized hashing algorithm) to a natural person is only possible for whoever has the key used for pseudonymization.

If the data is to be traced back to a specific person and if this tracing does not serve to comply with rights vis-a-vis the data subject, it requires the data subject's consent. The data processing controller has this key and must use it if the data subject exercises his/her rights. Depending on the data subject's specific choices, the controller must then provide information about the data, correct the data, or delete it. This may require the involvement of a service provider that processes data on behalf of the controller.

---

11 Unlike Article 11.2, Article 12.2 refers not only to data subjects' rights in Articles 15-20 but also to data subjects' rights in Articles 21-22. According to Article 12.2, line 1, the controller also seems to therefore have the right to refuse in the events of objections (Article 21) and automated individual decision-making (Article 22). Articles 11.2 and 12.2 are clearly contradictory.

12 This applies to personal data processing and pseudonymized data processing.

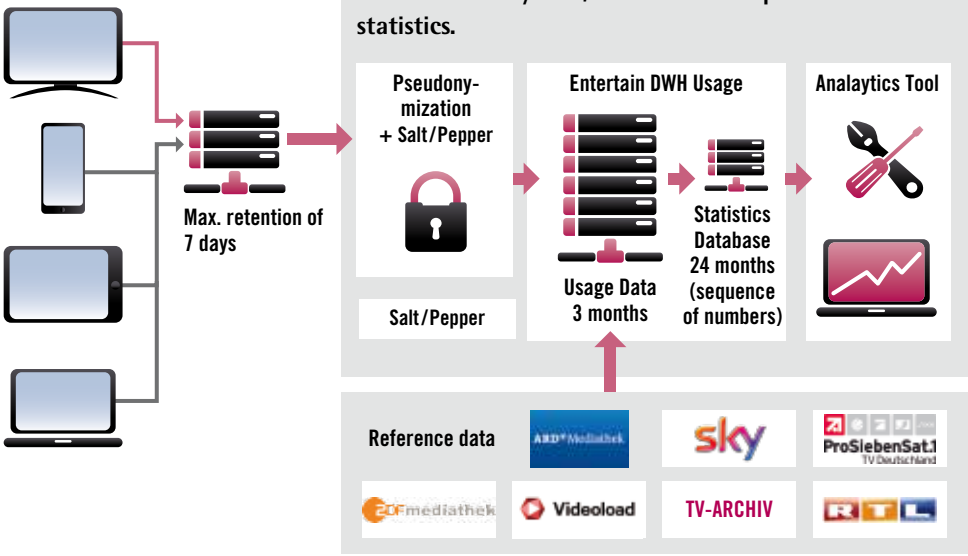
6. Application scenarios

6.1. Pseudonymization and Entertain TV (DTAG)

Deutsche Telekom markets access to television programs and films via the internet under the product name of Entertain TV. The company provides customers with a set-top box for using the product. Statistics on viewers' habits are maintained for various purposes, including obligations vis-a-vis broadcasters.

6.1.1. Overview

The following diagram shows the data flow through the individual system, from the set-top box to the statistics.



6.1.2. Data generation

Using the set-top box, i.e. when the consumer uses the system's remote control, generates a range of events depending on what button was pressed and the relevant context. These events form the basis of the analyses, which document activities such as activation/deactivation, channel changes, information about the programs watched, information about users' recording activities, or information about users watching recorded programs. Corresponding event

data sets contain a range of information, i.e. about the set-top box (device ID), the customer's account ID, date/time, and other specific subjects.

6.1.3. Pseudonymization

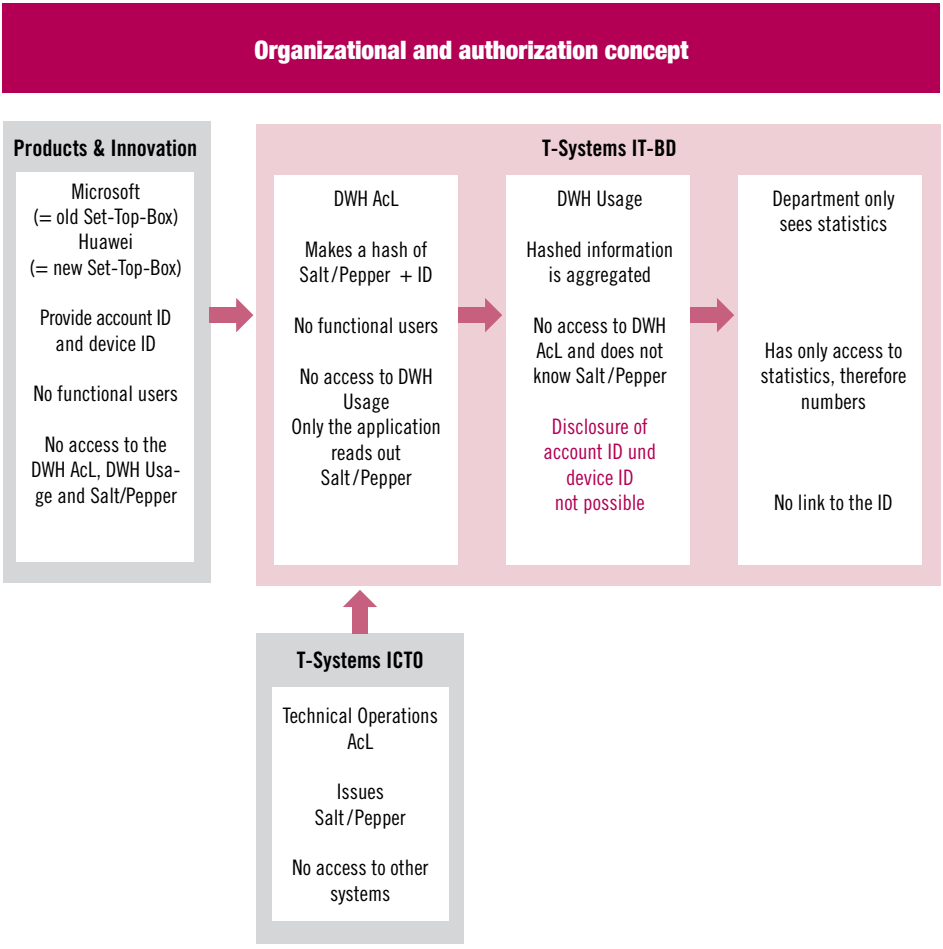
The account ID is a pseudonym for the customer, and the device ID is a pseudonym for the associated set-top box. The event data sets required for analysis purposes do not contain any attributes featuring personal data (of an immediate kind). Managed separately in organizational terms, there are allocation tables which permit the pseudonyms (account and device IDs) to be linked to customers or set-top boxes. Access to these tables would make it possible to ultimately trace the device and account IDs to trace back to the customers. Tracing is sometimes necessary, i.e. for billing services as per the contract.

However, as no party wants to trace details back to plaindata for the purpose of generating statistics, account and device IDs are subjected to additional pseudonymization before processing. In this instance, pseudonymization takes place within the Data Warehouse Acquisition Layer (DWH ACL) unit, and processing (statistics generation) takes place in Date Warehouse Usage (DWH Usage), a separate unit (see chart below).

The underlying pseudonymization process means that statistics are generated using pseudonyms that are linkable but non-disclosable for DWH Usage, the unit involved in processing. The pseudonyms are created using a deterministic cryptographic hashing process (SHA-512) and a uniform pseudonymization salt/pepper component. Pseudonym linkability is established because deterministic processes transpose identical plaintexts onto identical result values (pseudonyms) when salt/pepper is identical. The output length of SHA-512 means that the risk of collisions is negligible ($< 10^{-70}$). As the DWH Usage Department does not have access to the pseudonymization salt/pepper, it cannot trace pseudonyms back to people and so disclose plaintexts. This also applies if getting around a one-way function by creating a look-up table, which is possible without using or knowing the salt/pepper due to the limited number of possible input values.

Ultimately, the pseudonymization process used means that no Deutsche Telekom Group employee can view, analyze, or transfer anyone else information about specific customer's usage behavior.

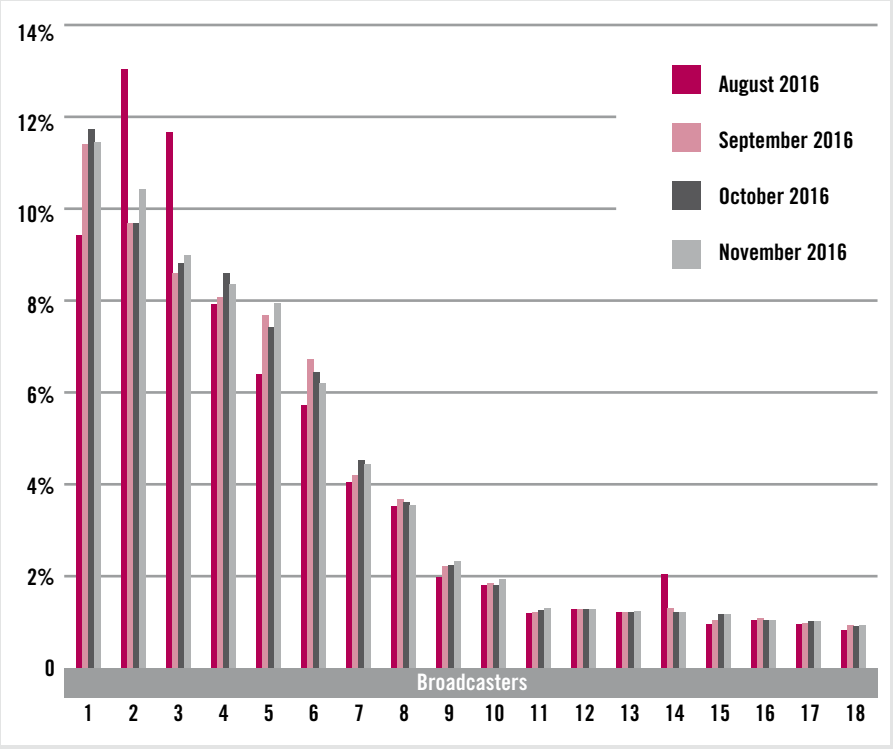
The following chart outlines the underlying organizational and authorization concept, and the various units involved.



6.1.4. Statistics generation

All attributes traceable to the end customer are pseudonymized using an account or device ID. Payments are possible as linkable pseudonyms are used. For example, this means that it is possible to answer a question about how many households or set-top boxes watched a certain channel at a certain time. The anonymous statistics no longer contain account and device IDs or the generated pseudonyms, which prevents the statistical figures from being traced back to the hashed IDs.

Deutsche Telekom must meet certain obligations towards broadcasters, so it transfers only anonymized statistics on viewers' user behavior, e.g. market share, using relative figures as illustrated in the following table.



6.1.5. Opt-out

Data protection notifications inform every Entertain customer that data is gathered for statistical purposes. Deutsche Telekom uses e-mails and pop-ups in Entertain itself to inform customers of this before introducing this analysis solution.

Every customer has the option of **objecting** (opt out) to his/her pseudonymized usage data being collected and analyzed. He/she can use his/her set-top box to perform this opt-out. This used to require the input of a PIN number regarding the previous product (old Microsoft set-top box), but it is no longer necessary for the new product (new Huawei set-top box). By opting out, the customer's usage data is not used either for a pseudonymized usage profile or for anonymous statistics. Customer can also use conventional communication channels to inform Telekom that they want to exercise their opt-out right.

6.2. Direct marketing

6.2.1. Promotional campaign

The objective of an advertising campaign is to deploy advertisements within a certain period of time in a certain target market or for a certain target group in order to increase sales or increase awareness of a product or service. Analyses based on pre-existing data are necessary to identify a target market or target group for these campaigns. If an advertiser's existing data pool is not sufficient, this data is augmented with additional characteristics. The desired characteristics that were previously not present, e.g. age and estimated income bracket, are often obtained from external sources.

To select a target group, a range of characteristics such as age, previous purchasing behavior, etc. can be used to assess the likelihood of members in this group buying a certain product. Similarly, offline (direct mailing, phone-based advertising, and e-mail advertising) and online options are available when defining target groups according to their affiliation with a specific customer segment. Pseudonymized data plays an important role in campaigns of these kinds. Further details are given below.

6.2.1.1. Using pseudonymized data for comparing and matching data with external sources – creating an analysis database

When augmenting the existing data with externally sourced data, the first step is to compare the databases of the advertiser and data provider and then match them. The relevant steps in this process are as follows:

1. A "data hygiene process" (standardization, homogenization, parsing, checking for duplicates, etc.) is used between the databases to correct **characteristics which instantly** identify data subjects (e.g. name and address) and match their quality levels before saving.

2. The same algorithm is used to pseudonymize the processed name and address fields in both databases to form a linkable ID number. The deployed pseudonymization process varies between a simple "memnum", i.e. a 17-figure number consisting of the first three letters – (distributed throughout the memnum) of the first name, last name, street and house number, and zipcode, and a proprietorially developed pseudonymization algorithm. The system then deletes the **direct identifiers**.
3. The ID numbers are used to compare the databases of the advertiser and data supplier against each other. If the same pseudonyms (ID numbers) are found ("linkability"), the relevant data sets can be matched. A company can use this procedure to save additional characteristics in an external source's marketing file.

The result is an analysis database comprising in-house and purchased characteristics and ID numbers, and without names or addresses. This database can be used to create a customer profile (i.e. age group of 65+ and living in a city). This information can then be used to address people with similar profiles (i.e. by selecting only people aged 65+ and living in cities from the consumer database).

6.2.1.2. Interest-based advertising: Using pseudonymized data for selecting data

1. The advertiser's analysts select the ID numbers for a promotional campaign from the in-house analysis database. They are selected according to a simple selection of characteristics (i.e. age 65+ and city = "Frankfurt am Main" or "Berlin" or "Munich" or "Stuttgart"), However, more complex statistical models can also be used.
2. The selected ID numbers are matched with the reference file which stores the plaindata (name, address, etc.) and ID numbers.
3. The advertising material is then dispatched to the selected customers (via e-mail, direct mailing, display, etc.).

6.2.1.3. Gauging the success of a campaign with pseudonymized data

A promotional campaign's success is normally evaluated after it finishes. Reactions to the campaign, i.e. information requests or purchases, are saved, and the advertising message and success of the advertising channel are evaluated. Reactions are saved as raw data in the advertiser's CRM system for future analyses. Just as when creating an analysis database, the direct identifiers are removed in this instance before the data is made available to the analysis unit.

6.2.2. Data marketing on behalf of an agency/data broker

If commercial data is marketed, the data owners (the entities responsible for collecting, processing, and using the data) integrate a control mechanism so that every time that data is to be deployed or sold, such an action may only be performed by the hired order data marketer (agency/broker), and that the data owner is aware of this. This ensures that the necessary licensing fee is paid correctly. Data marketing usually consists of the following steps:

1. The data owner concludes an agreement with the agency/data broker regarding order data processing.
2. A file with selection characteristics and pseudonyms (ID numbers) but without direct identifiers is made available to the agency/data broker.
3. The agency/data broker selects the pseudonyms (ID numbers) using the selection characteristics for its customers and sends this selection to the data owner.
4. The data owner transfers the data straight to the customers of the agency/data broker, which contacts its end customers in writing (or it uses a lettershop for sending printed advertising to end customers).

6.2.3. Using display advertising

For online services such as publishers, social media and e-retailers, advertising is an important revenue generator. In some sectors, it is the sole source of income, and it enables internet users to use services free of charge. Advertisements should be of relevance to users and not seem like a nuisance. The following example is one of the variants possible for interest-based display advertising, i.e. when an advertisement appears automatically on the screen of a user's device.

1. A user registers on or logs into the online service provider's website.
2. The online service pseudonymizes the user's contact data (P1) by creating a user ID.
3. The online service installs a cookie with the pseudonym/user ID (P1) on the user's computer.
4. In a parallel process, the online service sends the pseudonym/user ID number (P1), the user's name, and the user's address to an agency at regular intervals as part of its order data processing activities.
5. On behalf of the online service, the agency processes the data (pseudonym/user ID number (P1), name, address) provided separately in file form and generates an internal pseudonym (P2) based on this information. The system then deletes all information apart from the two pseudonyms (P1 and P2).
6. The agency now has a file with pseudonyms (P2) and consumers' characteristics (estimated/identified). Using the pseudonym (P2), further data is added to the consumer characteristics (estimated characteristics and affinities) saved by the agency for a user. The pseudonym (P2) is then deleted.
7. The pseudonym (P1), the agency characteristics, and a source ID are made available to a real-time bidding market for selection via an advertising place platform – a demand side platform (DSP), sell side platform (SSP)[13] or other. This enables an advertiser such as a garden center to target people with, for example, a house that has a garden. The source ID identifies the online service provider and ensures that the data stocks can be kept apart, and that separate invoicing is possible regarding the online services.
8. If the DSPs/SSPs or another party in the real-time bidding market identifies a cookie with a pseudonym (P1), it can use this cookie to zero in on the relevant user's display in an anonymized but still targeted manner.

Just like with written advertisements, the user is able to make an **objection** against interest-based advertising. The sectoral self-regulation of the EDAA (European Digital Advertisement Alliance), operated in Germany by DDOW (Deutsche Datenschutzrat Online-Werbung) offers users this option in the form of an icon.[14]

---

13 In their role as technical service providers, DSPs help advertisers to find the right advertising space to their target group at predetermined conditions: They are hub platforms that enable the efficient sale of advertising inventory via a range of channels (ad networks, ad exchanges, etc.). SSPs are different in that they handle the sale of advertising spaces.

14 For further information, see http://www.youronlinechoices.com/de/nutzungsbasierte-online-werbung/.

6.3. Pseudonymization in the guidelines on data protection in medical research – TMF's generic solutions 2.0

6.3.1. Overview: TMF's data protection guideline

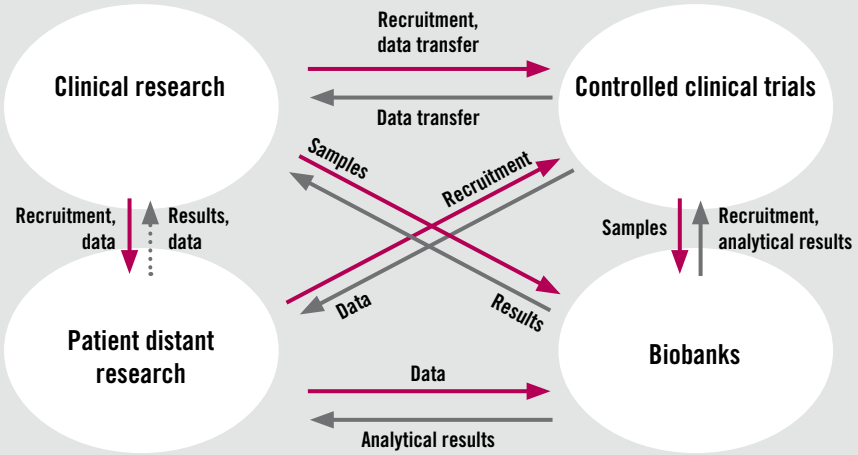
TMF[15] first published its data protection guideline for medical research projects in 2003, and it is now in its second edition (2014). The idea for developing the guideline was inspired by the creation of research networks throughout Germany in the 1990s. These networks were faced with uncoordinated data protection regulations issued by the different states in Germany as well as the national government, all of which were in turn overlaid by EU laws. The guideline's goal was to identify generic solutions for dealing with complex situations.

Medical research requires sensitive data that almost always takes the form of medical details or at least contains this kind of information. Data protection also therefore concerns doctor-patient confidentiality if data is sourced from treatment situations. As medical confidentiality is part of a doctor's professional rights and can be applied along with data protection rights, TMF's data protection guideline can only allude to this legal framework to a lesser degree.

The guideline has a modular structure. The individual modules depict typical situations so that these can be used as generic foundations for developing suitable specific data protection concepts. However, there is a certain interdependence between the modules, as the data is often used in different contexts or passed from one context to another.

15 Technologie- und Methodenplattform für die vernetzte medizinische Forschung e.V.

Medical research areas: The TMF data protection guideline's modules



The special feature of the TMF's data protection guidelines is the fact that the generic concepts were agreed in extensive negotiations with Germany's data protection authorities, and they have been recommended by the participants at the conference of Germany's state and national data protection officials since 2003. This establishes a level of legal security for researchers as the TMF's guidelines guarantee that a data protection concept meets the data protection requirements. Since TMF's foundation, a data protection working group operating within it also contributes to support for research institutes implementing the guideline. As part of a peer review, the working group presents a data protection concept that members study, thereby identifying weaknesses and proposing additional measures. The concept may be edited, and the members vote on its conformity. The relevant data protection authority then normally recognizes this vote as fit for purpose.

6.3.2. Pseudonymization as a technical and organizational measure for information-al separation of powers

The "informational separation of powers" is a cornerstone of TMF's generic data protection concept. Pseudonymization is the main tool for separating identifying data on the one hand and producing research data on the other. It serves as a technical and organizational measure for protecting test persons.

During research activities, data processing is generally performed on a consensual basis, i.e. test persons receive information, normally during the data's collection, explaining what the objectives and risks are. This takes place before the data is used, and it is known as "informed consent". The data can be used while it still reveals a link to the data subject: Anonymization is necessary only in connection with the principle of data minimization but is not required for the data's usage.

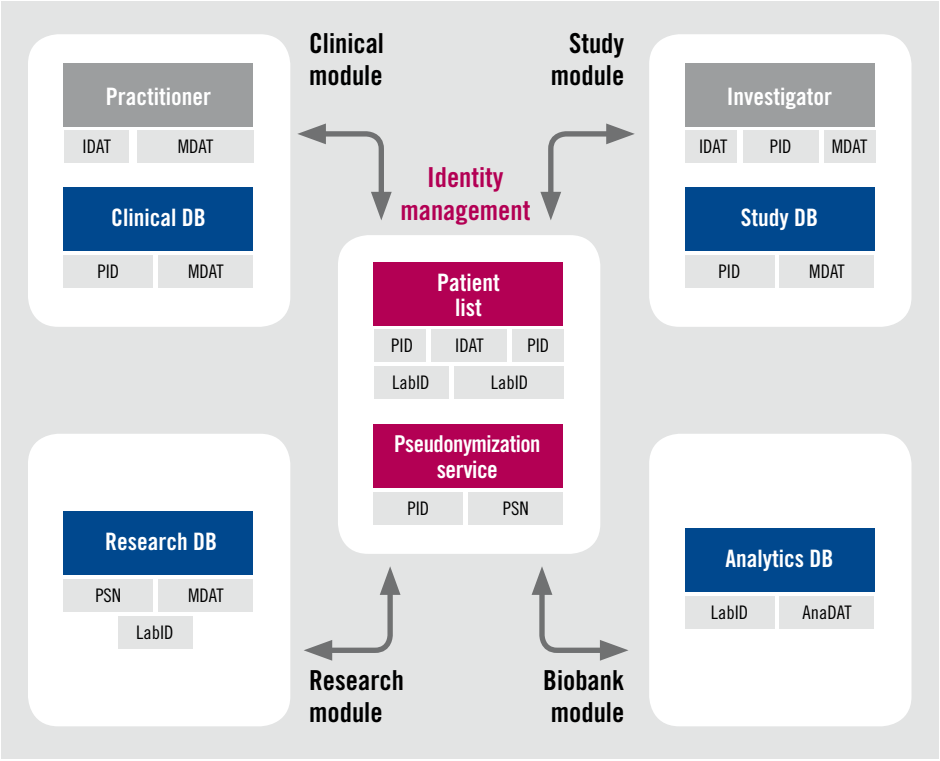
However, the GDPR[16] interprets pseudonymization differently from the Federal Data Protection Act[17]. The difference may seem slight, but it has considerable repercussions. While the Federal Data Protection Act merely requires re-identification to be difficult, the GDPR requires it to be impossible, which is identical to anonymization. Achieving de facto anonymity for pseudonymized data sets raises the requirements considerably. However, within the context of biomedical research, this is unlikely to be possible as the data sets are very often large and complex, and, increasingly, they feature genomic information, which would make data that is de-identified to the point of anonymity practically worthless for research purposes. Furthermore, the data's anonymity is not a feature that is secure in the long term – it depends on a range of factors, including unknown outside additional information, some of which might not be available until some time in the future. As a result, research data is generally not actually anonymous but instead largely de-identified so its usability can be maintained for the purpose in question. Data from which immediate identifiers have been removed normally contains a lot of additi-

16 Article 4.5 of the GDPR defines pseudonymization as "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 organizational measures to ensure that the personal data is not attributed to an identified or identifiable natural person".

17 Section 3.6a of the Federal Data Protection Act (6a) states that in pseudonymization, the name and other identifying features are replaced with code, so as to make it impossible or far more difficult to identify the affected individual.

onal details that, combined, do not prevent re-identification. Re-identification is merely made more difficult, but not de facto impossible. Pseudonymization is, however, just one of the options for protecting test persons. Others are conceivable and, given the situation, practical. The question arises here if a new term needs to be found for data that is "encoded" but are not pseudonymized as per the GDPR.

The following model is based on the Federal Data Protection Act's interpretation of pseudonymization. This model can also be applied in line with the GDPR, but details may needs adjusting because the remaining data sets generally do not meet the pseudonymization requirements laid out in the GDPR.



PID Patient identifier (pseudonym within the clinical database) · **MDAT** Medical data · **IDAT** Identifying data of a patient · **LabID** Laboratory data/sample number · **PSN** Pseudonym in the research context · **AnaDAT** Analysis data

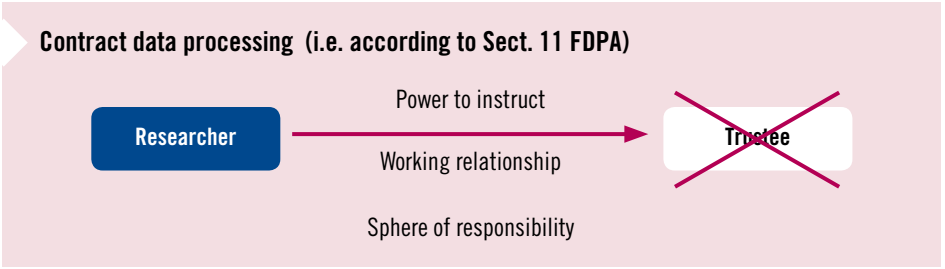
A pseudonym must not be revealing: It may not contain elements that reveal identification data, i.e. dates of birth, initials, etc. Pseudonyms must only be allocated using an assignment list or rule which must be guarded well.

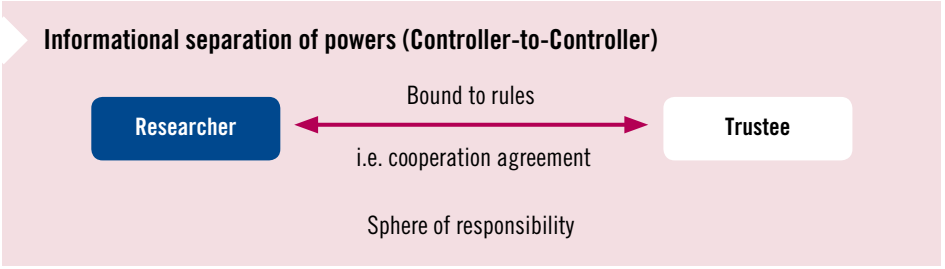
6.3.3. Involving a trustee

Involving a trustee is recommended for complying with to the order to store data separately as per Section 4.5 of the GDPR. The TMF guideline has considered this as useful as per existing data protection law. A trustee (TTP) that stores patients' identification data (IDAT) and pseudonyms (PID) must meet certain requirements.

>> It must be an independent legal entity (own permanent legal form).
>> It must have its own separate premises and staff.
>> It must be contractually obligated to implement the data protection concept.
>> It must be otherwise independent.

Order data management by the custodian for the research institute does not meet these requirements.



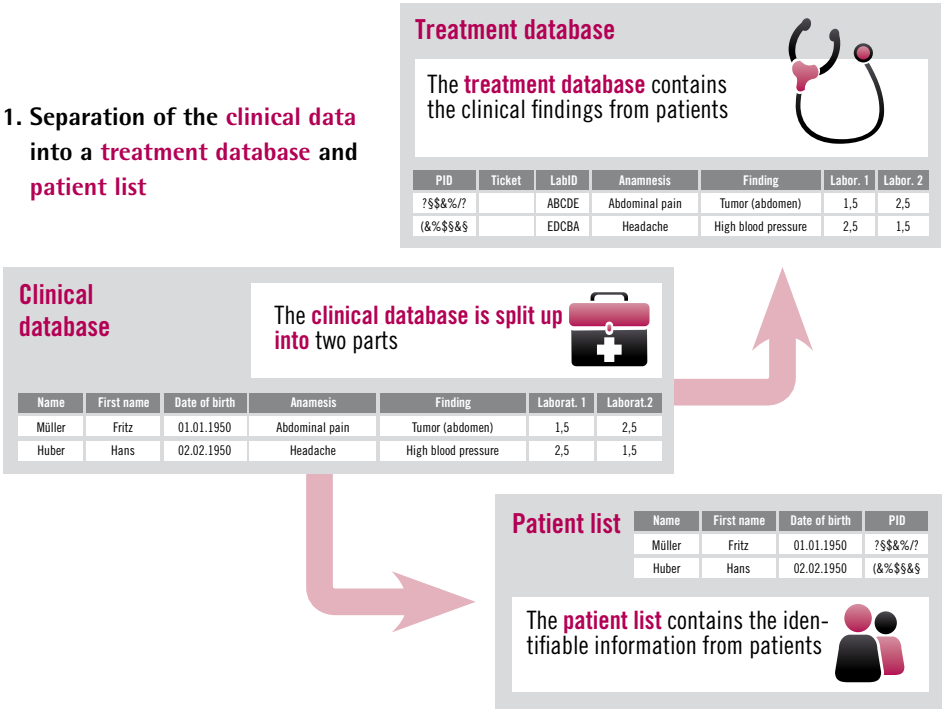


Example: Key elements of the clinical model

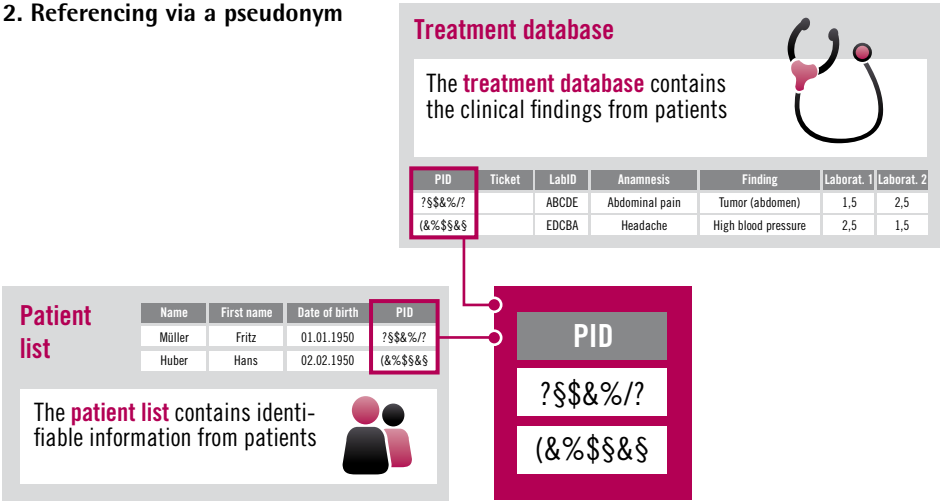
>> Central data pool (e.g. EHR repository or data warehouse)
>> Online access to all data for staff performing treatment (IDAT & MDAT)
>> Saving MDAT under pseudonym (PID)
>> Pseudonym (PID) known only to data warehouse staff and custodian
>> MDAT and IDAT both visible only at client workstations (in treatment context) and not on application server
>> IDAT and PID both visible only for custodian
>> Access controlled via 1x token in treatment context
>> No public use of data pool (no access and no search function from outside)
>> External research only with exported data

The process of transferring a clinical database to a treatment database available to medical staff for further patient treatment and also containing data for research purposes consists of the following steps:

1. Separation of the clinical data into a treatment database and patient list



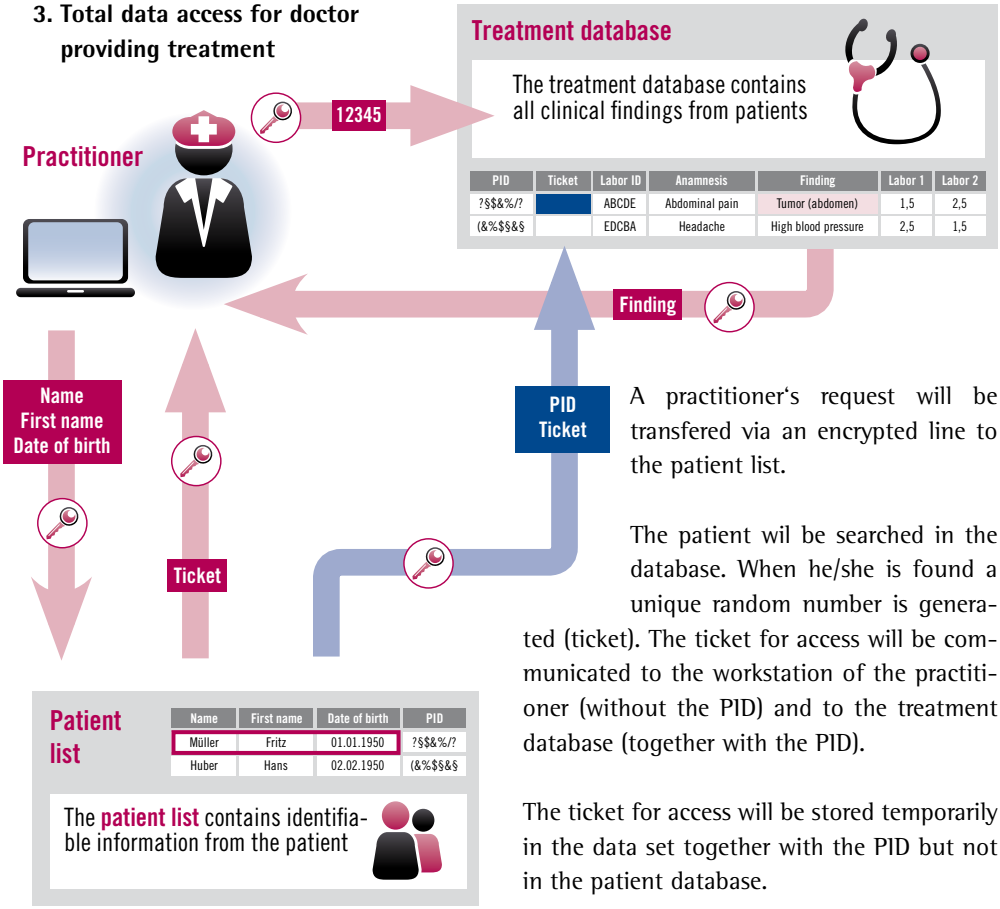
2. Referencing via a pseudonym



The patient list and the treatment database reference one another using a shared random secret identifier (PID).

The patient list and the treatment database are logically and physically seperated. They are administered independently.

3. Total data access for doctor providing treatment



A practitioner's request will be transfered via an encrypted line to the patient list.

The patient wil be searched in the database. When he/she is found a unique random number is generated (ticket). The ticket for access will be communicated to the workstation of the practitioner (without the PID) and to the treatment database (together with the PID).

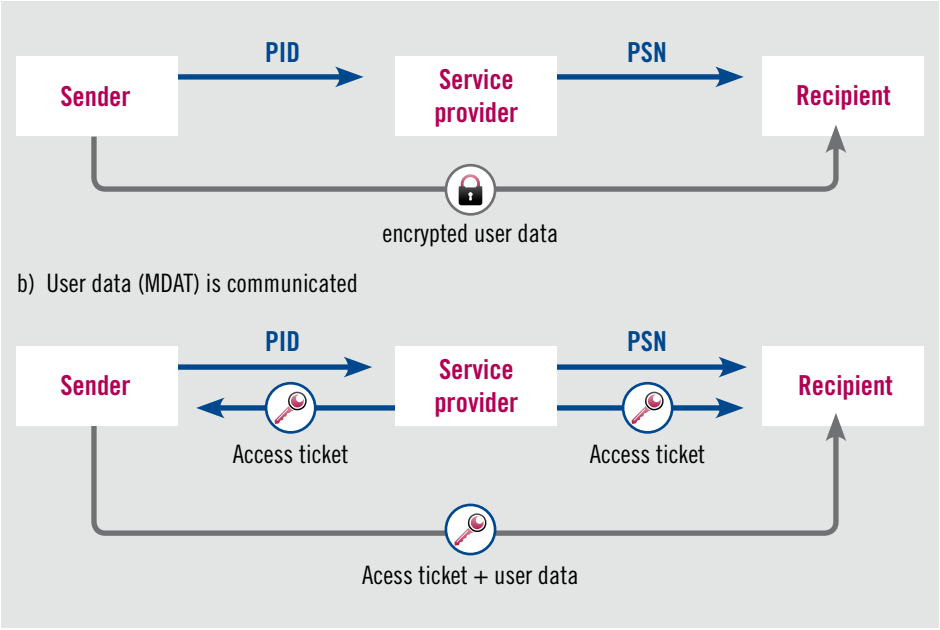
The ticket for access will be stored temporarily in the data set together with the PID but not in the patient database.

The practitioner can access the requested information from the treatment database using the ticket for access.

The ticket for access will be deleted in the treatment database after the practitioner's request or when the request has timed out.

6.3.4. Repseudonymizing by pseudonymizing service (custodian) when exporting data to research module

The goal of the pseudonymization service is to provide particular protection for data in a research database created for long-term storage. This is done by using a **cryptographic procedure** to transform the PID from the patient list into a pseudonym PSN that is used as identification in the research database. As the pseudonymization service will never need or even see the medical data (MDAT) again, these are transferred using **asymmetrical encryption**.



Pseudonymization is a purely machine-based procedure that requires no staff intervention. Data can only be accepted from authorized senders.

Notes