Issues of Verifying Anonymity: An Overview

Sarah Stummer[1]

Abstract: The processing of anonymous data has great potential for research, economy, and society. However, due to legal ambiguities regarding anonymization and the verification of anonymity, its potential is not exploited yet. This paper addresses the chances and issues regarding the concept of anonymization and discusses what needs to be clarified to enable organizations to verify whether data are anonymized in a GDPR-compliant way. Based on this discussion, the paper proposes the implementation of a metrics system as a possible preliminary solution to solve the issues with anonymization and the verification of anonymity.

Keywords: Anonymity, Anonymization, De-anonymization, GDPR, Personal data.

1 Introduction

Over the last years, technology advanced rapidly, and the world became more connected as well as data-driven. Connecting to friends, family members or colleagues via social media, instant messaging services or web conferencing tools, cooking and eating dinner with smart devices assisting or training with a smart watch recording the heart rate, training-progresses etc. – all these and numerous more activities nowadays include technology which allows to connect with other people, systems or applications. As a result of the technological advancement and increasing connectivity, a large amount of data are processed and data are becoming more and more significant.

However, even though the data are required to offer and improve the mentioned services, and therefore, required to serve the interests of individuals, they often refer to a certain natural person and thus are personal data. Hence, the technological advancement and increasing connectivity can also pose privacy risks for individuals. In particular, individuals, their behavior and interests as well as their connection to other individuals become traceable. This can lead, e.g., to identity theft, financial or social disadvantages [Da18], social engineering as well as decreasing personal privacy.

As there is such a tension between the significance of data and the added value data create on one hand and privacy risks that are posed by the use of personal data on the other hand,

[1] Fraunhofer Institute for Secure Information Technology SIT | ATHENE – National Research Center for Applied Cybersecurity, Cloud Computing, Identity & Privacy, Rheinstr. 75, 64295 Darmstadt, Germany, sarah.stummer@sit.fraunhofer.de, https://orcid.org/0000-0003-4015-4429.

doi:10.18420/inf2022_17

the European legislator tightened legal obligations in connection to the processing of personal data as well as monetary consequences in the event of an infringement of data protection obligations. As a result, companies must comply with multiple data protection obligations in order to process personal data and make use of the added value provided by data processing. In addition, limits are set insofar as the processing of personal data under certain conditions is not permitted.

To create a balance between the added value created by data processing and the protection of individual privacy as well as to evade the scope of data protection obligations, where natural persons do not need to be identified or identifiable, data anonymization and the processing of information that not (or no longer) relate to an identified or identifiable natural person ('anonymous data') may be a measure. However, due to ambiguities and issues with the concept of anonymization, its potential is not exploited, and the added value of anonymization is not used in practice yet. This paper discusses the chances and issues regarding the concept of anonymization and points out what needs to be clarified to anonymize data and verify anonymity in a GDPR-compliant way. Based on this discussion, the paper proposes the implementation of a metrics system as a possible solution to enable organizations to assess anonymity and thus, to create a positive sum between privacy and the significance of data.

2 Related Work

Anonymization is the subject of multiple publications in the technical and legal literature. This reaches from publications on the relevance of anonymization to publications on the legal and technical analysis of the terms 'anonymous' and 'anonymization' to issues regarding anonymization and anonymization techniques.

Relevance of anonymization Since anonymization is considered a measure to reduce the data subjects' risk for harms [An19], it is relevant for general data protection. Besides that, the potential of anonymization is seen e.g. for research projects and business models not requiring references to natural persons [Bu20], [Gi21], especially for social science surveys, statistical analyses, and the development of new services or products [Ar14]. A concrete application scenario, where anonymization can have an added value (inter alia by enabling the identification of needs, e.g., for medical care or parking spaces in residential areas) is the smart city [ST21a].

Concept of anonymization From a legal point of view the terms 'anonymous' and 'anonymization' are analyzed and discussed in multiple publications, especially with regards to whether the term 'anonymous' is to be understood absolute or relative (e.g. in [Bi20], [Gi21], [Ma16]). Additionally, multiple supervisory authorities have commented on anonymization and its legal requirements ([Eu20], [Ar14], [An19]). From an

interdisciplinary perspective, [Hö19] addresses differences and similarities in the legal and technical concept of anonymization as well as their implications for practice. Accordingly, the legal and technical concept of anonymization especially differ in the group of persons for whom data must be anonymous: Whereas from a legal perspective data must be anonymous for everyone, including the controller, from a technical perspective data must only be anonymous for external attackers. [AE20] propose a procedure for assessing identifiability, respectively, anonymity by the consideration of the five risk dimensions according to the Five Safes framework (safe projects, safe people, safe data, safe settings and safe outputs [DRW16]).

Issues with anonymization Especially the legal discussions regarding anonymization often lead to the conclusion that the terms and their conditions are not specific enough for anonymization being used in practice [WBH19], [MBH18], [ST21b]. Reasons for this are seen, inter alia, in the expansion of the term 'personal data' on data that initially do not appear to be personal [Oh10], missing processes for anonymization and a lack of ability to assess whether data are anonymous [ST21b]. Moreover, there are multiple publications regarding insufficient anonymization and de-anonymization (e.g. [NS08], [MHVB13], [Bo13]). For instance, [MHVB13] watched geolocation data of 1,5 million people and proved the possibility of identifying 95% of the people observed with only four location-time-points. As a result, for organizations it is very hard to determine when data are actually anonymized.

3 Chances of Anonymization

When personal data are processed, privacy regulations, such as the GDPR, become applicable. As a result, controllers must comply with data protection obligations. Under the GDPR, these obligations especially include the principles relating to the processing of personal data stated in Art. 5 (1) GDPR. Accordingly, the processing of personal data must be transparent and fair and requires a legal basis in order to be lawful. Additionally, the processing of personal data only is allowed for specified, explicit, and legitimate purposes and should be adequate, relevant, and limited to what is necessary for achieving the purposes of the processing. Further processing only is allowed in a manner that is compatible with those purposes the personal data originally were collected for.

Following from the principles of data protection in terms of Art. 5 (1) GDPR, it is not always lawful to process personal data. This may be the case when the purpose of the desired processing cannot be considered fair, for instance when personal data collected by a smart voice assistant (e.g. search requests, shopping lists, and behavior of a certain user) are to be disclosed to other parties (e.g. manufacturers) to show the demand for specific products (e.g. medicine). This may also be the case when the controller desires to process

data for another purpose than what they were collected for, for instance for research purposes, for the training of an artificial intelligence or for the improvement of the collectors' products and services, and/or because reference to a natural person is not required to achieve the desired purposes. Also, the processing of personal data could be unlawful due to the lack of a legal basis for processing.

To counter these legal barriers and enable organizations to achieve their desired purposes, when reference to a natural person is not required, anonymization and the processing of anonymous data might be a suitable solution.

Anonymization is the process of rendering personal data anonymous, so that the reference to a natural person of the data is removed [PP21]. Therefore, it enables organizations to make use of data without having to comply to data protection obligations and without posing a risk for privacy. Additionally, it potentially enables the processing of data for purposes for which it could not be processed if it was personal data. In this context, anonymization might also be an alternative to erasure (affirmative: [ESS19], [St20]; negative: [Ro21]).

The potential of anonymization is also seen in practice. This in particular is shown by an empirical interview of 30 smart city participants, from which 28 interviewees stated that anonymous data can create added value for smart cities [ST21a]. Such added value can be reached, for instance, by

- data exchange between public agencies and social networking platforms (e.g. to improve the responds to disaster alerts),
- data sharing between e.g. educational institutions, public health departments, and health care providers (e.g. to assist in the mitigation of spreading contagious diseases) and
- data exchange between health care providers, research institutions, and civil service offices (e.g. "to identify areas in communities where certain diseases are occurring more frequently, so that the cause can be researched and help can be offered") [ST21a].

Moreover, anonymization becomes increasingly important in research and economy [Bu20], [Gi21], e.g. for training, testing and evaluating artificial intelligence [RG21].

4 Issues regarding the Concept of Anonymization

Even though anonymization has great potential and can be an appropriate solution to create a balance between privacy and data use, there are certain issues with the concept of anonymization – especially regarding the verification of anonymity – that must be concretized for organizations to make use of its potential. These issues will be discussed in the following.

4.1 Boundaries between Personal Data and Anonymous Data

Firstly, there is an obscurity in the boundaries between personal and anonymous data as well as the allocation of data as personal or anonymous.

Legal definitions of the GDPR

According to Art. 4 (1) GDPR personal data are any information relating to an identified or identifiable natural person (so called 'data subject'), whereas an identifiable natural person is one who can be identified, directly or indirectly. Hence, for data to be considered personal data it is not required that the data subject is identified directly through direct identifiers, such as passport numbers, social security numbers, unique customer numbers or – depending on the context – names and addresses [EM15] (identifying personal data). Rather, the possibility of identifying a natural person through so-called quasi-identifiers is sufficient as well (identifiable personal data). Quasi-identifiers are datasets consisting of attributes which individually do not identify a natural person but in their combination become identifying. For instance, information on the date of birth, gender, and zip codes in their combination can be considered a quasi-identifier, depending on the context [Sw00].

A special form of personal data are pseudonymized data. The term pseudonymized data is not directly defined within the GDPR, however, from the definition of pseudonymization according to Art. 4 (5) GDPR it can be concluded that pseudonymized data according to the GDPR are personal data that can no longer be attributed to a specific data subject without the use of additional information, which is kept separately and is subject of technical and organizational measures.

Regarding anonymous data the European legislator did not introduce a definition. Instead, the term only is mentioned in the Recitals of the GDPR, especially in Recital 26. Accordingly, the principles of data protection only apply to information concerning an identified or identifiable natural person and thus, are not applicable to anonymous data. Hence, anonymous data are differentiated from personal data insofar as anonymous information is information not relating to an identified or identifiable natural person or

personal data rendered anonymous so that the data subject no longer is identifiable. Anonymous data therefore can be understood as the opposite of personal data [Ro21].

Differences between personal data and anonymous data

Based on the legal definitions of the terms ‘anonymous data’ and ‘personal data’, it becomes clear that the terms differ in their ability to identify a natural person as well as in their degree of identifiability. In this regard, the boundaries between the terms seem to be clear: Personal data generally allow to identify a natural person and therefore, exhibit a higher degree of identifiability than anonymous data, where the identification of a specific natural person is not possible (anymore). However, as shown in Figure 1, with regards to personal data, it needs to be further distinguished between identifying personal data and identifiable personal data. Since identifying personal data allow the identification of a natural person directly, whereas identifiable personal data only allow identification in combination with other information and anonymous data do not allow identification at all, the degree of identifiability increases with each layer shown in Figure 1.

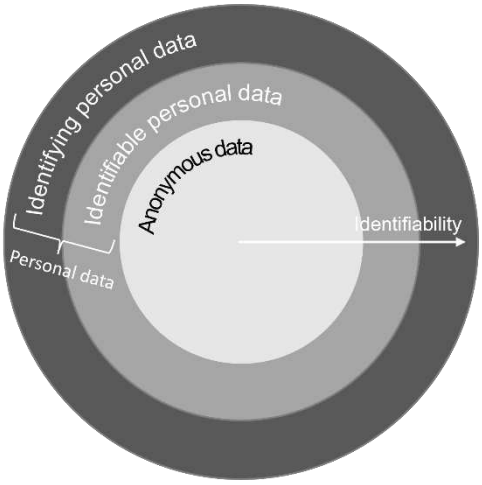

Figure 1: Abstract view on the boundaries between personal data and anonymous data as well as their sub-categories based on their ability to identify a specific natural person.

Assessing whether data are personal data or anonymous data

Though, the legal definitions and the differences in the terms are clear, the boundaries between the terms are blurring, when it comes to assessing whether data refer to an identified or identifiable natural person and determining if a dataset is anonymous or not.

First, whether personal data are identifying or identifiable depends on the specific context, especially the size of the group of eligible natural persons (‘anonymity set’ [PKS05]) [Ar07]. For instance, in a group of 50 natural persons let the persons‘ eye colors be given and distributed like this: 10x green, 15x blue, and 25x brown. In this case, the attribute ‘blue eyes’ alone would not identify one specific natural person but would be applicable to fifteen persons. Hence, for identifying a specific natural person, the combination of the eye color with other attributes – such as date of birth, hair color or initials – would be required. However, if the group would only consist of 7 individuals, with only one person having blue eyes (e.g. 2x green eyes, 4x brown eyes, 1x blue eyes), the attribute ‘blue eyes’ would directly identify a natural person and thus, would qualify as identifying personal data. Therefore, when assessing whether data are directly or indirectly identifying, the context of the processing, especially with regards to the anonymity set, must be taken into account [Ar07]. A categorization of data as identifying personal data or identifiable personal data based solely on the data is not sufficient.

It becomes even more complex when it further is assessed whether data are personal data at all or if they qualify as anonymous data. Especially the boundary between anonymous data and identifiable personal data is blurring. This is because pursuant to Recital 26 GDPR different aspects need to be considered when assessing whether a natural person is identifiable and thus, data are personal or anonymous. Accordingly, when determining whether a natural person is identifiable, all the means reasonably likely to be used by the controller or by another person to identify the natural person should be taken into account. The likelihood of means used is depending on all objective factors, such as the costs, the amount of time as well as available identification-technologies. Hence, instead of a clear distinction between identifiable personal data and anonymous data, the European legislator provides different factors that must be considered when determining whether data have a reference to a natural person or are anonymous. As a result, from a legal point of view, it is unclear when data are anonymous. Instead, there is a fine line between (identifiable) personal data and anonymous data, without a specification on where exactly this line runs (see Figure 2).

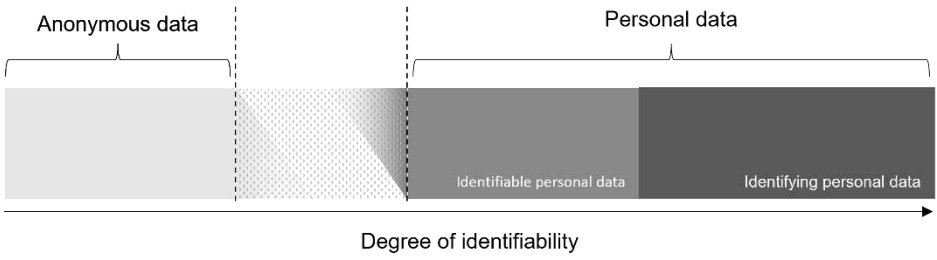

Figure 2: Blurring boundaries between anonymous data and personal data.

In order to create legal certainty for the organizations anonymizing data and to enable them to make use of anonymization and its potential to create added value, it therefore must be concretized where the line between personal data and anonymous data runs, which requirements must be met to consider data sufficiently anonymous and thus, under which conditions data are not personal data but qualify as anonymous.

Issues to solve:

Where does the boundary between anonymous data and personal data lie? When are data anonymous?

Which requirements must be met to consider data sufficiently anonymous?

4.2 Possibility of Anonymity

Even though it is not conclusively resolved where the boundary between anonymous data and personal data lies and when data are anonymous from a legal point of view, anonymity and techniques to render personal data anonymous (anonymization techniques) are discussed sufficiently.

Anonymization techniques

From a technical perspective, there are numerous techniques to anonymize data. The foundation for anonymization generally is the removal or masking of direct identifiers to ensure that direct identification is not possible anymore. However, to fully anonymize data, it furthermore requires that indirect identification of a natural person is ruled out practically. Therefore, additional techniques to generalize and/or randomize data must be applied to such an extent that natural persons are not identifiable anymore. Through generalization, data are abstracted and generalized in such a way that singling out a natural person becomes less likely. Through randomization, data are distorted in such a manner that their connection to a natural person is removed [Ar14].

Insufficient anonymization and de-anonymization

However, as multiple studies have shown, if anonymization techniques are applied naively or insufficiently, identifying natural persons may still be possible and thus, the resulting data may not be anonymous. For instance, [NS08] analyzed attributes of Netflix' customer database with more than 100 million movie ratings from a scale of 1 to 5. Except for the rating and its point of time, every identifier was removed. Additionally, randomization was conducted. However, the researchers still were able to identify 99% of the records within the dataset with only eight ratings and the dates of their delivery (with a 14-day error). 68% of the users even could be identified with only two ratings and the time of

their delivery, known to within three days. Another example for the identifiability of data is given by [MHVB13], who watched geolocation data of 1,5 million people for a period of 15 months and found out that it requires only four location-time-points to identify 95% of the people observed.

The possibility of identifying a natural person is amplified by the technological advancement and interconnectivity over the internet – both resulting in an increasing amount and accessibility of data. For instance, when contacting friends, family members or other contacts, in multiple cases, direct messenger services, such as WhatsApp or Instagram Direct Messaging, are used. Additionally, personal meetings and leisure activities often are recorded by phone cameras and shared – potentially connected to a geolocation – with other individuals via social media. Another example are household activities: An increasing number of households use smart home devices, such as vacuum robots, smart refrigerators, smart washing machines and voice assistants (e.g. Alexa), which support by turning the light on or setting timers but also by looking for recipes, answering questions or ordering goods. All these and numerous more activities process data and therefore result in an increasing amount and accessibility of data allowing inferences to individuals. The question therefore is whether in an era of big data, where by the use of services a large amount of data is collected, information is shared in the internet and a digital trace is left by individuals, anonymity is even possible.

Hence, it needs to be assessed whether the legal requirements for anonymity, especially due to the increasing accessibility of data, can be met and thus, whether sufficient anonymity is possible from a legal point of view. In addition to the question of whether anonymity generally is possible, it needs to be discussed whether and to what extent the (loss of) value of data can be considered when it comes to assessing anonymity. Usually, the more the data are generalized and randomized, the less value they have. However, since anonymization is not only carried out to increase privacy, but also to use the anonymized data for certain purposes, the resulting data must be meaningful and sufficient to fulfill the desired purposes. Otherwise, the anonymized data loses their value, and their processing becomes useless. Thus, it is questionable whether anonymity is possible without losing the data's information value and to what extent such considerations can be taken into account when it comes to anonymization and verifying anonymity (e.g. by requiring anonymization only to such an extent to which the resulting data still are useful).

Issues to solve:

Is anonymity possible from a legal point of view?

Can loss of information value be considered when assessing anonymity, so that data must only be anonymized to an extent to which they still are useful?

4.3 Relevance of the Context and Risk

Connected to the possibility of anonymity and the blurring boundaries between anonymous and personal data, it is also questionable whether and to what extent the context of the processing as well as the risk that is connected to the processing can be considered when assessing anonymity.

For instance, when an anonymized dataset is only to be used internally within the anonymizing organization and appropriate measures to ensure confidentiality are to be taken, it is also only possible for the anonymizing organization to try carrying out de-anonymization. In this case, perhaps less means to identification need to be considered when determining whether data are anonymous or not. The same applies for cases, in which data is collected in a form that is anonymous for the data controller, even though it cannot be ruled out that other parties could be able to identify a natural person.

An argument for considering risk and context is the jurisdiction of the European Court of Justice. Accordingly, all legal means enabling a party to identify a natural person [Eu16] (including the involvement of third parties which have the information on identity and are legally compelled to provide this information [Ge17]) must be considered when assessing whether data are personal or anonymous. From this follows that only the lawful means of the parties having access to the anonymized dataset must be taken into account. Another argument for considering risk and context are the requirements of the GDPR in connection to technical and organizational measures, especially the GDPRs' risk-based approach. Art. 24, 25, 32 GDPR require the controller to take technical and organizational measures in order to implement data protection principles stated in Art. 5 (1) GDPR as well as to ensure an appropriate level of security. When taking such measures, the controller shall, inter alia, take the risk as well as the context and purposes of processing into account. Since data anonymization removes the data's reference to a natural person and therefore ensures data minimization in terms of Art. 5 (1) (c) GDPR, it serves as a measure to implement data protection principles [Go18].

However, in contrast to this is that data anonymization results in data protection obligations not being applicable anymore. Thus, it is questionable whether such a risk- and context-based approach can be followed (affirmative e.g. [Ag21], [An19]; negative e.g. [PP21]). This consideration also is connected to the dispute of whether anonymity must be absolute or relative. According to the absolute theory of identifiability, data only can be considered anonymous if it is impossible for anyone to identify a natural person. The relative theory of identifiability, on the other hand, considers data anonymous when identification due to all objective factors, such as the costs and the amount of time required for identification, reasonably is not likely (even though theoretically it was possible) [SHS19]. Regarding this dispute, the GDPR is not unambiguous. In contrast, Recital 26 GDPR contains elements of the absolute theory as well as elements of the relative theory:

In favor of the relative theory of identifiability is that for the determination of whether a natural person is identifiable, only the means reasonably likely to be used for the identification of natural person should be taken into account. On the other hand, the GDPR addresses all the means used by the controller or another person which is characteristic for the absolute theory of identifiability. Thus, the European legislator seems to generally support the relative theory but restricts it through elements of the absolute theory [SHS19], [Ma16].

Issues to solve:

Can a risk- and context-based approach be followed when assessing anonymity?

(To what extent) must anonymity be absolute or relative?

4.4 Consequences of De-Anonymization

As already mentioned, anonymity is not static but can change dynamically [Ma16], [HW19]. Hence, in the past there have been multiple examples where alleged anonymity has been reversed (so called 'de-anonymization'), so that conclusions to a specific natural person have been possible (again) and natural persons became (re-)identifiable.

De-anonymization can be caused by anonymization techniques being applied naively or insufficiently, by new, faster or cheaper technologies [HW19], [HW20] as well as by increased accessibility of data [Bo13].

As a result of de-anonymization, data becomes personal data (again) and data protection obligations must be fulfilled when the data are processed. However, since de-anonymization can also result in natural persons being identifiable rather than identified (for instance in cases where in an insufficient anonymized dataset only a few natural persons are identified and it cannot be ruled out that more natural persons are identifiable), data protection obligations which require the identification of a natural person cannot always be fulfilled. The questions therefore are, what the consequences of insufficient anonymization or de-anonymization are and whether and to what extent legal obligations of the GDPR must be fulfilled when natural persons become re-identifiable.

Issue to solve:

Must data protection obligations be fulfilled in consequence of data-de-anonymization and to what extent?

5 Approach for a solution

As we have seen, there are multiple open issues regarding the concept of anonymization, all of which relate to clarifying whether a particular dataset is anonymous or not, and how to deal with a dataset that was assumed to be anonymous, but which has been found as being personal data by a detailed analysis of the dataset. As a result of these issues, it can hardly be determined whether a dataset is anonymous. The chances of anonymization therefore are not exploited in practice yet, even though many important added values for society and economy could be achieved.

To enable organizations to assess and verify anonymity of a dataset and hence, enable them to exploit the chances of anonymization, one possible way could be the development and use of a metrics system. Metrics are mathematical functions that allow to systematically measure the adherence to requirements and to compare the as-is state with the target state [AS12]. Metrics therefore could also be used to assess and verify whether the requirements for anonymity are fulfilled by an anonymized dataset.

The development and use of a metrics system for the assessment and verification of anonymity, however, first requires the mentioned issues to be solved. Therefore, anonymity and its requirements must be analyzed from an interdisciplinary view considering law and technology. In particular, it must be assessed under which conditions means for identification are reasonably likely to be used, to what extent a risk- and context-based approach can be followed when the likelihood of means to be used for identification is assessed and where the boundary between personal data and anonymous data lies. Based on the findings of the analysis, the legal requirements for anonymity become more concrete and verifiable requirements and measures for anonymization can be derived. These verifiable measures then serve as the foundation for a set of metrics, through which anonymity can be assessed and verified.

Since anonymity can change dynamically, organizations furthermore should be provided with a guideline addressing the privacy-compliant handling of non-anonymous data. Therefore, the legal consequences of de-anonymization need to be analyzed (possibly depending on the use case) and a procedure for organizations to handle non-anonymous data needs to be elaborated.

This contribution therefore proposes the following approach:

1. Answering the mentioned issues regarding the concept of anonymization with an interdisciplinary perspective considering law and technology.
2. Deriving verifiable metrics from the results of (1).

3. Implementing a metrics system to verify the degree of anonymity of a dataset as well as a guideline that shows the privacy-compliant handling of non-anonymous data, based on the results of (2).

6 Conclusion and Outlook

As there are multiple issues with the determination of anonymity, which are not sufficiently clarified yet, even though anonymization has great potential for society, research, and economy and is increasingly important for the privacy of individuals, its potential is not exploited yet.

To enable organizations to benefit from the added value anonymization and the processing of anonymous data can create, the issues mentioned in this paper need to be clarified and a process for verifying anonymity must be elaborated. Initial indications for solving some of the mentioned issues could be provided by the position guideline for anonymization and pseudonymization, which the European Data Protection Board announced for 2021/2022 [Eu21]. Additionally, I will work on answering the mentioned issues with an interdisciplinary perspective considering law and technology to elaborate a metrics system for the assessment and verification of anonymity.

This research work has been funded by the German Federal Ministry of Education and Research and the Hessen State Ministry for Higher Education, Research and the Arts within their joint support of the National Research Center for Applied Cybersecurity ATHENE.

References

[AE20] Arbuckle, Luk; El Emam, Khaled: Building an Anonymization Pipeline. Creating Safe Data, O'Reilly Media, 2020.

[Ag21] Agencia Española de Protección de Datos: 10 Misunderstandings related to Anonymisation. At: https://edps.europa.eu/system/files/2021-04/21-04-27_aepd-edps_anonymisation_en_5.pdf, 2021.

[An19] An Coimisiún um Chosaint Sonraí: Guidance Note: Guidance on Anonymisation and Pseudonymisation. At: https://www.dataprotection.ie/sites/default/files/uploads/2020-09/190614%20Anonymisation%20and%20Pseudonymisation.pdf, 2019.

[Ar07] Article 29 Data Protection Working Party: Opinion 4/2007 on the concept of personal data (WP 136). At: https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2007/wp136_en.pdf, 2007.

[Ar14] Article 29: Data Protection Working Party: Opinion 05/2014 on Anonymisation Techniques (WP 216), Brüssel, 2014.

[AS12] Ammann, Franz-Ernst; Sowa, Aleksandra: Systematische Entwicklung von Metriken zur Beurteilung der Datensicherheit. In: Datenschutz und Datensicherheit (DuD), p. 247-251, 2012.

[Bi20] Bischoff, Claudia: Pseudonymisierung und Anonymisierung im Rahmen klinischer Prüfungen von Arzneimitteln (Teil I). In: Pharma Recht (PharmR), p. 309-315, 2020.

[Bo13] Bohannon, John: Genealogy Databases Enable Naming of Anonymous DNA Donors, Science, 339 (6117), p. 262, 2013.

[Bu20] Der Bundesbeauftragte für den Datenschutz und die Informationsfreiheit: Positionspapier zur Anonymisierung unter der DSGVO unter besonderer Berücksichtigung der TK-Branche. Bonn, 2021.

[Da18] Datenschutzkonferenz: Kurpapier Nr. 18 Risiko für die Rechte und Freiheiten natürlicher Personen. At: https://www.datenschutzkonferenz-online.de/media/kp/dsk_kpnr_18.pdf, 2018.

[DRW16] Desai, Tanvi; Ritchie, Felix; Welpton, Richard: Five Safes: designing data access for research. Economics Working Paper Series. University of the West of England, Bristol, England. Faculty of Business and Law, 2016.

[EM15] El Emam, Khaled; Malin, Bradley: Sharing Clinical Trial Data: Maximizing Benefits, Minimizing Risk. Washington D.C.: National Academies Press, 2015.

[ESS19] Enzmann, Matthias; Selzer, Annika; Spychalski, Dominik: Data Erasure under the GDPR – Steps towards Compliance. In: European Data Protection Law Review (EDPL), p. 416-420, 2019.

[Eu16] European Court of Justice, C-582/14, 19.10.2016.

[Eu20] European Data Protection Board: Guidelines 04/2020 on the use of location data and contact tracing tools in the context of the COVID-19 outbreak. At: https://edpb.europa.eu/sites/default/files/files/file1/edpb_guidelines_20200420_contact_tracing_covid_with_annex_en.pdf, 2020.

[Eu21] European Data Protection Board: EDPB Work Programme 2021/2022. At: https://edpb.europa.eu/system/files/2021-03/edpb_workprogramme_2021-2022_en.pdf, 2021.

[Ge17] German Federal Court of Justice, VI ZR 135/13, 16.05.2017.

[Gi21] Gierschmann, Sybille: Gestaltungsmöglichkeiten durch systematisches und risikobasiertes Vorgehen – Was ist schon anonym? Planung und Bewertung der Risiken der Anonymisierung. In: Zeitschrift für Datenschutz (ZD), p. 482-486, 2021.

[Go18] Gola, Peter: Datenschutz-Grundverordnung. 2. Auflage, C.H. Beck, München, 2018.

[Hö19] Hölzel, Julian: Differential Privacy and the GDPR. In: European Data Protection Law Review (EDPL), p. 184-196, 2019.

[HW19] Hornung, Gerrit; Wagner, Bernd: Der schleichende Personenbezug: Die Zwickmühle der Re-Identifizierbarkeit in Zeiten von Big Data und Ubiquitous Computing. In: Computer und Recht (CR), p. 565-574, 2019.

[HW20] Hornung, Gerrit; Wagner, Bernd: Anonymisierung als datenschutzrelevante Verarbeitung? Rechtliche Anforderungen und Grenzen für die Anonymisierung personenbezogener Daten. In: Zeitschrift für Datenschutz (ZD), p. 223-228, 2020.

[Ma16] Marnau, Ninja: Anonymisierung, Pseudonymisierung und Transparenz für Big Data: Technische Herausforderungen und Regelungen in der Datenschutz-Grundverordnung. In: Datenschutz und Datensicherheit (DuD), p. 428–433, 2016.

[MBH18] Marnau, Ninja; Berrang, Pascal; Humbert, Mathias: Anonymisierungsverfahren für genetische Daten, In: Datenschutz und Datensicherheit (DuD), p. 83-88, 2018.

[MHVB13] de Montjoye, Yves-Alexandre; Hidalgo, César A.; Verleysen, Michel; Blondel, Vincent D.: Unique in the Crowd: The privacy bounds of human mobility. In: Scientific Reports 3:1376, p. 1-5, 2013.

[NS08] Narayanan, Arvind; Shmatikov, Vitaly: Robust de-anonymization of large sparse datasets. In IEEE Symposium on Security and Privacy, p. 111-125, 2008.

[Oh10] Ohm, Paul: Broken promises of privacy: Responding to the surprising Failure of Anonymization. In: UCLA Law Review (57), p. 1701–1777, 2010.

[PKS05] Pfitzmann, Andreas; Köhntopp, Marit; Shostack Adam et. al: Anonymity, Unlinkability, Unobservability, Pseudonymity, and Identity Management – A Consolidated Proposal for Terminology. At: https://dud.inf.tu-dresden.de/literatur/Anon_Terminology_v0.22.pdf, 2005.

[PP21] Paal, Boris P.; Pauly, Daniel A.: Datenschutz-Grundverordnung Bundesdatenschutzgesetz. 3. Auflage, C.H. Beck, München, 2021.

[RG21] Roßnagel, Alexander; Geminn, Christian L.: Vertrauen in Anonymisierung. Regulierung der Anonymisierung zur Förderung Künstlicher Intelligenz. In: Zeitschrift für Datenschutz (ZD), p. 487-490, 2021.

[Ro21] Roßnagel, Alexander: Datenlöschung und Anonymisierung – Verhältnis der beiden Datenschutzinstrumente nach DS-GVO. In: Zeitschrift für Datenschutz (ZD), p. 188-192, 2021.

[SHS19] Simitis, Spiros; Hornung, Gerrit; Spiecker gen. Döhmann, Indra: Datenschutzrecht. 1. Auflage, C.H. Beck, 2019.

[St20] Stürmer, Verena: Löschen durch Anonymisieren? In: Zeitschrift für Datenschutz (ZD), p. 626-631, 2020.

[ST21a] Selzer, Annika; Timm, Ingo: Chances and Limitations of Personal and Anonymized Data Processing. In: INFORMATIK 2021. Gesellschaft für Informatik, Bonn, p. 773-787, 2021.

[ST21b] Selzer, Annika; Timm, Ingo: Potenziale anonymer Datenverarbeitungen nutzen. In: Datenschutz und Datensicherheit (DuD), p. 816-820, 2021.

[Sw00] Sweeney, Latanya: Simple Demographics Often Identify People Uniquely. In: Carnegie Mellon University, Data Privacy Working Paper 3. Pittsburgh, 2000.

[WBH19] Winter, Christian; Battis, Verena; Halvani, Oren: Herausforderungen für die Anonymisierung von Daten. Technische Defizite, konzeptuelle Lücken und rechtliche Fragen bei der Anonymisierung von Daten. In: Zeitschrift für Datenschutz (ZD), p. 489-493, 2019.