Personal Data and Encryption in the European General Data Protection Regulation

by **Gerald Spindler and Philipp Schmechel***

Abstract: Encryption of personal data is widely regarded as a privacy preserving technology which could potentially play a key role for the compliance of innovative IT technology within the European data protection law framework. Therefore, in this paper, we examine the new EU General Data Protection Regulation's relevant provisions regarding encryption – such as those for anonymisation and pseudonymisation – and assess whether encryption can serve as an anonymisation technique, which can lead to the non-applicability of the GDPR. However, the provisions of the GDPR regarding the material scope of the Regulation still leave space for legal uncertainty when determining whether a data subject is identifiable or not. Therefore, we inter alia assess the Opinion of the Advocate General of the European Court of Justice (ECJ) regarding a preliminary ruling on the interpretation of the dispute concerning whether a dynamic IP address can be considered as personal data, which may put an end to the dispute whether an absolute or a relative approach has to be used for the assessment of the identifiability of data subjects. Furthermore, we outline the issue of whether the anonymisation process itself constitutes a further processing of personal data which needs to have a legal basis in the GDPR. Finally, we give an overview of relevant encryption techniques and examine their impact upon the GDPR's material scope.

Keywords: GDPR; Encryption; Anonymisation; Pseudonymisation; Personal Data; Material Scope; ECJ; Advocate General; Data Protection; Secure Multiparty Computation

Recommended citation: Gerald Spindler and Philipp Schmechel, Personal Data and Encryption in the European General Data Protection Regulation, 7 (2016) JIPITEC 163 para 1.

A. Introduction

1 Seventeen years ago, *Lawrence Lessig* wrote that "encryption technologies are the most important technological breakthrough in the last one thousand years".[1] This might be a slight exaggeration, but it emphasises the importance of encryption technologies in today's digital world. Encrypted data plays a significant role in the protection of data subjects' privacy. Its legal problems are closely related to the scope of the data protection laws and the legal effects of anonymisation and pseudonymisation.

2 Encrypting personal data is becoming increasingly important for many business models in a data-driven economy and for preserving data subjects' privacy with regard to today's monitoring and profiling possibilities – both of government institutions and of high-tech companies. Be it for the processing of sensitive health data, for the *Internet of Things* or for connected cars, for the privacy preserving use of Big Data or cloud computing technologies[2], encryption can be a key to protect an individual's privacy and

1 *Lessig*, Code and Other Laws of Cyberspace, 1999, p. 35.

2 See e.g. the PRACTICE project, funded by the EU-FP7-programme, which aims to build a secure cloud framework that allows for the realization of advanced and practical cryptographic technologies, <https://practice-project.eu/>.

can make several IT innovations possible, which would otherwise conflict with the data protection framework. For many years, the discussion about the material scope of the European Data Protection Directive[3] (DPD) and about the exact definition of personal data and the interpretation of the term "identifiable" has been one of the "key issues"[4] of European data protection law. [5] Additionally, the legal effects of encrypted data for the applicability of data protection law and for the personal references of data have still not sufficiently been examined. These questions regarding personal data and encryption once again occur in the new EU General Data Protection Regulation[6] (GDPR).

3 Encrypting personal data and deleting any personal reference from the data could also be a way to work with this information when it is transferred to third countries outside of the EU.[7] EU standard contractual clauses or compliance to the new Privacy Shield when transferring data to the U.S. would therefore not be necessary if the data lost all of its personal reference. However, legal uncertainty concerning whether the encryption of personal data has the effect that such data loses its personal reference or not may discourage controllers to use these privacy preserving measures. Thus, in this article we will examine the legal effects of encryption in regards to the applicability of the GDPR.[8]

4 The GDPR only applies if "personal data" is processed. Thus, the notion of *personal* data is crucial for the application of the GDPR. Depending on how "personal data" is defined and interpreted, the effect a valid encryption of this data takes may be different. Furthermore, we will examine how encrypted data is treated in the GDPR – as anonymised or pseudonymised data – and where and how in the GDPR encryption can be used as a technical and organisational measure.[9] With regard to some important encryption tools for the transport, storage and processing of personal data we will demonstrate the effect of encryption on the material scope of the GDPR.[10]

B. The General Data Protection Regulation and Encryption

5 After years of intensive negotiations, the GDPR has now been passed and will finally come into force from 25 May 2018 (see Article 99 Par. 2 GDPR) and will, according to Article 94 Par. 1 GDPR, repeal the old Directive 95/46/EC.[11] Due to its legal form of as a Regulation, the GDPR will be binding in its entirety and will be directly applicable in all Member States of the European Union.[12] We will examine the material scope of the GDPR and the effect of encryption on personal data.

6 The importance for controllers of knowing the exact scope of the Regulation and whether the data they process will be considered as personal data or not, e.g. due to the use of encryption, increases with the GDPR's very broad territorial scope, especially the rules for controllers not established in the EU will be changed dramatically.[13] According to Article 3 Par. 1 the GDPR "applies to the processing of personal data in the context of the activities of an establishment of a controller or a processor in the Union, regardless of whether the processing takes place in the Union or not". Moreover, Par. 2 states that the Regulation applies even to the processing of personal data where the processing activities are related to the offering of goods or services or the sheer monitoring of the data subject's behaviour as long as their behaviour takes place within the Union. "Monitoring" means *inter alia* the online tracking of natural persons to create profiles in order to take decisions, for analysing or predicting personal preferences, behaviours and attitudes (see Recital 24 S. 2 GDPR).[14]

3 Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data, Official Journal of the European Communities, L 281, pp. 31-50.

4 *Boehme-Neßler*, Datenschutz und Datensicherheit 2016, p. 419.

5 See e.g. *Article 29 Data Protection Working Party*, Opinion 4/2007 on the concept of personal data, WP 136, pp. 6 et seq.; *Hon/Millard/Walden*, Queen Mary University of London – Legal Studies Research Paper No. 75/2011, pp. 8 et seq., available at: <http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1783577##>, accessed 26 August 2016; *Article 29 Data Protection Working Party*, Opinion 05/2014 on Anonymisation Techniques, WP 216, pp. 5 et seq.

6 Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), Official Journal of the European Union, L 119, pp. 1-88.

7 Cf. *Esayas*, European Journal of Law and Technology, Vol 6, No 2 (2015), p. 13; Commission Decision 2000/520/EC of 26 July 2000, Official Journal of the European Communities L 215/7 (24).

8 See *infra* B.II.2.d.).

9 See *infra* B.II.1.

10 See *infra* B.II.3.

11 See for an overview of the legislative process of the GDPR *Albrecht*, Computer Law Review International 2016, pp. 33 et seq.

12 *Reding*, International Data Privacy Law 2012, p. 119 (121).

13 See *Kindt*, CiTiP Working Paper 26/2016, pp. 13 et seq., available at: <http://papers.ssrn.com/sol3/JELJOUR_Results.cfm?form_name=journalbrowse&journal_id=1781425>, accessed 8 August 2016.

14 Under the DPD, according to Article 4 Par. 1 (c) controllers targeting EU data subjects only had to comply with the DPD if they made use of "equipment" situated in the EU to process personal data.

7 Thus, the GDPR's broad territorial scope leads towards a new awareness of data controllers (also established outside the Union) regarding their processing of personal data. Therefore, technologies which minimise the use of personal data – especially encryption – and which avoid the application of the GDPR become even more important.

I. Personal Data: The Material Scope of the GDPR

8 As already outlined, the characteristics of personal data are crucial for the application of the GDPR. However, the GDPR does not introduce major changes to the concept of personal data in comparison to the DPD. Just like the DPD, the GDPR follows a "black/white approach", hence the data are either personal or not, which means that if the data has a personal reference, all data protection rules apply and if not, it is outside the GDPR's scope.[15] According to Article 2 Par. 1 GDPR

> *"This Regulation applies to the processing of personal data wholly or partly by automated means and to the processing other than by automated means of personal data which form part of a filing system or are intended to form part of a filing system."*

9 Article 4 No. 1 S. 1 GDPR defines that "'personal data' means any information relating to an identified or identifiable natural person[16] ('data subject')" which is the same wording as Article 2 (a) DPD. In this regard, "any information" means virtually *any* information, even publicly available information; when a reference to a natural person can be made the data protection principle of the GDPR always applies regardless of the data's content.[17] However, Article 4 No. 1 S. 2 GDPR introduces a new definition of the concept of an "identifiable natural person", which refers to a person "who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person". Thus, the definition of an "identifiable natural person" distinguishes between identifiability on the basis of a reference to an identifier which can clearly identify a natural person, or due to special personal characteristics such as a person's sexual preferences or medical condition.[18]

10 However, it is still highly controversial whether or not a so-called *absolute* or *relative* approach has to be applied for assessing the data controller`s abilities to identify a natural person.

1. The "Identifiable Natural Person"

11 Crucial to understanding the exact scope of the concept of "personal data" is how much a potential data controller has to do in order to establish a link between a natural person and the data, in other words what efforts are required to identify a person.

a.) Absolute Approach

12 The *absolute* approach takes into account all possibilities and chances in which the data controller would be able to identify the data subject individually. Thus, all ways and means for a data controller without any regard to expenses etc. are taken into account. Even theoretical chances of combining data so that the individual is identifiable are included. If identifiability is assessed *absolutely*, then it is sufficient for the application of personal data acts if *anyone* in the world is able to decrypt or decode the encrypted data.[19]

13 In terms of encryption, as long as *anyone* in the world is able to decrypt the data set, the operations of the controller or processor using this encrypted data are subject to data protection legislation, even if they don't possess the key for decryption. Based on this approach data protection legislation is applicable, regardless of the applied encryption technique, as long as one entity holds the key for decoding.[20]

b.) Relative Approach

14 In contrast, the *relative* approach considers the necessary effort required by the data controller

15 *Forgó*, International Data Privacy Law 2015, p. 54 (59).

16 Like in Article 1 Par. 1 of the DPD, the material scope of the GDPR only applies to the processing of personal data of *natural* persons according to Article 1 Par. 1 GDPR.

17 Cf. *Kranenborg*, in: Peers/Hervey/Kenner/Ward (eds.), The EU Charter of Fundamental Rights, 2014, Art 8, Recital 08.85; *Article 29 Data Protection Working Party*, WP 136 (*supra* Note 5), pp. 6 et seq; *Karg*, Datenschutz und Datensicherheit 2015, p. 520 (521).

18 Cf. *Härting*, Datenschutz-Grundverordnung, 2016, Recital 275 et seq.

19 *Kuner*, European Data Protection Law: Corporate Compliance and Regulation, 2nd Ed. 2007, p. 92; *Pahlen-Brandt*, Datenschutz und Datensicherheit 2008, p. 34 (38); *Nink/Pohle*, Multimedia und Recht 2015, p. 563 (565), who criticize that consequently this approach would lead to the result that there would virtually be no more anonymous data.

20 Cf. *Meyerdierks*, Multimedia und Recht 2009, p. 8 (10).

in order to identify the data subject.[21] Therefore, only realistic chances of combining data in order to identify an individual are taken into account – and not highly theoretical identification risks.[22] With regards to encryption issues, data protection legislation is only applicable if the data *controller* is able to decrypt a certain data set[23] – or, at least has reasonable chances of obtaining the decrypting key. In the case law of some courts, the trend is beginning to lean towards favouring a *relative* understanding.[24]

c.) The GDPR's Approach

15 The GDPR utilises a broad approach regarding the interpretation of "identifiable natural person" however, some terms can also be interpreted in a *relative* way. Additionally, both Article 7 and Article 8 of the Charter of Fundamental Rights of the EU (CFR) always have to be taken into account when interpreting the data subject`s rights.[25]

16 Recital 26 S. 3 of the GDPR states that "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." On the one hand, the Recital refers to means reasonably likely to be used "by another person" which have to be taken into account, which veers towards an *absolute* approach, because this *third person* could be any person in the world.[26] This is also in tune with the scope of Article 8 CFR, according to which "identifiable" has to be interpreted widely.[27]

17 Moreover, stating in Article 4 No 1 S. 2 GDPR that every "identifier" shall contain personal references is another hint for a rather *absolute* approach of the Regulation regarding the identifiability of a natural person.[28] Additionally, Recital 26 states that using means for "singling out" the natural person directly or indirectly may make this person identifiable. Thus, a data subject may now be singled out for data processing even if it is unlikely that his or her name can be tied to the data, because even this could result in harming his or her privacy.[29]

18 On the other hand, the term "means reasonably likely to be used" suggests limitations through *relative* elements, in particular the notion of "reasonably".[30] Additionally, if a zero risk threshold would be applied for any potential data user, no existing technique could achieve the required level of anonymisation.[31] Moreover, according to the *Article 29 Data Protection Working Party* (interpreting the Data Protection Directive), "a mere hypothetical possibility to single out the individual is not enough to consider the person as 'identifiable'".[32]

19 Recital 26 GDPR continues by stating objective factors which shall be relevant for the interpretation of the means used to identify a natural person:

> *"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."*

20 These factors illustrate a further attempt to limit the broad *absolute* elements of the GDPR's material

21 *Roßnagel/Scholz*, Multimedia und Recht 2000, p. 721 (723); *Meyerdierks* (*supra* Note 20), pp. 8 et seq.; *Voigt*, Multimedia und Recht 2009, p. 377 (379); *Lundevall-Unger/Tranvik*, International Journal of Law and Information Technology 2010, p. 53 (58); *Hon/Millard/Walden* (*supra* Note 5), p. 14.

22 *Esayas* (supra Note 7), p. 6.

23 Cf. *Spindler*, Verhandlungen des 69. Deutschen Juristentages, Band I, Gutachten, 2012, pp. 115 et seq.

24 England and Wales High Court (Administrative Court), [2011] EWHC 1430 (Admin), Case No. CO/12544/2009, Recital 51 f.; Upper Tribunal (Administrative Appeals Chamber), [2011] UKUT 153 (AAC), Appeal Number: GI/150/2011, GI/151/2011, gi/152/2011, Recital 128; House of Lords, [2008] UKHL 47, recital 27; The Paris Appeal Court, decision of 15 May 2007 – *Henri S. vs. SCPP*; Local Court of Munich, decision of 30 September 2008 – 133 C 5677/08, Recital 26; District Court of Wuppertal, decision of 19 October 2010 – 25 Qs 10 Js 1977/08-177/10; District Court of Berlin, decision of 31 January 2013 – 57 S 87/08; different point of view: The Stockholm Lænsrætt, reference No. 593-2005, publication date 8 June 2005; Local Court of Berlin-Mitte, decision of 27 March 2007 – 5 C 314/06, Recital 20; Administrative Court of Wiesbaden, decision of 27 February 2009 – 6 K 1045/08.WI, Recitals 52 et seq.

25 Cf. *Vedsted-Hansen*, in: Peers/Hervey/Kenner/Ward (eds.) (*supra* Note 17), Art 7, Recital 07.72A.

26 Cf. *Zuiderveen Borgesius*, Computer Law & Security Review 2016, p. 256 (267) who interprets Recital 26 as "an absolute approach to identifiability"; *Polonetsky/Tene/Finch*, Santa Clara Law Review, (Forthcoming) 2016, p. 593 (614).

27 Cf. *Kranenborg*, in: Peers/Hervey/Kenner/Ward (eds.) (*supra* Note 17), Art 8, Recital 08.85.

28 *Brink/Eckhardt*, Zeitschrift für Datenschutz 2015, p. 205 (208); *Buchner*, Datenschutz und Datensicherheit 2016, p. 155 et seq.; *Härting* (*supra* Note 18), Recital 279.

29 *Hon/Kosta/Millard/Stefanatou,* Tilburg Law School Legal Studies Research Paper Series No. 07/2014, p. 9, available at: <http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2405971>, accessed 15 August 2016; *Zuiderveen Borgesius* (*supra* Note 26), p. 256 (267); *Marnau*, Datenschutz und Datensicherheit 2016, p. 428 (430).

30 Cf. *Esayas* (*supra* Note 7), p. 6; *Härting* (*supra* Note 18), Recital 282.

31 *Esayas* (*supra* Note 7), p. 6; regarding "anonymisation of personal data in the GDPR" see *infra* B.II.2.d.).

32 *Article 29 Data Protection Working Party*, WP 136 (*supra* Note 5), p. 15; *Article 29 Data Protection Working Party*, WP 216 (*supra* Note 5), pp. 8 et seq.

scope.[33] Significant objective factors will be *inter alia* the state of science and technology, including future technological developments as well as the time and costs needed to identify somebody.[34]

21 In October 2014, the German Federal Court of Justice (BGH) requested the European Court of Justice (ECJ)[35] for a preliminary ruling on the interpretation of the dispute regarding whether a dynamic IP address can be considered as personal data,[36] in particular if the relevant additional information is held by a third party, such as an internet service provider. The ECJ will most likely resolve the dispute between an *absolute* or *relative* approach regarding dynamic IP-addresses by interpreting Article 2 (a) DPD and especially recital 26 of the DPD.[37] Since Article 2 (a) DPD and Article 4 No. 1 GDPR are very similar, the ECJ's decision will certainly also have a major influence on the general interpretation of defining "identifiability" in the GDPR.[38] On 12 May 2016 the *Advocate General (AG)*, *Campos Sánchez-Bordona* published his opinion regarding this case, however, whilst the ECJ is not bound to follow his opinion, it often does so.[39]

22 In his opinion, the *AG* contradicts an interpretation of "means likely reasonably to be used ... by any other person" in such a way that it would be sufficient that *any* third party might obtain additional data in order to identify a person[40], since this "overly strict interpretation would lead, in practice, to the classification as personal data of all kinds of information, no matter how insufficient it is in itself to facilitate the identification of a user".[41] Moreover, the *AG* emphasises that otherwise "it would never be possible to rule out, with absolute certainty, the possibility that there is no third party in possession of additional data which may be combined with that information".[42] This can be interpreted as a tendency of the *AG* towards a *relative* approach. Furthermore, according to the *AG*, "(j)ust as recital 26 refers not to any means which may be used by the controller (...), but only to those that it is likely 'reasonably' to use, the legislature must also be understood as referring to 'third parties' who, also in *a reasonable manner*, may be approached by a controller seeking to obtain additional data for the purpose of identification".[43] The *AG* concludes that contacting third parties shall not be reasonable when it is "very costly in human and economic terms, or practically impossible or prohibited by law".[44] Otherwise, distinguishing between the different means would be nearly impossible, since it would always be possible to imagine the hypothetical contingency of a third party who could — now or in the future — have additional relevant data to assist in the identification of a data subject.[45]

23 Although the *AG* states that in the future advances in technical means will "significantly facilitate access to increasingly sophisticated instruments for collecting and processing data" and thus, the safeguards put in place in defence of privacy are justified, this shall not result in a failure to take account of "the means likely reasonably to be used" by certain third parties.[46] Consequently, the *AG's* opinion includes several *relative* elements which clearly advocate against an *absolute* approach that would lead to an indefinite scope of the GDPR.

24 Nevertheless, according to the *AG*, it would be sufficient to obtain information "reasonably" if the legal *possibility* of retaining and transferring it to others exists. The possibility that the data *may* be transferred shall itself transform the dynamic IP address into personal data for the provider of services on the Internet.[47] The reasonable means of access shall be *lawful means*, therefore, "the legally relevant means of access are reduced significantly,

33 *Spindler*, Der Betrieb 2016, pp. 937 et seq.

34 *Härting* (*supra* Note 18), Recital 284; *Zuiderveen Borgesius* (*supra* Note 26), p. 256 (262).

35 ECJ, Case C-582/14 – *Patrick Breyer v Bundesrepublik Deutschland*.

36 German Federal Court of Justice (BGH), decision of 28 October 2014 – VI ZR 135/13.

37 German Federal Court of Justice, (*supra* Note 36), Recitals 27, 29 et seq.

38 *Härting*, Der IT-Rechts-Berater 2016, pp. 36 et seq.; *Keppeler*, Computer und Recht 2016, p. 360 (364).

39 Opinion of *Advocate General Campos Sánchez-Bordona*, delivered on 12 May 2016, Case C-582/14 – *Patrick Breyer v Bundesrepublik Deutschland*.

40 Opinion of the *Advocate General* (*supra* Note 39), Recital 64.

41 Opinion of the *Advocate General* (*supra* Note 39), Recital 65.

42 Opinion of the *Advocate General* (*supra* Note 39), Recital 65.

43 Opinion of the *Advocate General* (*supra* Note 39), Recital 68 (emphasis added).

44 Opinion of the *Advocate General* (*supra* Note 39), Recital 68; see also in favour of an "unreasonableness" of using illegal means *Spindler/Nink* in: Spindler/Schuster (eds.), Recht der elektronischen Medien, 3rd Ed. 2015, § 11 TMG Recital 8; *Brisch/Pieper*, Computer und Recht 2015, p. 724 (728), who argue that the wording of „reason" is not compatible with the use of illegal means, but who are, however, against a strict classification of illegal means as unreasonable and thus recommend a consideration of each individual case.

45 Cf. Opinion of the *Advocate General* (*supra* Note 39), Recital 68.

46 Opinion of the *Advocate General* (*supra* Note 39), Recital 66 et seq.

47 Opinion of the *Advocate General* (*supra* Note 39), Recital 72, who additionally names this *possibility* "perfectly reasonable"; cf. regarding the classification of dynamic IP addresses as personal data for *access providers* judged by the EJC, Case C-70/10, judgement of 24 November 2011 – *Scarlet Extended SA v Sabam*, Recital 51, which states that "[IP] addresses are protected personal data because they allow those users to be precisely identified".

since they must be exclusively lawful"[48], however, according to the *AG* it shall not matter how restrictive they may be in their practical application for constituting "reasonable means".[49] Allowing even the *possibility* of obtaining the data is a significant limitation of the above mentioned *relative* elements of the *AG's* interpretation and widens the material scope of the DPD and, consequently, also that of the GDPR significantly.

25 A further broadening of the scope and an orientation towards an *absolute* interpretation of identifiable can be found in the *GA's* statement that alone the sheer *potential* possibility of identification shall be sufficient and not that the dynamic IP address *only* becomes personal data when the Internet service provider receives it.[50] Hence, the *AG's* opinion can be interpreted as a vote for a rather *absolute* approach, which would lead to an even wider scope of the GDPR.

26 However, extending the scope of the Regulation too widely could lead to burdening regulations for data-processing entities which would be incommensurate with the actual risks to the privacy of the data subjects[51] and would thus not be compatible with the purpose of data protection law.[52] Because if the ECJ followed this broad – and nearly absolute – approach of the *AG*, virtually all data would have to be considered as personal data, which would, in the end, weaken the data protection framework and could make it unworkable[53], for instance because of an increase of informed consents and legal permissions to process the data.[54] If all data should be treated as personally identifiable and subjected to the GDPR, this could result in creating "perverse incentives" for controllers to abandon anonymisation and therefore increase, rather than relieve, privacy risks.[55] Thus, the very opposite of the protective intention would occur. Hence, we still hope that the ECJ will not follow the lines of argumentation of the *AG*.

2. Non Personal Data

27 Data which does not have any personal references, for instance sheer machine data or so called *attribute data*, does not fall under the material scope of the GDPR. Sensors that collect data for applications, e.g. made for climate analysis or the monitoring of industrial complexes do not process personal data at any stage.[56]

28 However, this attribute data can still turn into personal data when related to a natural person, for instance in the case of a worker's shift or when being linked with other information in a *Big Data* scenario.[57] Data from the *Internet of Things*[58], e.g. from cars, machines ("Industry 4.0"), smart homes or household applications will in many cases be connected to natural persons and thus be considered as personal data.[59] Moreover, the huge amounts of data can be used in connection with technologies like radio frequency identification tags ("RFID-tags") or monitoring and personal profiling so that identification might be easier than before.[60] How easy a re-identification is was demonstrated by a study carried out by computer science professor *Latanya Sweeney* which showed that the combination of a postal code, date of birth, and gender, is sufficient to identify 87% of individuals in the U.S[61], despite the fact that such data that are usually considered to be non personal data[62].

48 Cf. *supra* Note 44.

49 Opinion of the *Advocate General* (*supra* Note 39), Recital 73.

50 Opinion of the *Advocate General* (*supra* Note 39), Recital 77: "(...) their potential as a means of identifying — by themselves or together with other data — a natural person"; cf. *Keppeler* (*supra* Note 38), p. 360 (362).

51 Cf. *Schwartz/Solove*, California Law Review 2014, p. 877 (887).

52 Cf. Recital 4 S. 2 GDPR: "The right to the protection of personal data is not an absolute right; it must be considered in relation to its function in society and be balanced against other fundamental rights, in accordance with the principle of proportionality".

53 Cf. *Tene/Polonetsky*, Stanford Law Review 2012, p. 63 (66).

54 *Keppeler* (*supra* Note 38), p. 360 (364), who points out the practical problem that an increase of informed consents could mean that the text of the consents will be read even less by the data subjects and that a consent can be withdrawn by the data subject at any time, Article 7 Par. 3 GDPR.

55 Cf. *Tene/Polonetsky* (*supra* Note 53), p. 63 (66).

56 See *Rouvroy*, Council of Europe, T-PD-BUR(2015)09REV, Of Data and Men: Fundamental Rights and Freedoms in a World of Big Data, p. 20, available at: <https://www.coe.int/t/dghl/standardsetting/dataprotection/TPD_documents/T-PD-BUR(2015)09REV_Big%20Data%20report_A%20%20Rouvroy_Final_EN.pdf>, accessed 28 July 2016.

57 Cf. *Karg* (*supra* Note 17), p. 520 (522).

58 See for more use cases of the *Internet of Things*: *Vermesan/Friess* (eds.), Digitising the Industry - Internet of Things Connecting the Physical, Digital and Virtual Worlds, pp. 15 et seq., available at: <http://www.internet-of-things-research.eu/pdf/Digitising_the_Industry_IoT_IERC_2016_Cluster_eBook_978-87-93379-82-4_P_Web.pdf>, accessed 8 August 2016.

59 *Härting* (*supra* Note 18), Recital 268.

60 See regarding RFID and data protection law *TAUCIS*, Technikfolgenabschätzung: Ubiquitäres Computing und Informationelle Selbstbestimmung, 2006, pp. 198 et seq., available at: <https://www.datenschutzzentrum.de/taucis/ita_taucis.pdf>, accessed 29 July 2016; *Schmid*, Radio Frequency Identification Law Beyond 2007, in: Floerkemeier et al., The Internet of Things, 2008, pp. 196 et seq.

61 *Sweeney*, Carnegie Mellon University, School of Computer Science, Data Privacy Lab, Working Paper No. 3, 2000, available at: <http://dataprivacylab.org/projects/identifiability/paper1.pdf>, accessed 15 August 2016.

62 *Schwartz/Solove*, N.Y.U. L.Q. Rev. 2011, p. 1814 (1842),

29 Recital 30 of the GDPR now explicitly states that:

> *"(n)atural persons may be associated with online identifiers provided by their devices, applications, tools and protocols, such as internet protocol addresses, cookie identifiers or other identifiers such as radio frequency identification tags. This may leave traces which, in particular when combined with unique identifiers and other information received by the servers, may be used to create profiles of the natural persons and identify them."*

30 Thus, a lot of the data which originally was *attribute data*, e.g. produced by *Internet of Things* technologies, will become personal data due to the association of online identifiers with natural persons. Data of machines connected to the internet and operated by factory workers, of customers being tracked by RFID-tags, smart grid data, or of devices in smart homes or connected household appliances (e.g. toothbrushes, fridges, watches or TVs) will therefore be considered as personal data.[63] Additionally, natural persons can often be identified or be identifiable by "singling out"[64] their data. Thus, because of the broad material scope of the GDPR and of *Big Data* technologies, there are fewer and fewer possibilities to process data without a personal reference, in particular in the *Internet of Things* era.

3. Conclusion

31 The GDPR's material scope contains several parts which can be interpreted as *relative* approaches regarding the identifiability of natural persons, most prominently with the duty to include means only, if they are "reasonably likely to be used". Moreover, according to the *AG*, illegal means shall not be considered. Nevertheless, several other terms indicate a rather *absolute* approach of the GDPR, be it the wide scope of the online identifiers, the incorporation of "singling out" or that information obtained by a *third* person shall be sufficient to make the data personal for a controller. If the *AG's* opinion that the mere *possibility* of retaining and transferring the data to others is sufficient for a personal reference of data will prevail, the GDPR's material scope will have to be interpreted widely, using a mix of *relative* and *absolute* elements – an approach which could turn out to be a *pyrrhic victory*.

II. Encrypted Data and the GDPR

32 Encrypting personal data is a data security technique which has the effect of rendering data unintelligible to any person who is not authorised to access it due to encoding the information into a mutilated state, so that only parties with access to a decoding mechanism and a secret decryption key can access the information.[65] Encryption of data seems to be one of the promising solutions in order to ensure privacy particularly in cloud computing environments. When a controller encrypts the data before uploading it to a cloud, the data is regarded as personal data for the controller who holds the decryption key and the controller thus remains accountable for the data.[66] As encrypted personal data makes sure that no unauthorized person is able to use the sensitive data, only the original data controller is able to identify the persons related to data stored in the cloud – and not the cloud operator nor third persons. Hence, encryption may serve as a tool to safeguard data protection. Furthermore, when processing is carried out on behalf of the controller, such as in a cloud computing scenario, the GDPR introduces several new obligations to comply with - especially for processors and not only for controllers. Encrypting personal data can thus be a useful way to avoid these obligations for the processor.

33 In the cases where third parties are able to decrypt the data but the controller cannot, the question whether the GDPR shall be applicable for this controller is a point of controversy.[67] Moreover, if decryption has been achieved only by the use of illegal means, the controller who has not used those means shall not be subjected to the GDPR.[68]

34 In this section, we examine the provisions of the GDPR regarding encryption, anonymous and pseudonymous data in order to be able to assess the effect of encrypted personal data on the material scope of the Regulation.

available at: <http://scholarship.law.berkeley.edu/facpubs/1638>, accessed 10 August 2016; *Ohm*, UCLA Law Review 2010, p. 1701 (1705) with further examples.

63 See *International Working Group on Data Protection in Telecommunications*, Working Paper on Big Data and Privacy – Privacy principles under pressure in the age of Big Data analytics, 55th Meeting, 2014, Skopje, p. 4, available at: <http://dzlp.mk/sites/default/files/u972/WP_Big_Data_final_clean_675.48.12%20%281%29.pdf>, accessed 28 July 2016.

64 Regarding singling out people without knowing their names (for behavioural targeting) see *Zuiderveen Borgesius* (*supra* Note 26) pp. 256 et seq. and *supra* B.I.1.c.).

65 Cf. *ENISA*, Privacy by design in big data – An overview of privacy enhancing technologies in the era of big data analytics, 2015, p. 38; available at: <https://www.enisa.europa.eu/publications/big-data-protection>, accessed 9 August 2016; *Gürses/Kundnani/van Hoboken*, Media, Culture & Society, Crypto and empire: the contradictions of counter-surveillance advocacy, 2016, p. 7.

66 *Hon/Kosta/Millard/Stefanatou* (*supra* Note 29), p. 10.

67 See *infra* B.II.2.d.).

68 Cf. the Opinion of the *Advocate General*, *supra* Note 44.

1. Encryption in the GDPR

35 Unlike the proposal of the Parliament[69], the final version of the GDPR does not provide a further definition of encrypted data, but mentions encryption in several provisions as a compliance requirement. According to Article 32 Par. 1 (a) GDPR, encryption is regarded as an appropriate technical and organisational measure to ensure the security of processing. It is apparent that this does not deal with the applicability of the GDPR, but rather with the protection of personal data.[70]

36 Moreover, in case of a data breach, the controller is not required to communicate to the data subject if he or she has implemented encryption as a technical and organisational protection measure (Article 34 Par. 3 (a) GDPR).

37 Additionally, it is one of the "appropriate safeguards" of Article 6 Par. 4 (e) GDPR, which have to be taken into account when assessing the compatibility of a processing for a purpose other than that for which the personal data have been collected. Finally, depending on the classification of encryption as pseudonymisation or not[71], the provisions of the GDPR regarding pseudonymous data[72] may be applicable for encrypted data, too.

2. Is Encrypted Data Anonymised or Pseudonymised?

38 Since the GDPR does not define "encrypted data", we have to examine if encryption is a technique which anonymises or just pseudonymises personal data. In this regard, again the dispute regarding the material scope of the Regulation, as described above, plays an important role. To assess whether encrypted data has to be treated as anonymised or pseudonymised data, we first have to provide an overview of the GDPR's provisions regarding these privacy preserving techniques.

a.) "Anonymous Information" in the GDPR

39 Although technologies to anonymise personal data are considered to be of high value to protect the fundamental privacy rights of the data subjects, the GDPR does not provide a specific article to regulate "anonymous information" in the Regulation, it is only mentioned in one Recital. According to Recital 26 S. 4 and 5 GDPR the:

> *"principles of data protection should (...) 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."*

40 Thus, the GDPR is not applicable to anonymous data. To examine whether data can be considered as anonymous; once again the problem of the identifiability of data subjects arises.[73] In this regard, the possibility to anonymise personal data in the GDPR can be seen as another hint in favour of a *relative* approach, because given the possibilities to re-identify and combine data (*Big Data*), anonymous information could not be established when following a pure *absolute* approach.[74] However, to determine whether encrypted data may be considered as anonymous data, we will first take a look at the GDPR's provisions regarding pseudonymisation.

b.) "Pseudonymisation" and "Pseudonymous Data" in the GDPR

41 Unlike in the DPD, the GDPR includes a definition of "pseudonymisation". According to Article 4 No. 5 GDPR, pseudonymisation:

> *"means the processing of personal data in such a manner that the personal data can no longer be attributed to a specific data subject without the use of additional information, provided that such additional information is kept separately and is subject to technical and organisational measures to ensure that the personal data are not attributed to an identified or identifiable natural person".*

42 Moreover, "(t)he application of pseudonymisation to personal data can reduce the risks to the data subjects concerned and help controllers and processors to meet their data-protection obligations" (Recital 28 S. 1 GDPR). Furthermore, Recital 28 S. 2 GDPR emphasises that the explicit introduction of "pseudonymisation" does not intend to preclude any other measures of data protection. Thus, the connection between a

69 Article 4 No. 2b of the proposal of the European Parliament for a GDPR (LIBE proposal) defines encrypted data as "personal data, which through technological protection measures is rendered unintelligible to any person who is not authorised to access it", thus, according to LIBE, encrypted data shall just be a subcategory of personal data, which shall not lose its personal reference due to encryption.

70 See Recital 83 GDPR for more details regarding these measures.

71 See *infra* B.II.2.c.).

72 See *infra* B.II.2.b.).

73 See *supra* B.I.1.

74 *Härting* (*supra* Note 18), Recital 291.

natural person and the information on the basis of a corresponding rule remains – pseudonymised data is still qualified as personal data.[75] Hence, pseudonymisation is merely a method which can reduce the likelihood of identifiability of individuals, but does not exclude this data from the material scope of the GDPR. It is handled by the Regulation primarily as a data security measure,[76] and its use is encouraged in several articles of the GDPR; Article 32 Par. 1 (a) names it an appropriate technical and organisational measures to ensure a level of security appropriate to the risk.

43 Moreover, pseudonymisation shall, like encryption, be one of the "appropriate safeguards" of Article 6 Par. 4 (e) GDPR.[77] In addition, in accordance with Article 89 Par. 1 S. 3 GDPR, pseudonymisation is a safeguard to ensure that technical and organisational measures are applied when personal data is being (further) processed for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes. Finally, pseudonymisation is a technical and organisational measure that shall be implemented by the controller as a way to comply with the principle of data minimisation for the newly introduced provisions for "data protection by design and by default".[78] However, the GDPR does not distinguish between the quality of the possible pseudonymisation measures and its consequences for the controller. Nevertheless, to clearly define the unclear provision and the use of pseudonymisation, associations and other bodies representing categories of controllers or processors may prepare "codes of conduct" according to Article 40 Par. 2 (d).[79]

c.) Encrypted Data as Pseudonymised Data or Anonymous Data?

44 When encrypting personal data, in accordance with Article 4 No. 5 GDPR, the encryption key is the "additional information" which is "kept separately" and "subject to technical and organisational measures". Hence safety measures such as a secure key management and the respective encryption method used by the controller have to be used "to ensure that the personal data are not attributed to an identified or identifiable natural person". Therefore, because of its existing assignment rule encryption is an example of pseudonymisation.[80]

45 However, it is controversial whether encrypted personal data, and thus pseudonymised data, can be regarded as anonymised[81] data. Encrypted personal data should nevertheless undisputedly remain personal data to a person who holds the decryption key.[82] The relevant question is whether encrypted data shall also be personal data for a controller or processor who does not have access to the decryption key, for instance a cloud provider. Some academics have argued in this direction[83]; far more important, the *Art. 29 Data Protection Working Party* opines that believing that a pseudonymised dataset is anonymised is a "common mistake".[84] Additionally, the wording of Recital 26 S. 2 GDPR states that "personal data which have undergone pseudonymisation, which could be attributed to a natural person by the use of additional information should be considered to be information on an identifiable natural person".

46 At first sight, this is a clear statement of the EU legislator that pseudonymised data shall always be personal data. Nevertheless, to resolve this dispute, once again the question is crucial whether an *absolute* or a *relative* approach regarding the identifiability of a data subject has to be applied. According to the *absolute* approach, encrypted data will consequently always be personal data, because somebody, at least the key holder or any other party given sufficient time, economic resources and computing power, will always be able to decrypt the data, since no system of encryption can be completely secure[85]. According to this logic, encryption is merely a technical and organisational measure to ensure that data is not accessible to unauthorised persons rather than changing the data's quality. However, with a *relative* approach the data could be regarded as anonymous for the controller.

75 *Karg* (*supra* Note 17), p. 520 (522); see *infra* B.II.2.c.).

76 *Zuiderveen Borgesius* (*supra* Note 26), p. 256 (267).

77 See *supra* B.II.1.

78 According to Recital 78 GDPR, personal data should be pseudonymised "as soon as possible".

79 *Marnau* (*supra* Note 29), p. 428 (431).

80 *Article 29 Data Protection Working Party*, WP 216 (*supra* Note 5), p. 21; *Esayas* (*supra* Note 7), p. 8; *Hennrichs*, Cloud Computing - Herausforderungen an den Rechtsrahmen für Datenschutz, 2016, p. 137.

81 For an overview of existing anonymization techniques such as randomization or generalization see the Opinion of the *Article 29 Data Protection Working Party*, WP 216 (*supra* Note 5), pp. 12 et seq.; *International Working Group on Data Protection in Telecommunications* (*supra* Note 63), pp. 13 et seq., which provides guidelines for procedures for robust anonymisation; *ENISA* 2015 (*supra* Note 65), pp. 27 et seq.; *Lagos*, Indiana Law Review 2014-2015, pp. 187 et seq.

82 *Article 29 Data Protection Working Party*, WP 216 (*supra* Note 5), p. 29; *Borges*, in: Borges/Meents (eds.), Cloud Computing, 2016, § 6 Recital 33; *Polonetsky/Tene/Finch* (*supra* Note 26) p. 593 (613).

83 *Wagner/Blaufuß*, Betriebs-Berater 2012, p. 1751; *Esayas* (*supra* Note 7), p. 8.

84 *Article 29 Data Protection Working Party*, WP 216 (*supra* Note 5), p. 21.

85 Cf. *Kuner*, International Business Lawyer 1996, p. 186.

d.) Requirements for Encrypted Data in order to be considered as Anonymous Data

47 Consequently, we have to examine which level of encryption is sufficient so that with a *relative* approach the encrypted personal data can be considered as anonymous data. As mentioned above, only the knowledge and possibilities of the controller to identify the data subject shall be taken into account, therefore, processing encrypted data without affecting the scope of the data protection law might be possible.[86]

48 In order to concretise whether the means to decrypt the dataset and identify the data subject are reasonably likely to be used, one should take account of objective factors. There are three relevant factors that have to be considered when assessing the level of security of encrypted data against decryption, namely the strength of the encryption algorithm used, the length of the encryption key (the longer the key the safer the encryption will be) and the security of the key management.[87] Obviously, the key always has to be stored separately from the encrypted data in a secure way. If not, attackers may easily be able to decrypt the data[88] and thus, the personal data would no longer be anonymous. The simplest and most common way of decryption is using *exhaustive key search* or *brute-force attacks* which means to try all possible keys and eventually guessing correctly.[89] However, if a secure encryption technology is used, this way of decrypting the dataset cannot be considered as very likely for the controller.[90]

49 Other approaches to get access to the secret key are e.g. legally getting access to a key via a court decision, extracting the key from software or hardware, or by using accidental errors or systematic *backdoors* implemented in the encryption technique for law enforcement.[91] These ways are only considered to be likely for the controller if they do not violate the law or if they can be achieved by the use of computational power which can be reasonably expected. However, if a *backdoor* is implemented by the government into an encryption technology, the GDPR would be applicable for the controller who knows about this (governmental) possibility of accessing the personal data.

50 Additionally, as outlined above[92], the available encryption technology at the time of the processing has to be considered: applying the *AG's* opinion on encryption it would not be reasonably likely if it were practically impossible to decrypt the dataset, thus, if a state of the art encryption technology is enabled, in most of the cases, decrypting will be virtually impossible and therefore not likely and only possible with unreasonable efforts.[93] However, if according to the *AG* even the *potential* possibility of obtaining the decryption key from another person in a lawful way would be sufficient for an identification, the possibilities to avoid the applicability of the GDPR due to anonymisation via encryption would be very restricted.

51 Arguments against this wide interpretation could be sustained by Recital 57 GDPR, which deals with the data subject's right to access personal data held by the controller, where "the personal data processed by a controller do not permit the controller to identify a natural person". Then, "the data controller should not be obliged to acquire additional information in order to identify the data subject for the sole purpose of complying with any provision of this Regulation". This could be a hint against a too wide interpretation of getting access to a key obtained by a third party.

52 Additionally, future technological developments of decryption, e.g. due to more computing power or improved algorithms have to be considered (cf. Recital 26 GDPR), especially regarding the lengths of the secret key. The controller has to assess whether the future development is evidently foreseeable and thus ought to be regarded as a present information.[94] According to the *Article 29 Data Protection Working Party* (regarding the DPD), the controller should

86 Cf. *Hon/Kosta/Millard/Stefanatou* (*supra* Note 29), p. 10; *Borges,* in: Borges/Meents (eds.) (*supra* Note 82), § 6 Recital 33; *Hennrichs* (*supra* Note 80), 2016, p. 137.

87 *Hon/Millard/Walden* (*supra* Note 5), p. 22.

88 Cf. *Article 29 Data Protection Working Party*, WP 216 (*supra* Note 5), p. 22; *Hon/Kosta/Millard/Stefanatou* (*supra* Note 29), p. 10.

89 See with further examples *Gürses/Preenel*, in: van der Sloot/Broeders/Schrijvers (eds.), Exploring the Boundaries of Big Data, 2016, Part I, 3, Cryptology and Privacy in the Context of Big Data, p. 49 (62) available at: <http://www.ivir.nl/publicaties/download/1764.pdf>; accessed 29 August 2016; *Kroschwald*, Zeitschrift für Datenschutzrecht 2014, p. 75 (77).

90 *Cahsor/Sorge*, in: Borges/Meents (eds.) (*supra* Note 82), § 10 Recital 32, who state that using the 128 bits key lengths of AES encryption would make such an attack nearly impossible and thus not likely.

91 *Gürses/Preenel*, in: van der Sloot/Broeders/Schrijvers (eds.) (*supra* Note 89), p. 49 (63); the German and French government are currently deliberating on legal obligations to implement backdoors in encryption techniques for law enforcement reasons, see <http://www.interieur.gouv.fr/Actualites/L-actu-du-Ministere/Initiative-franco-allemande-sur-la-securite-interieure-en-Europe>, accessed 27 August 2016.

92 *Supra* B.I.1.c.).

93 Different opinion: A*rticle 29 Data Protection Working Party*, WP 216 (*supra* Note 5), p. 10, according to which the intentions of the data controller or recipient shall not matter, as long as the data are identifiable, data protection rules shall apply; see regarding the effect of different encryption technologies upon the applicability of the GDPR *infra* B.II.3.

94 *Borges*, in: Borges/Meents (eds.) (*supra* Note 82), § 6 Recital 38.

take into account the technological development for the period of time in which the data is meant to be processed, therefore, if the data shall be processed for ten years, he or she has to take the technological possibilities for these ten years into account; if the data can be decrypted in the ninth year, the data shall become personal data from that date on only.[95]

53 Therefore, due to technical developments, encrypted data will only be anonymous for a certain period of time and thus, the level of encryption has to be checked constantly by the controller and not only when the controller processes the data for the first time.[96] Moreover, if a controller receives an already encrypted dataset, he or she has to obtain further information regarding whether the original dataset has included personal data; if yes, the controller has to regularly check the state-of-the-art of the encryption technique.[97]

54 Thus, if the controller does not have the key to decrypt the data or other means to make it legible, it is in most cases reasonably likely that he or she cannot access the personal information, which consequently has to be regarded as anonymous information. Therefore, according to the GDPR, when using state-of-the-art encryption technique, encrypted personal data can be anonymous data, with the limitation that a potential possibility of obtaining the key, also by a third party and especially due to decryption, always has to be considered, but only if the means used are *reasonably likely*.

e.) Anonymisation as (Further) Processing of Personal Data

55 There is legal uncertainty regarding the lawfulness of the anonymisation process, more precisely whether anonymising personal data means "further processing" of personal data.[98] The *Working Party* states in its WP 216 (still regarding the DPD), that "anonymisation constitutes a further processing of personal data; as such, it must satisfy the requirement of compatibility by having regard to the legal grounds and circumstances of the further processing".[99]

56 Nevertheless, this further processing of personal data is considered to be compatible with the original purposes of the processing but only if the anonymisation process leads to "reliable (...) anonymised information".[100] Furthermore, the data controller's legitimate interest always has to be balanced against the data subject's rights and fundamental freedoms.[101] Consequently, according to the *Working Party*, anonymisation can be compatible with the original purposes of the processing, but it would be a violation of data protection law if personal data was anonymised for purposes that are not compatible with the original purpose and if there were no other legitimate grounds for processing the data, such as the data subject's consent.[102]

57 The *Working Party* clarifies this by providing as an example the anonymisation of the contents of traffic data immediately after its collection by mobile operators which performed deep packet inspection technologies. It was lawful in accordance with Article 7 (f) DPD, because of a legal permission stipulated in Article 6 Par. 1 of the e-Privacy Directive for certain traffic data which has to be erased or made anonymous as soon as possible when it is processed and stored by the provider of a public communications network or publicly available electronic communications service.[103]

58 Another example is given by *Esayas*, according to which the anonymisation of personal data for the purpose of using this data for advertising would constitute a violation of data protection law (unless there are other legitimate grounds for the processing – for instance the data subject's consent), if the data has originally been collected to provide a certain service for the data subject.[104]

59 Applying the *Working Party's* interpretation to the GDPR, the Regulation's requirements regarding further processing need to be fulfilled when anonymising personal data. Thus, it has to be analysed whether anonymisation is a *compatible* use according to the GDPR, then no legal basis separate from that which allowed the collection of

95 *Article 29 Data Protection Working Party*, WP 136 (*supra* Note 5), p. 18.

96 *Spindler*, (*supra* Note 23), p. 115; *Borges*, in: Borges/Meents (eds.) (*supra* Note 82), § 6 Recital 40; different opinion *Lundevall-Unger/Tranvik* (*supra* Note 21), p. 53 (71) who call it "a burden [for the controllers] that they probably cannot be expected to bear" and state that it "will not make controllers in a wired world more inclined to comply with the provisions of the [European data protection law]".

97 *Borges*, in: Borges/Meents (eds.) (*supra* Note 82), § 6 Recital 41.

98 See *El Emam/Álvarez*, International Data Privacy Law 2015, p. 73 (79); *Hon/Kosta/Millard/Stefanatou* (*supra* Note 29), p. 12; *Esayas* (*supra* Note 7), pp. 4 et seq.

99 *Article 29 Data Protection Working Party*, WP 216 (*supra* Note 5), pp. 3, 7.

100 *Article 29 Data Protection Working Party*, WP 216 (*supra* Note 5), p. 7.

101 Cf. *Article 29 Data Protection Working Party*, WP 216 (*supra* Note 5), p. 8.

102 Cf. *Walden*, International Journal of Law and Information Technology 2002, p. 224 (233); *Esayas* (*supra* Note 7), p. 4.

103 *Article 29 Data Protection Working Party*, WP 216 (*supra* Note 5), p. 8.

104 *Esayas* (*supra* Note 7), p. 4.

the personal data would be required (Cf. Recital 50 GDPR). Recital 49 GDPR states that:

> *"(t)he processing of personal data to the extent strictly necessary and proportionate for the purposes of ensuring (...) information security (...) constitutes a legitimate interest of the data controller concerned. This could, for example, include preventing unauthorised access to electronic communications networks and malicious code distribution (...)."*

60 According to this, the anonymisation of personal data could be interpreted as necessary for ensuring information security and be, in accordance with Article 6 Par. 1 (f) GDPR, of legitimate interest to a controller.[105] Apart from this, according to Article 5 Par. 1 (b) GDPR a "further processing for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes shall, in accordance with Article 89 (1), not be considered to be incompatible with the initial purposes".

61 Furthermore, according to the compatibility test of Article 6 Par. 4 GDPR, account should be taken *inter alia* of the possible consequences of the intended further processing for data subjects. Since anonymisation, pseudonymisation and encryption are privacy preserving technologies[106], in most cases applying these tools on the data subject's personal data will be in their interest.

62 However, regarding a possible re-identification of the personal data, the consequences of anonymising personal data for a data subject could also be serious (e.g. when processing special categories of personal data according to Article 9 GDPR) and thus not in its interest if the anonymous data, which is not affected by the scope of the GDPR, is transferred unrestricted from controller to controller.

63 Even though these concerns have to be taken seriously, the *Working Party's* opinion implies a non-existent weakness of the data protection law. Because as long as the data is anonymous, there is no threat to the privacy of the data subjects and as soon as a re-identification of the data is possible the GDPR with all its protective instruments is applicable again. Moreover, the need to justify the process of anonymisation itself could discourage the use of anonymisation and pseudonymisation as privacy-enhancing techniques.[107] However, with the use of Recital 49 GDPR, this dispute could possibly come to an end as soon as the GDPR comes into effect.

3. The Impact of different Encryption Techniques upon the GDPR's Material Scope

64 Finally, we will give a short overview of significant encryption technologies and examine the effect of these technologies on the applicability of the GDPR by determining *inter alia* which technical level of encryption has to be achieved to avoid a decryption or de-anonymisation of personal data and thus the applicability of the Regulation.

65 We have to distinguish between encrypted *transport* of data (e.g. encryption of e-mails or messages of messenger services via *end-to-end encryption*[108]) and encrypted *storing* of data (e.g. online backups in a cloud). However, if personal data is encrypted whilst being stored, applications and programs may not be able to handle and further process that encrypted data unless the data is decrypted and thus once again personal data. Processing stored encrypted data (e.g. in the cloud) in a secure and useful way – hence without the need of spending too much time or computer power – might be possible by using *Fully Homomorphic Encryption*[109] or *Secure Multiparty Computation*[110].

66 However, first of all, a distinction is made between *symmetric cryptography* and *asymmetric cryptography* techniques.

a.) Symmetric Cryptography – Secret Key Encryption

67 In a *symmetric cryptography* scenario, the parties use a publicly known encryption algorithm to transform the personal data into ciphertext or to later decrypt the dataset, the encryption is performed by a secret key which both parties have access to.[111] The level of security of the encrypted data depends significantly upon the secure storing, management, and transportation of the key which often cannot be transmitted safely.[112] Thus, a safe key management is

105 Cf. *Esayas* (*supra* Note 7), p. 5; *Hon/Kosta/Millard/Stefanatou* (*supra* Note 29), p. 12, who criticise that this legitimate interest should also refer to processors.

106 Cf. Recital 29 S. 1 GDPR which gives incentives for controllers to apply pseudonymisation when processing personal data; Article 5 Par. 1 (c) which regulates the principle of data minimisation, which is fulfilled by these technologies that reduce the amount of personal data.

107 *Hon/Kosta/Millard/Stefanatou* (*supra* Note 29), p. 12; *Esayas* (*supra* Note 7), p. 5.

108 See for details regarding WhatsApp's end-to-end encryption <https://www.whatsapp.com/security/?l=en>, accessed 26 August 2016.

109 See *infra* B.II.3.c.).

110 See *infra* B.II.3.d.).

111 Cf. *Gürses/Preenel*, in: van der Sloot/Broeders/Schrijvers (eds.) (*supra* Note 89), p. 49 (53).

112 *Maisch*, Informationelle Selbstbestimmung in Netzwerken, 2015, p. 322.

a necessary condition for avoiding the applicability of the GDPR, however, this can hardly be achieved when only using *symmetric cryptography*, because any holder of the key can easily re-identify the data subjects through decryption of the dataset.[113]

68 However, a safe transportation can be achieved when encrypting the *symmetric* key with an *asymmetric* encryption technique[114] (*hybrid cryptosystem*). Thus, a decryption in this scenario, when not *asymmetrically* encrypting the key, will in many cases be reasonably likely and the data protection law would thus be applicable for the controller or processor of the *symmetrically* encrypted database.

b.) Asymmetric Cryptography – Public Key Encryption

69 In *asymmetric* or public-key cryptography, two different keys are used, the first key (the public key) is used by the sender to encrypt the information, the second key is a private and secret key used by the recipient to decrypt the information.[115] Therefore, the encryption key can be made public, a common secret is not needed to be agreed on by the parties in advance as the second secret key is only known by the recipient.[116]

70 This technique is used mostly for *end-to-end encryption*. Thus, in an *asymmetric* encryption scenario the private key has to be kept secret. The risk that a third party could obtain the key consequently arises e.g. if the secret key is stored at a cloud provider which also holds the public key or by *man-in-the-middle attacks*, if a third party misleads the other parties by pretending to be the respective counterpart. If all necessary security measures are complied with – in the sense of the *relative* approach – it is not reasonably likely that a *man-in-the-middle attack* occurs.

71 However, in light of the *AG's* wide approach it may be sufficient that there is a *potential* possibility of identification by a third party. Thus, if the secret key is held safely by the recipient, a third party, e.g. a cloud provider which stores or transports the encrypted data does not have access to the private key and will, provided that a state-of-the-art key is used, not be able to decrypt the data (with reasonable efforts) and therefore does not fall under the scope of the GDPR. However, the controller always has to monitor the technological development regarding the key used and possible innovative technological ways of decryption.[117] Since *asymmetric* encryption has a significantly lower performance than *symmetric* encryption, in practice *hybrid* encryption is mostly used.

c.) Fully Homomorphic Encryption

72 *Fully Homomorphic Encryption (FHE)* is an encryption technology that allows the performance of an analysis "in the ciphertext in the same way as in the plaintext without sharing the secret key".[118] Therefore, for the processing of the data, no decryption and thus no knowledge of the private key is needed. Moreover, even the result of the processing is encrypted, which can only be decrypted by the user and not by the cloud provider.[119] The cloud provider will never see the data in plaintext. Thus, when processing personal data with the use of *FHE*, the GDPR is not applicable to the cloud provider which consequently does not process personal data. Unfortunately, due to its still very low performance, *FHE* is at present still highly inefficient and currently not a practical alternative to the processing of personal data on plaintext.[120]

d.) Secure Multiparty Computation

73 *Secure Multiparty Computation (SMC)*[121] allows for secure computation of sensitive data sets, such as tax or health data, without having to trust a centralised entity (such as a trusted third party).[122] It refers to a field of cryptography that deals with protocols involving two or more participants who want to mutually compute a useful result without having to

113 *Article 29 Data Protection Working Party*, WP 216 (*supra* Note 5), p. 20.

114 See *infra* B.II.3.b.).

115 *Gürses/Preenel*, in: van der Sloot/Broeders/Schrijvers (eds.) (*supra* Note 89), p. 49 (53).

116 *Gürses/Preenel*, in: van der Sloot/Broeders/Schrijvers (eds.) (*supra* Note 89), p. 49 (53).

117 See *supra* B.II.2.d.).

118 *ENISA* 2015 (supra Note 65), p. 40; *FHE* was first shown to be possible by *Gentry*, A fully homomorphic encryption scheme, 2009; another type of *homomorphic encryption* is *Somewhat Homomorphic Encryption (SHE)* which has a better performance than *FHE* but limits the number of operations.

119 *Gürses/Preenel*, in: van der Sloot/Broeders/Schrijvers (eds.) (*supra* Note 89), p. 49 (58).

120 *ENISA*, Privacy and Data Protection by Design – from policy to engineering, 2014, p. 43, available at: <https://www.enisa.europa.eu/publications/privacy-and-data-protection-by-design>, accessed 10 August 2016; *Gürses/Preenel*, in: van der Sloot/Broeders/Schrijvers (eds.) (*supra* Note 89), p. 49 (58).

121 *MPC* was first introduced by *Yao*, Proceedings of the 23rd Annual IEEE Symposium on Foundations of Computer Science, 1982, pp. 160 et seq.; for further details about *MPC* see *Cramer/Damgård/Nielsen*, Secure Multiparty Computation and Secret Sharing, 2015.

122 *Gürses/Preenel*, in: van der Sloot/Broeders/Schrijvers (eds.) (*supra* Note 89), p. 49 (60).

trust each other with their sensitive data.[123] Every party will provide an input value and learn only the result of their own individual value so that nobody is able to access all the information.[124] A data donor distributes the data into shares using secret-sharing and sends one random share of each value to a single server.

74 When using *Sharemind*[125], each party will receive one share of every secret value. The original secret can only be reconstructed by collecting all the shares of a value and adding them up.[126] After the data has been transmitted and stored, the server can perform computations on the shared data; however, the server does not share the information with other servers so that none of them can reconstruct the input values.[127] An increase of servers reduces the risk of collusions. After finishing the computation, the results of the servers are transmitted and published to the client of the computation. The servers send the share of the results to the client who reconstructs the real result.[128]

75 Thus, *Sharemind* requires three steps: the donors have to be informed whose data shall be provided; the data has to be divided; and then stored on the different servers. If it is necessary for one data donor to specify whose information the other donor has to provide, this has to be considered as processing of personal data. It would then be inevitable to identify the data subjects whose information is needed for the purposes of computation. Alternatively, to reduce the amount of personal data shared, all data can be loaded to *Sharemind* and securely joined using ciphertexts.

76 As outlined above, before the data is stored on the different servers, it has to be divided into the shares. This process must be carried out in plaintext using personal data. Although Article 4 No. 2 GDPR mentions the "alteration" of data as processing, the division of data does not entangle the application of the GDPR as "alteration" refers to the alteration of content, not of its appearance.[129] The secret-sharing of personal data by dividing it thus does not fall under the GDPR's scope. Once the data has been divided, it will be stored on the different servers. Applying an *absolute* approach on the identifiability of data subjects, these data chunks would have to be considered as personal data and this kind of processing would be processing of personal data – however, with the approach opined in this article, the data chunks are not considered to be personal data since it is highly unlikely for a party to receive the other shares.

77 *SMC* is advantageous due to the fact that simply random fragments of personal data are used. The original data can only be restored (and thus turned into personal data) if all fragments are put together. Hence, it is crucial to determine whether the GDPR is applicable to the computation over data fragments. Without the other parts, the file cannot be read in any way. One fragment itself does not contain information regarding a person and thus cannot be regarded as personal data. Only if all fragments of the data were gathered and put together, the Regulation would be applicable. Theoretically, all server providers may collude and reengineer the personal data. However, this is highly unlikely since the providers of the server themselves have a high interest in ensuring safety and confidentiality of the *SMC* and may be legally bound by contract.[130] This unreasonable chance of collusion leads to ruling out the applicability of the GDPR.

78 In contrast to *SMC*, when using *FHE* the parties do not need to be available online and the result is always encrypted.[131] However, *FHE* and *SMC* are special cases that still are not widespread, thus, when processing encrypted data without using these or similar technologies on some point it will always be necessary to decrypt the information with the consequence that in this moment the GDPR will be applicable to the controller again.[132]

C. Conclusion

79 Encrypting personal data can lead to the non-applicability of the GDPR and might thus be an important privacy preserving technology for controllers – however, since the provisions of the GDPR regarding its material scope also include several elements which can be interpreted in an *absolute* point of view and since the *Advocate General* of the ECJ has widened the scope in his opinion a lot, there is still legal uncertainty regarding the applicability of the GDPR for encrypted data. Therefore, controllers

123 Cf. *Bogdanov*, Sharemind: programmable secure computations with practical applications, 2013, available at: <http://dspace.ut.ee/bitstream/handle/10062/29041/bogdanov_dan_2.pdf?sequence=5&isAllowed=y>, accessed 27 August 2016.

124 *Kamm/Willemson*, International Journal of Information Security, 2015, p. 531 (532).

125 See <https://sharemind.cyber.ee//>.

126 *Bogdanov* (*supra* Note 123), p. 34.

127 *Kamm/Willemson* (*supra* Note 124), p. 531 (532).

128 *Kamm/Willemson* (*supra* Note 124), p. 531 (533).

129 Cf. for the German Federal Data Protection Act *Gola/Klug/Körffer*, in: Gola/Schomerus, Bundesdatenschutzgesetz, 12th Ed. 2015, § 3 Recital 30.

130 Regarding the risks for the confidentiality if parties pool their information see *ENISA* 2015 (*supra* Note 65), p. 41.

131 *Gürses/Preenel*, in: van der Sloot/Broeders/Schrijvers (eds.) (*supra* Note 89), p. 49 (60).

132 *Hoppen*, Computer und Recht 2015, p. 802 (804).

have to analyse each encrypted dataset on its own and determine whether a decryption might be reasonably likely, also taking continuously into account the use of future decryption technologies and the security of the key management. We hope that the ECJ does not follow this nearly *absolute* interpretation of the identifiability of natural persons since it would tremendously harm future incentives of controllers to implement privacy preserving technologies.

80 Additionally, encryption serves as a technical and organisational measure to ensure the security of processing in several parts of the Regulation. Controllers have to consider that the process of encryption as well as anonymisation might constitute a further processing of personal data.

81 Using state-of-the-art *asymmetric* encryption technologies especially for transporting personal data is a method which will in most of the scenarios be unlikely to be decrypted and can according to our interpretation prevent the applicability of the GDPR. Storing encrypted data in a cloud can also be done in a secure way without falling within the material scope of the GDPR. Although existing technologies such as *FHE* and *SMC* can exclude the applicability of the GDPR for the processing of encrypted data, processing encrypted data in most cases still has to be undertaken in plaintext by decrypting the data and thus by the use of personal data.

---

* *Prof. Dr. Gerald Spindler* is holder of the chair of Civil Law, Commercial and Economic Law, Comparative Law, Multimedia- and Telecommunication Law and head of the Institute for Business Law at the University of Göttingen, Germany.

Philipp Schmechel, Dipl.-Jur., is a Ph.D. student and research assistant for the EU-PRACTICE project at Prof. Spindler's chair at the University of Göttingen. His doctoral thesis deals with "Big Data and European data protection law".

The research leading to these results has received funding from the European Union Seventh Framework Programme ([FP7/2007-2013]) under grant agreement number ICT-609611 (PRACTICE). The information in this document is provided "as is", and no guarantee or warranty is given that the information is fit for any particular purpose subject to any liability which is mandatory due to applicable law. The user uses the information at its sole risk and liability.