Disentangling encryption from the personalization debate: On the advisability of endorsing the "relativist approach" underpinning the identifiability criterion.

Pier Giorgio Chiara*

Contents

*Pier Giorgio Chiara is a doctoral researcher in the Law, Science and Technology Joint Doctorate - Rights of Internet of Everything, funded by Marie Skłodowska-Curie Actions, at the University of Luxembourg, Bologna and Turin; piergiorgio.chiara@uni.lu.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie ITN EJD "Law, Science and Technology Rights of Internet of Everything" grant agreement No 814177.

I also would like to thank professor Micheal Veale for the fruitful insights and remarks he gave me at the annual conference IDLaw2020, organized by the University of Wien, where I presented this paper.

University of Vienna Law Review, Vol. 4 No 2 (2020), pp. 168-188, https://doi.org/10.25365/vlr-2020-4-2-168.

1. Introduction

"Encryption is just a bunch of math, and math has no agency"[1]

Since ancient times, society has felt the need to hide sensitive information: cryptography in its various forms has always been a way to guarantee information security. Back in the days, an increasingly data-driven society[2] has not changed the paradigm: European data protection law, i.e. the GDPR, envisages encryption as means to assure the principles underpinning information security[3]. Recently, the European Data Protection Supervisor claimed that encryption is "natural mean for data protection, and for personal data protection as well: GDPR, in this sense, is reflecting a natural state"[4]. Indeed, ongoing research efforts in the field of Internet of Things (IoT), acknowledge that one of the most pressing concerns is developing lightweight encryption protocols suited to storage and computational power capacities of IoT devices, which are mostly resource-constrained, in order to secure data flow[5].

This paper aims to ascertain whether and to what extent state of the art cryptography may challenge the distinction, albeit increasingly blurred, between personal and non-personal data, underlying the European data protection legal framework.

From a methodological viewpoint, the main research question, namely how encrypted data may be classified from a data protection viewpoint, lays down the context of the legal analysis.

Section number two focuses on a technical perspective rather than a legal one, on the understandings of cryptography, encryption and pseudonymisation. Such technical taxonomy is necessary, given the blurring epistemological boundaries between personal and non-personal data.

The third section sets out the different interpretations of the GDPR to ascertain whether "encrypted" data qualifies as personal or not. Against this backdrop, three different

---

[1] Bruce Schneier, *Data and Goliath: The Hidden Battles to Collect Your Data and Control Your World* (WW Norton & Company, 2015), p. 131.

[2] Alex Pentland, "The data-driven society" (2013) 309 Scientific American 4, pp. 78–83.

[3] GDPR refers to encryption as a data protection safeguard several times: in article 32, encryption is enlisted prominently as security measure for personal data; when a data breach is notified, data protection authorities has to assess not only the impact that breach has created on human rights but also, from a technical viewpoint, what kind of mitigation measures were involved and eventually are going to be adopted.

[4] Wojciech Wiewiórowski, *Keynote: Data protection needs encryption,* EDPS, 1st Online IPEN Workshop, 3 June 2020.

[5] Ammar Rayes and Samer Salam, *Internet of Things: from Hype to Reality* (2019) 2nd edition, Springer, pp. 211ff; Dimitrios Serpanos and Marilyn Wolf, *Internet-of-Things (IoT) Systems: Architectures, Algorithms, Methodologies* (2018) Springer, pp 59-81; Peter Marwedel, *Embedded System Design: Embedded Systems Foundations of Cyber-Physical Systems, and the Internet of Things* (2018) 3rd edition, Springer, pp. 191-193.

understandings of Recital 26 GDPR are considered, i.e. the absolutist, the relativist and the risk-based approach.

The fourth section will cast the light on the polymorphic encryption framework and methodology. The legal analysis, integrated with technical considerations, aims at ascertaining whether the endorsement of one of the possible readings of Recital 26 GDPR is a kind of zero-sum game or, conversely, the relativist and risk-based approach can coexist alongside each other: if the relativist approach were to be applied, then many actors, such as cloud service providers, would eschew an intensive data protection regime. Moreover, the stance on the ambivalent relation of encryption towards human rights[6] will be discussed by proposing a decentralised key management scheme.

The fifth section then discusses the desirability of the outcome of the fourth section. The relativist approach, as regards the case-study under scrutiny, would result in having cloud service providers off the hook when it comes to GDPR enforcement. The question that this section seeks to address is whether this effect is advisable, in the light of recent business models.

Finally, the conclusion sums up the findings regarding the balance between the rationale of so-called digital security technologies (DSTs), i.e. to ensure confidentiality of communication in digital environments, and the problem at stake in the context of business trends, models and societal power, i.e. individuals profiling over aggregated, *allegedly* non personal data.

2. The quest for encryption, beyond pseudonymisation: a technical overview

The question whether encrypted data are personal or not, hinges on proper epistemological classification of the technical concepts of encryption and pseudonymisation. The aim of this section is thus to shed light on the relationship of the former with the latter.

It is worth clarifying from the outset that the word "encryption" itself could be misleading or simplistically broad: without dwelling on the various cryptographic algorithms too extensively, a classification is needed. Amongst cryptographic tools, a difference can be made between symmetric encryption, asymmetric encryption and cryptographic hash function.

The first type under scrutiny is symmetric or single-key encryption. Through mathematical operations, it aims at hiding the content (the so-called plain text, which is named cyphertext after the process is completed), by rendering the information unintelligible to anyone who does not have the cryptographic key[7]. Encryption is deemed as a two-way function: what is encrypted, *via* an algorithm called "cipher", can be decrypted with the proper key. As a result, lacking the

[6] Mireille Hildebrandt, "Digital security and human rights: a plea for counter-infringement" (2019) in Mart Susi (ed) *Human Rights, Digital Society and the Law: A Research Companion,* Taylor & Francis, p. 262.

[7] Bruce Schneier, *Applied Cryptography: Protocols, Algorithms, and Source Code in C* (1995), Wiley Inc., p. 21.

University of Vienna Law Review, Vol. 4 No 2 (2020), pp. 168-188, https://doi.org/10.25365/vlr-2020-4-2-168.

cryptographic key, even the most powerful actor (e.g. States or big tech companies[8]), theoretically, is not able to decrypt a message. The most widely used symmetric encryption algorithms are block ciphers[9].

The main difference between symmetric and asymmetric encryption is that the latter involves the use of two separate keys, whilst the former uses only one key[10]. In this approach, one of the keys (public key) can be widely distributed while the other key (secret key) must be kept secret[11].

Finally, unlike encryption, cryptographic hash functions are commonly referred to as one-way functions, i.e. an irreversible process aiming at scrambling variable-size plain text to produce a unique fixed-size message digest (i.e. the message output)[12], collision resistant[13]. In recent years, the Secure Hash Algorithm (SHA), developed by the federal US Agency NIST, became the most popular hash function[14].

The rationale of cryptography is therefore to protect *potentially* any kind of information. Indeed, it encompasses either data at rest (storage encryption) or data in motion (transmission encryption)[15].

---

[8] Leander Kahney, "The FBI Wanted a Back Door to the iPhone. Tim Cook Said No" *Wired* (16Apr 2019): "Apple and the government had been at odds for more than a year, since the debut of Apple's encrypted operating system, iOS 8, in late 2014. iOS 8 added much stronger encryption than had been seen before in smartphones. It encrypted all the user's data—phone call records, messages, photos, contacts, and so on—with the user's passcode. The encryption was so strong, not even Apple could break it".

[9] William Stallings and Lawrence Brown, *Computer Security: Principles and Practice* (2018) 4th edition, Pearson, p. 55: "a block cipher processes the plaintext input in fixed-size blocks and produces a block of ciphertext of equal size for each plaintext block. The algorithm processes longer plaintext amounts as a series of fixed-size blocks. The most important symmetric algorithms, all of which are block ciphers, are the Data Encryption Standard (DES), triple DES, and the Advanced Encryption Standard (AES)".

[10] Id., p. 67: authors immediately clear the field from potential misconceptions concerning public-key encryption. First, there is nothing in principle about either symmetric or public-key encryption that makes one superior to another from the point of view of resisting cryptanalysis or brute-force attacks; second, there seems no foreseeable likelihood that symmetric encryption will be abandoned for the benefit of asymmetric encryption.

[11] Richard R Brooks, *Introduction to Computer and Network Security: Navigating Shades of Gray* (2014) CRC Press, Talyor & Francis Group, p. 100.

[12] EDPS and AEPD, *Introduction to the Hash Function as a Personal Data Pseudonymisation Technique* (2019), pp. 8-10; Quynh Dang, *Recommendation for Applications Using Approved Hash Algorithms* (2012), NIST Special Publication 800-107, pp. 6-9; Richard R Brooks (fn 11), p. 91.

[13] William Stallings and Lawrence Brown (fn 9), p. 65: this property concerns the impossibility to "find an alternative message with the same hash value as a given message. This prevents forgery when an encrypted hash code is used".

[14] Bart Preneel, "The First 30 Years of Cryptographic Hash Functions and the NIST SHA-3 Competition" (2010) CT-RSA 2010: Topics in Cryptology, Springer, 1-14.

[15] Karen Scarfone, Murugiah Souppaya and Matt Sexton, *Guide to Storage Encryption Technologies for End User Devices* (2007), NIST Special Publication, 800-111.

University of Vienna Law Review, Vol. 4 No 2 (2020), pp. 168-188, https://doi.org/10.25365/vlr-2020-4-2-168.

Pseudonymisation, instead, is conceived by the GDPR[16] as a means of *reducing* risks to data subjects[17] by hiding the identity of individuals in a dataset, e.g. by replacing one or more personal data identifiers with the so-called pseudonyms (and appropriately protecting the link between the pseudonyms and the initial identifiers)[18]. Even though the (re)identification's risk is reduced, pseudonymised data fall nevertheless within the scope of GDPR: it is certainly true that such processing prevents direct identification through attribution, but not through the test set by Recital 26 and article 4 GDPR[19], which clearly specifies the personal nature of such data.

The relation link between those techniques, might be built upon a speciality criterion, based on the values they aim at securing. Encryption, thus, *can also be* used to protect the identities of individuals, whereas pseudonymisation's scope cannot consider other identifiers.

A second variable of the analysis of the techniques under scrutiny predicates on the value of the resulting text after such operations: contrary to encrypted data, pseudonymised data still provide some legible information and, thus, a third party (i.e. other than the controller or processor) may still understand the semantic (structure) of the data. Pseudonymisation is indeed considered by GDPR as an appropriate safeguard for any personal data processing for scientific, historical or statistical research[20].

Thirdly, these operations potentially overlap: encryption and hash function are in turn two possible ways to reach pseudonymisation's goals. Article 29 Working Party listed, among the most used pseudonymisation techniques, encryption with secret key (i.e. traditional two-way encryption), hash function, keyed hash function with stored key (so-called *pepper*[21]), keyed-hash

---

[16] GDPR, article 4(5): "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".

[17] GDPR, Recital 28.

[18] Konstantinos Limniotis and Marit Hansen, *Recommendations on Shaping Technology According to GDPR Provisions - An Overview on Data Pseudonymisation* (ENISA, 2018), p. 17.

[19] Miranda Mourby et al, "Are 'pseudonymised' data always personal data? Implications of the GDPR for administrative data research in the UK" (2018) 34 Computer Law and Security Review 2, p. 225: "[i]t is Recital 26 GDPR which must be used to establish whether data are personal".

[20] GDPR, Article 89 and Recital 156.

[21] William Stallings and Lawrence Brown (fn 9), p. 983: in cryptography, a *pepper* is a "random [and secret value] that is concatenated with a password before applying the one-way encryption function used to protect passwords that are stored in the database of an access control system". It differs from *salt,* since the latter is not secret (merely unique) and can be stored alongside the hashed output; Paul A Grassi et al., *Digital Identities Guidelines* (NIST Special Publication, 800-63B, 2017), sec. 5.1.1.2: NIST does not make any formal difference between *salt* and *pepper*, by referring to both as *salt*. The *pepper* value recommended is at least of 32 bits: the US agency assures that if the *pepper* value is kept secret, brute-force attacks on the hashed memorized secrets are impractical.

University of Vienna Law Review, Vol. 4 No 2 (2020), pp. 168-188, https://doi.org/10.25365/vlr-2020-4-2-168.

function with deletion of the key and tokenization[22]. A classificatory misconception underlies these notions: if GDPR mentions encryption and pseudonymisation as security measures alongside each other[23], Article 29 Working Party seems to categorize the former as a means of achieving the latter[24]. However, the main scopes of these techniques are different, albeit they may sometimes overlap, as it is well described in a recent ENISA report[25].

3. The personalisation debate in the cryptographic domain

In order to ascertain whether encrypted data qualifies as personal or not, the legal test to look at in the GDPR's dichotomous architecture, that is between personal and non-personal data, is based on Recital 26 GDPR[26]: in accordance with the GDPR, data is personal when the controller *or another person* can identify the data subject by using the "means reasonably likely to be used". The Recital goes further on specifying the personal nature of data undergone under pseudonymisation, whilst anonymisation techniques render GDPR inapplicable[27]. Even though encryption *could* be a means of pseudonymisation, (en)crypted data are not aprioristically categorised.

However, some tenets of this test are left in a space of uncertainty, resulting from conflicting interpretations and understandings by different supervisory authorities. Albeit article 4(1) GDPR lays down the definition of personal data, Recital 26 GDPR, as well as the earlier Recital 26 Directive 95/46, further specifies the test determining the scope of data protection[28]. Therefore, the analysis focuses on three possible readings of Recital 26.

---

[22] Article 29 Working Party, *Opinion 05/2014 on Anonymisation Techniques* (2014), WP 216, pp. 20-21.

[23] GDPR, Article 32(1).

[24] Article 29 Working Party (fn 22), pp. 20ff.

[25] Konstantinos Limniotis and Marit Hansen (fn 18), pp. 17ff.

[26] GDPR, Recital 26: "[p]ersonal 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. To determine whether a natural person is identifiable, account should be taken of all the means reasonably likely to be used, such as singling out, either by the controller or by another person to identify the natural person directly or indirectly. To ascertain whether means are reasonably likely to be used to identify the natural person, account should be taken of all objective factors, such as the costs of and the amount of time required for identification, taking into consideration the available technology at the time of the processing and technological developments.

[27] Id, "[n]amely 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.

[28] Case C-582/14 *Patrick Breyer* [2016] EU:C: 2016:779; see fn 19.

University of Vienna Law Review, Vol. 4 No 2 (2020), pp. 168-188, https://doi.org/10.25365/vlr-2020-4-2-168.

The first interpretation is predicated on the so-called absolute approach[29]. This understanding envisages all possibilities and chances in which anyone would be able to identify the data subject: while GDPR explicitly refers only to the possibility of "singling out"[30] an individual, the Working Party goes further, by adding to the de-identification test the criteria of "linkability"31 and "inference" 32. It results that "even theoretical chances of combining data so that the individual is identifiable are included"33. Thus, A29WP sets a high threshold to meet: as noted by Finck and Pallas, it seems to establish its own "zero-risk test" 34, hence implying that the "outcome of anonymisation as a technique applied to personal data should be, in the current state of technology, as permanent as erasure"35, i.e. making it impossible to process personal data. In the context of encryption, either symmetric or asymmetric, i.e. two-way cryptography, if anyone is theoretically able to decrypt the dataset, then "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"36. Interestingly, the Working Party, in an earlier opinion, stated that one-way cryptography (hash function) creates in general anonymised data37, but afterwards it claimed that hash functions are "usually designed to be relatively fast to compute"38. In so doing, A29WP implicitly grouped all cryptographic means under the personal label, so to speak. The absolute approach, nonetheless, can hardly be sustained: there is thriving literature about the non-absolute nature of anonymisation39. Thus, if we could never rely on the non-personality of data, then any information that was once within the scope of the GDPR would always remain personal data.

---

[29] Gerald Spindler and Philipp Schmechel, "Personal Data and Encryption in the European General Data Protection Regulation" (2016) 7 JIPITEC 163, p. 165; Christopher Kuner, *European Data Protection Law: Corporate Compliance and Regulation* (2nd edition, Oxford University Press, 2007), p. 92.

[30] Article 29 Working Party (fn 22), p. 11: singling out refers to "the possibility to isolate some or all records which identify an individual in the dataset".

[31] Id, p. 11: *linkability* refers to the risk where attacker can establish (e.g. by means of correlation analysis) that two records are assigned to a same group of individuals but cannot single out individuals in this group

[32] Id, p. 12: *inference* has been envisaged as "the possibility to deduce, with significant probability, the value of an attribute from the values of a set of other attributes".

[33] Gerald Spindler and Philipp Schmechel (fn 29), p. 165.

[34] Michèle Finck and Frank Pallas, "They who must not be identified—distinguishing personal from non-personal data under the GDPR" (2020) 10 International Data Privacy Law 1, p. 15

[35] Article 29 Working Party (fn 22), p. 6.

[36] Gerald Spindler and Philipp Schmechel (fn 29), p. 165.

[37] Article 29 Working Party, *Opinion 4/2007 on the concept of personal data* (2007), WP 136, p. 18.

[38] Article 29 Working Party (fn 22), p. 20.

[39] Paul Ohm, "Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization" (2010) 57 UCLA Law Review 2, pp. 1701ff; Latanya Sweeney, "Simple Demographics Often Identify People Uniquely" (2000)

University of Vienna Law Review, Vol. 4 No 2 (2020), pp. 168-188, https://doi.org/10.25365/vlr-2020-4-2-168.

The second understanding of Recital 26 considers the relativist approach. As underlined by Spindler and Schmechel, the role of the data controller and her effort in establishing a link between a person and the data is crucial in order to understand the extent of the concept of personal data[40]. This reading takes only into account the efforts required by data controllers to identify an individual, thus setting aside mere theoretical possibilities[41]. In the context of cryptography, several authors argued that data resulting from such operations should not be considered personal data if two requirements are met: the cryptographic method must be effective, solid and up to date[42] and the data controller (or any other third party) is either not in possession of the decryption key or she has no reasonable chances to obtain the key[43]. Even though the relativist stance has been endorsed in other contexts as well (e.g. when the Commission decided that transferring key-coded data to the U.S. would not be a transmission of personal data, as the decryption key had not been sent together with the data[44]; the ECJ, when ruling on "Safe Harbour's" validity[45], has not altered the Commission's view on that point[46]), this reasoning was particularly successful in the field of cloud computing[47]: leaving aside cryptography as a means of pseudonymisation[48], two-way encryption (either symmetric or asymmetric) of full

671 Health, pp. 1–34; Arvind Narayanan and Vitaly Shmatikov, "Robust De-anonymization of Large Sparse Dataset" (2008) IEEE Symposium on Security and Privacy, pp. 111–125.

[40] Gerald Spindler and Philipp Schmechel (fn 29), p. 165.

[41] Id, p. 166.

[42] The evaluation would mainly consider three factors: algorithm's cryptographic strength; encryption key length; security of decryption key storage.

[43] Samson Esayas, "The Role of Anonymisation and Pseudonymisation Under the EU Data Privacy Rules: Beyond the 'All or Nothing' Approach" (2015) 6 European Journal of Law and Technology 2, pp. 8-9; Patrick Lundevall-Unger and Tommy Tranvik, "IP Addresses – Just a Number?" (2011) 19 International Journal of Law and Information Technology 1, p. 53.

[44] Commission Decision 2000/520/EC of 26 July 2000 pursuant to Directive 95/46/EC of the European Parliament and of the Council on the adequacy of the protection provided by the safe harbour privacy principles and related frequently asked questions issued by the US Department of Commerce [2000] OJ L215/7, p. 24.

[45] Case C-362/14 *Maximiliam Schrems* [2015] EU:C: 2015:650.

[46] Commission Implementing Decision (EU) 2016/1250 of 12 July 2016 pursuant to Directive 95/46/EC of the European Parliament and of the Council on the adequacy of the protection provided by the EU-U.S. Privacy Shield [2016] OJ L207, 66. The ECJ, when ruling the so-called *Schrems II* (Case C-311/18 *Facebook Irland and Schrems* [2020] EU:C:2020:559), has once again invalidated the Commission's decision: nonetheless, the Court has not taken a stance on the so-called key-coded data.

[47] OPTIMIS Project D7.2.1.2., *Cloud Legal Guidelines: Data Security, Ownership Rights and Domestic Green Legislation* (2011), 8; Kuan W Hon, Christopher Millard and Ian Walden, "The problem of 'personal data' in cloud computing: what information is regulated? - the cloud of unknowing" (2011) 1 International Data Privacy Law 4, p. 217.

[48] Kuan W Hon, Christopher Millard and Ian Walden (fn 47), pp. 218-219: "[i]rreversibly hashing direct identifiers cannot prevent identification through indirect identifiers, other information in the dataset, and/or other sources.

University of Vienna Law Review, Vol. 4 No 2 (2020), pp. 168-188, https://doi.org/10.25365/vlr-2020-4-2-168.

datasets could be deemed non-personal data in the provider's control if "within the specific scheme in which those other controllers (e.g. cloud service providers) are operating reidentification is explicitly excluded and appropriate technical measures have been taken in this respect"[49]. From the reading of Finck and Pallas, the understanding of A29WP is granitic, crystallised and flattened, so to speak, on an absolutist interpretation of Recital 26[50]. Nevertheless, the abovementioned A29WP Opinion on the concept of personal data, precisely regarding *key-coded* data, arguably assumes a relativist understanding. Moreover, in the context of irreversible hashing, "[e]ven if identification of certain data subjects may take place despite all those protocols and measures (due to unforeseeable circumstances such as accidental matching of qualities of the data subject that reveal his/her identity), the information processed by the original controller may not be considered to relate to identified or identifiable individuals taking account of all the means reasonably likely to be used by the controller or by any other person"[51]. It is an open question whether A29WP's absolutist stance, clearly highlighted in the Opinion on Anonymisation Techniques, is to resist.

The last reading of Recital 26 hinges on the risk-based approach. The recent work of Fink and Pallas revolves around the thesis whereby the safest and most correct interpretation of Recital 26 needs to be risk-orientend, thus completely discarding an absolutist view, supported *inter alia* by A29WP. It has been briefly said above that anonymisation, which has as an outcome non-personal data, can never be absolute; rather, risks remain. They argue that Recital 26 GDPR embraces a risk-based approach, following the very inspiration of the Regulation, to assess the personal nature of data. "If data can be matched to a natural person with reasonable likelihood, it qualifies as personal data and falls within the GDPR's scope of application. If de-personalization has been sufficiently strong so that identification is no longer reasonably likely, this is non-personal data and accordingly falls outside the Regulation's scope of application"[52]. They take as an argument the leading case in this matter, i.e. *Breyer*[53]. It is worth noticing that *Breyer* confirmed such an approach as the Court evaluated the actual risk of identification[54]. In a nutshell, the Court assessed the personal nature of Mr. Breyer's dynamic IP address even if the data required for his

Thus, personal data where identifiers have been deleted or one-way hashed may, after considering such 'means likely reasonably to be used', remain 'personal data'".

[49] Article 29 Working Party (fn 37), p. 20.

[50] Michèle Finck and Frank Pallas (fn 34), p. 15: "[t]his strict position is in line with earlier guidance from 2007 according to which anonymized data is data "that previously referred to an identifiable person, but where that identification is no longer possible".

[51] Article 29 Working Party (fn 22), p. 20

[52] Michèle Finck and Frank Pallas (fn 34), p. 34.

[53] see fn 28.

[54] Frederik Z Borgesius, "The Breyer Case of the Court of Justice of the European Union: IP Addresses and the Personal Data Definition" (2017) 3 European Data Protection Law Review 1, pp. 130–137.

identification were not held by German authorities, i.e. the data controller in such scenario, but by the internet service provider, i.e. a third party[55]. This stance dismisses a *pure* relativist approach, as "there is no requirement that all the information enabling the identification of the data subject must be in the hands of one person"[56]. Nevertheless, as pointed out by Finck and Pallas[57], an English court recently adopted a prudent understanding of *Breyer*, arguing that the possibility accorded by the law to the controller (i.e. German authorities, in *Breyer* case) to gain access to data to "identify a natural person would [not] make that procedure a means reasonably likely to be used"[58]. Once again, the relative criterion is getting closer.

As regard to cryptography, they value of the study of Finck and Pallas lies precisely in demonstrating that state-of-the-art cryptographic hash functions, such as SHA-3[59], alongside an appropriate pepper length[60], shall eschew the Regulation's scope of application. Thus, "with the even more resistant bcrypt-hashes, however, 32 bits of pepper would lead to more than 140 years with 10,000 current GPUs"[61], which is definitely not an effort that can possibly be deemed as "reasonably likely". EDPS and AEPD (Spanish Agency for Data Protection) ended up with the same result, albeit with a more cautious approach[62]. Therefore, it seems safe to assume that there is no more room for upholding the absolutist approach, when it comes to categorise data from a data protection viewpoint.

Provided that up-to-date one-way encryption should be considered non-personal data, what about two-way encryption? Is that *tout-court* meant to be personal data?

---

[55] Case C-582/14 (fn 52), para. 49.

[56] Id, para. 43; see also Article 29 Working Party, *Opinion 03/2003 on Purpose Limitation* (WP203, 2003), p. 31: "a very significant grey area, where a data controller may believe a dataset is anonymised, but a motivated third party will still be able to identify at least some of the individuals from the information released".

[57] Michèle Finck and Frank Pallas (fn 34), p. 18.

[58] Mircom International Content Management & Consulting Ltd v Virgin Media Ltd (EWHC 1827, 2019), p. 27.

[59] Morris J Dworkin, *SHA-3 Standard: Permutation-Based Hash and Extendable-Output Functions* (NIST FIPS, 202, 2015), sec. v: together with SHA-1 and SHA-2 families, thesesis standards will "[p]rovide resilience against future advances in hash function analysis, because they rely on fundamentally different design principles. In addition to design diversity, the hash functions in this Standard provide some complementary implementation and performance characteristics to those in FIPS 180-4".

[60] See fn 23: pepper is a random [and secret] value that is concatenated with a password before applying the one-way encryption function used to protect passwords that are stored in the database of an access control system.

[61] Michèle Finck and Frank Pallas (fn 34), p. 26.

[62] EDPS and AEPD (fn 12), pp. 17-18.

University of Vienna Law Review, Vol. 4 No 2 (2020), pp. 168-188, https://doi.org/10.25365/vlr-2020-4-2-168.

4. One size does not fit all: the case of polymorphic encryption as an argument in favour of the relativist approach.

The aim of this section is to investigate whether a relative understanding of Recital 26 may be applied alongside the risk-based approach. It will discuss whether the endorsement of one of the above-mentioned readings of Recital 26 GDPR is a kind of zero-sum game.

The truly interdisciplinary nature of the claim to be asserted makes it necessary to look for arguments to support it outside the social sciences; this would justify the recourse to computer science. The case of polymorphic encryption, indisputably a two-way function (i.e., reversible), may be a case in point. In their white paper, the authors (i.e. Verheul, Jacobs, Meijer, Hildebrandt and de Ruiter) present a novel approach for the management of sensitive personal data, especially in health care: the core concept is to put the data subject at the heart of encryption's operational and decision-making process[63], by providing for the necessary security and privacy infrastructure for big data analytics where data comes from various sources[64], likewise in IoT scenarios[65]. Even though a proper analysis on differential privacy[66] falls outside the scope of the paper, it is worth mentioning that Apple, as Pagallo observes[67], mastered such techniques in its endeavours to process health data for statistical purposes, eschewing therefore GDPR's regime[68].

Polymoprhic Encryption and Pseudonymisation methodology (hereafter, PEP) hinges on both encryption and pseudonymisation: albeit fundamental for the functioning of the process, the aim of this study is to focus on the former. The assumption is that traditional asymmetric encryption (therein called "public key encryption") is too rigid: there is only one "key" able to decipher the information encrypted. The ground-breaking concept behind this process is predicated on two consequential steps. First, this technology enables strong encryption at the source, so to speak, thus ensuring that either transport or Cloud facility's storage occurs to data which have been already encrypted: "polymorphic encryption works in a generic manner, and the decisions about

[63] Eric Verheul et al, "Polymorphic Encryption and Pseudonymisation for Personalised Healthcare" (White Paper, Institute for Computing and Information Sciences Radboud University Nijmegen, 2016).

[64] Eric Verheul and Bart Jacobs, "Polymorphic encryption and pseudonymisation in identity management and medical research" (2017) 5 NAW 18, p. 168.

[65] Lukas Malina et al, "A Privacy-Enhancing Framework for Internet of Thing Services" (2019) International Conference on Network and System Security, Springer, pp. 77-98.

[66] Aaron Roth and Cynthia Work, "The Algorithmic Foundations of Differential Privacy, Foundation and Trends" (2014) 9 Theoretical Computer Science 3-4, pp. 211-407.

[67] Ugo Pagallo, "The Legal Challenges of Big Data: Putting Secondary Rules First in the Field of EU Data Protection" (2017) 3 European Data Protection Law Review 1, p. 36.

[68] Pursuant to Recital 162 GDPR, statistical purposes are "any operation of collection and the processing of personal data necessary for statistical surveys or for the production of statistical results".

who can decrypt need not be taken at the time of encryption"[69]. It follows that "no data-management staff, hosting partner or cloud-service provider has the ability to access and decrypt the data"[70]. For each data subject, it shall be assumed that there is a private master key *x,* securely stored by a trusted Key Server, in secure hardware: it is never used for decryption[71]. Second, the data subject plays a crucial role in deciding who can decrypt such data (most likely the decision will be based upon policy's terms)[72]: anyone chosen for decrypting (personal) data will be given a *unique* private key, derived from the one and only private master key *x*. Afterwards, an intermediate party, called the Tweaker, grants access to the chosen parties by *re-keying* messages encrypted with the public master key: the Tweaker "is a central converter who exclusively knows how to turn the wheel on a polymorphic lock so that keys of specific parties fit"[73]. It is however crucial noting that the Tweaker works blindly: it does not know *x*; hence it cannot see the actual content of the information she is giving the access for[74].

This implementation of polymorphic encryption therefore seems to explicitly recall a relativist vision of Recital 26 GDPR: the perspective from which identifiability shall be assessed is that of data controller[75] precisely because until the tweaker grants her access, neither she nor any other third party can decrypt the data encrypted at the source. The application protocols of this cryptographic means are by their essence "relative", since cyphertext, hence possibly sensitive personal data, is disclosed upon data subject's choice to specific (and trusted) parties.

Assuming that it is not a zero-sum game, one can possibly apply the risk-based understanding *within* this relativist approach: it has to be analysed therefore to what extent PEP security assumptions can be deemed solid so identification is reasonably unlikely *but for* the chosen parties who can decrypt data. Verheul et al. delve into this from a technical viewpoint by

---

[69] Eric Verheul et al (fn 62), p. 8.

[70] Bart Jacobs and Jean Popma, "Medical research, Big Data and the need for privacy by design" (2019) 6 Big Data & Society 1, p. 4.

[71] Eric Verheul et al (fn 62), p. 32.

[72] Eric Verheul and Bart Jacobs (fn 63), p. 168; Eric Verheul et al (fn 62), p. 19: "PEP can be based on the first ground, consent, which is usually combined with a privacy policy or terms of service. We believe that its aims are better achieved by the introduction of a (modular) `data licensing agreement' (DLA) that makes sure that data are only processed insofar as necessary for the performance of the agreement by a party that is not allowed to share the data with other parties. The latter should always conclude their own DLA to obtain their own key. This ensures that data subjects have a clear overview of the parties that process their sensitive data".

[73] Eric Verheul et al (fn 62), p. 7: the Tweaker has a central role in assigning *local* or *different* pseudonyms as well. Thus, each data subject will have different pseudonym at different parties, since "[t]hese parties could somehow lose their data, or even maliciously combine data with others. If different parties use different pseudonyms for the same patient, it is in principle not possible to combine the data, at least not on the basis of identifiers" (11).

[74] Eric Verheul and Bart Jacobs (fn 63), p. 168; Eric Verheul et al (fn 62), p. 8.

[75] Michèle Finck and Frank Pallas (fn 33), p. 8.

University of Vienna Law Review, Vol. 4 No 2 (2020), pp. 168-188, https://doi.org/10.25365/vlr-2020-4-2-168.

concluding that their "re-randomized versions of polymorphic pseudonyms, encrypted pseudonyms and encrypted data [...] are not linkable to the original version"[76]. Following a relativist tenet, these encrypted data should be considered non-personal to anyone who does not have access to them, e.g. Key Server, Tweaker, Cloud Service Provider and ideally *any other third* party. Nevertheless, the security of the system rests inevitably on having two separate trusted parties, one holding the master private key, and one 'transformer', i.e. the Tweaker, holding the key factors for each service provider. If these two trusted parties collude, the system breaks down[77]. The question, now, is whether collusion between those parties is reasonably unlikely. If it were to follow the reasoning of the Court in *Breyer,* identification would not be reasonably likely if achieved by means prohibited by the law[78]. However, as Purtova pointed out, since re-identification results indeed from illegal acts, "a more nuanced reasoning would be that a legal prohibition to combine data for identification would make the means of identification 'less reasonably likely to be used', rather than 'not reasonably likely'"[79]. Anyhow, the *Breyer* test is passed.

4.1 Increasing trust by implementing a distributed key-management scheme

Albeit encryption is widely acknowledged as a guarantee for data protection and information security, the dependency on trusted third parties for key management (e.g. the master key server and the Tweaker, in the vein of PEP framework) and certification led Hildebrandt to conclude that even such *digital security technology* (hereafter called DSTs) is not neutral to human rights, as it spurs "new vulnerabilities that require further DSTs to detect fraudulent third parties or attacks against trusted platforms"[80]. The consideration that the higher the number of keys, the higher the risk would relate to a systemic problem in the context of polymorphic encryption. Risk, thus, is in a dependency relation with trust: where trust is, there is risk[81].

[76] Eric Verheul et al (fn 62), p. 55.

[77] Eric Verheul and Bart Jacobs (fn 63), p. 170.

[78] Case C-582/14 (fn 52), para 46.

[79] Nadezhda Purtova, "The law of everything. Broad concept of personal data and future of EU data protection law" (2018) 10 Law, Innovation and Technology 1, pp. 64-65.

[80] Mireille Hildebrandt, "Digital security and human rights: a plea for counter-infringement" (2019) in Mart Susi (ed) Human Rights, Digital Society and the Law: A Research Companion, Taylor & Francis, p. 266.

[81] Massimo Durante, "Safety and Security in the Digital Age. Trust, Algorithms, Standards, and Risks" (2019) in Don Berkich and Matteo Vincenzo D'Alfonso (eds.) *On the Cognitive, Ethical, and Scientific Dimensions of Artificial Intelligence,* Springer Nature, p. 380.

University of Vienna Law Review, Vol. 4 No 2 (2020), pp. 168-188, https://doi.org/10.25365/vlr-2020-4-2-168.

Yet, the decentralised re-keying scheme can arguably lower the risk by avoiding potential collusions[82]: the trust in the overall system would increase accordingly. Through the adoption of several security protocols, such as Shamir's secret sharing[83], Ismail et al. propose a scheme allowing the data subject to get total control over the generation and management of the decryption keys without relying on a trusted authority[84]: if two or more parties (eg, Cloud service providers) combine the keys, they cannot decrypt data[85]. Thus, there is much thriving literature confirming that secret sharing algorithms can be successfully applied either to distribute the encryption key among a number of cloud nodes[86] or to divide data subject's encrypted data into shares to be stored in different cloud service providers[87]. Notwithstanding, is an outcome that softens data protection legal framework towards actors with tremendous computational power even just desirable?

5. A reason for concern: paving the way to large data-driven companies heaven?

Adopting the relativist approach could bring a scenario where cloud service providers, for instance, would see their legal obligations lowered, so to speak, since they will not have to follow data protection principles. All in all, European personal data protection regime, i.e. the GDPR, embraces many overarching principles such as availability, integrity, accountability and plenty more: data confidentiality is not the only value at stake. The problem of our increasingly data-driven society is primarily with those who wish to analyse data without considering them personal. Indeed, there are reasons for concern to go down the road of finding ways to exempt intermediaries and cloud services from personal data obligations because it allows them to process data for their own ends coupled with lowered safeguards for data subjects.

Against this backdrop, attention shall be turned to what is happening between two known big data-driven corporations: Google and Mastercard. the former knows who actually viewed an

---

[82] Michael Egorov, Wilkison MacLane and David Nuñez, "NuCypher KMS: Decentralized key management system" (2017) arXiv preprint arXiv:1707.06140.

[83] Adi Shamir, "How to Share a Secret" (1979) 22 Communications of the ACM 11, pp. 612-613.

[84] Tayssir Ismail et al, "Hybrid and Secure E-Health Data Sharing Architecture in Multi-Clouds Environment" (2020) International Conference on Smart Homes and Health Telematics, Springer, Cham, p. 257.

[85] Id, p. 253.

[86] Roy D'Souza et al, "Publicly verifiable secret sharing for cloud-based key management" (2011) Progress in Cryptology– INDOCRYPT Berlin, Springer, 290–309; Jonathan Gill et al, "SYSTEM: Secure Cloud Storage, Auditing, and Access Control for Electronic Health Records" (2012) Working Paper, Department of Computer Science University of Illinois at Urbana-Champaign.

[87] Tatiana Ermakova and Benjamin Fabian, "Secret sharing for health data in multiprovider clouds" (2012) International Conference on Business and Informatics, 93–100; Hanlin Zhang et al, "Cloud Storage for Electronic Health Records Based on Secret Sharing With Verifiable Reconstruction Outsourcing" (2018) 6 IEEE Access, pp. 40713 – 40722.

advertisement and the latter knows who purchased a product and how much has been spent[88]As highlighted by Michael Veale, the question ultimately boils down to whether it is feasible to build a model tailored for targeted and personalised advertisement by combining the two aggregated datasets without seeing personal data of the other party[89]. Long story short, yes. It happened.

According to Veale, by using a cryptosystem, called private set intersection[90], partially based upon homomorphic encryption[91], Google and Mastercard eventually paired up, resulting as output of such collaboration an aggregate non-personal dataset on total expenditures of those who saw advertisements. Their next logical step seems to be getting rid of the initial personal data from the process, to further distance the firm from data protection legal regime: for example, these data could be put on a browser or on a user device[92]. The latter perfectly casts the light on recent trends in data management within the world of (big) data-driven companies. There is a thriving research in designing models for users profiling without data leaving their devices: using MPC or homomorphic encryption, companies train shared models based on tracking data that never leaves an individual's phone. This trend is particularly exacerbated by walled gardens, like IoS, resulting in practical inability of users to check what code is running on their systems. Against this backdrop, a secure multi-party computation protocol could be considered, since it allows many actors to collectively compute a function over aggregated data which may not be considered personal[93]. Each subject holds pieces of, without revealing what she effectively holds to any other player, i.e. the content of data. Accordingly, large-scale actors are investing more and more in training and improving machine learning classifiers to extract knowledge from such aggregated data. As in the case of polymorphic encryption scheme, they would not have access to such data but they would have access to the end result of the aggregation process (in the Google/Mastercard example, they could see users' total expenditures based on viewed advertisements). Thus, several authors proposed to conceive such machine learning models trained and working over allegedly non personal data *as personal data* in order "to re-balance or at least disrupt the power relations

---

[88] Mark Bergen and Jennifer Surane, "Google and Mastercard cut a secret ad deal to track retail sales" *Bloomberg* (30 Aug 2018).

[89] Michael Veale, "Knowing without seeing: informational power, cryptosystems and the law" (2019) available at: https://suri.epfl.ch/slides/2019/michael-veale.pdf.

[90] Yan Huang, David Evans and Jonathan Katz, "Private Set Intersection: Are Garbled Circuits Better than Custom Protocols?" (2012) NDSS Symposium.

[91] Craig Gentry, "Fully homomorphic encryption using ideal lattices" (2009) STOC '09: Proceedings of the forty-first annual ACM symposium on Theory of computing, pp. 169-178.

[92] Weisong Shi, Jie Cao, Quan Zhang, Youhuizi Li and Lanyu Xu, "Edge computing: vision and challenges" (2016) 3 IEEE Internet of Things Journal 5, pp. 637-646.

[93] Peter Laud and Liina Kamm (eds), *Applications of Secure Multiparty Computation* (2015) Cryptology and information security series volume 13, IOS Press.

between those holding models and those whose data are used to train them"[94]. Reducing liability and accountability is the actual dream of large companies, dangerously paved by models that aim at narrowing the scope of data protection legal regime.

6. Conclusion

In sketching some conclusive remarks, the risk-based approach, as opposed to absolutist understanding, should be recognised as the correct interpretative key for the reading of recital 26. Yet, an attempt has been made to illustrate how the case of polymorphic encryption can be theoretically used as an argument for the revival of a debate on the relativistic approach in the post-Breyer period, in a perspective of complementarity and not exclusion of the risk-based one. Thus, a solid technical risk assessment shall always be carried out to ascertain to what extent re-identification is reasonably likely. What has been tried to argue is that those data, encrypted with specific two-way functions, such as PEP methodology, shall be considered non-personal to all those who are denied the possibility of decryption. By putting the data subject at the very centre of operational decisions in the context of data encryption, the control's paradigm would be rebalanced towards the data subjects' side, albeit data controllers remain accountable for the processing: the user is empowered to decide each time, and not necessarily at the time of encryption, who can access (parts of) her data and for which purpose.

A combined application of the two criteria mentioned above could ultimately mitigate the highly intensive and non-scalable regime enforced by GDPR: the technical instruments could therefore restore a renewed equilibrium in the unbalanced relationship between personal and non-personal data[95], eschewing for the moment the scenario outlined by Purtova[96].

Notwithstanding, encryption is done for a purpose. The last section aims at casting the light on the balance between the rationale of so-called digital security technologies (DSTs),[97] and the problem at stake in the context of business trends, desires, business models and societal power. When discussing encryption schemes, it is more and more necessary to holistically understand the processing purpose, what actors are trying to achieve, what are the possible outcomes if different legal regimes apply. All in all, altering the personal data definition could be dangerous, as it would reduce responsibilities (e.g. security principles ex art.32 GDPR; data protection by

[94] Michael Veale, Reuben Binns and Lilian Edwards, "Algorithms that remember: model inversion attacks and data protection law" (2018) Philosophical transactions of a royal society, p. 12.

[95] Inge Graef, Raphael Gellert and Martin Husovec, "Towards a holistic regulatory approach for the European data economy: Why the illusive notion of non-personal data is counterproductive to data innovation" (2019) 44 European Law Review 5, pp. 605-621.

[96] Nadezhda Purtova (fn 78), p. 75.

[97] See fn 7, p. 262.

University of Vienna Law Review, Vol. 4 No 2 (2020), pp. 168-188, https://doi.org/10.25365/vlr-2020-4-2-168.

design ex art. 25 GDPR; limitation and accountability principles ex art. 5 GDPR) for certain actors over others. Definitional changes in the context of the personalisation debate need to be seen in terms of much wider changes in encrypted computation to avoid a scenario where conflating competition, privacy and data protection would result into a huge unsolvable problem.

7. Bibliography

Article 29 Working Party, Opinion 4/2007 on the concept of personal data (2007), WP 136

Article 29 Working Party, Opinion 05/2014 on Anonymisation Techniques (2014), WP 216

Bergen M and Surane J, "Google and Mastercard cut a secret ad deal to track retail sales" Bloomberg (30 Aug 2018)

Borgesius FZ, "The Breyer Case of the Court of Justice of the European Union: IP Addresses and the Personal Data Definition" (2017) European Data Protection Law Review 3(1)

Brooks RR, Introduction to Computer and Network Security: Navigating Shades of Gray (2014) CRC Press, Talyor & Francis Group

Case C-362/14 Maximiliam Schrems [2015] EU:C: 2015:650

Case C-311/18 Facebook Irland and Schrems [2020] EU:C:2020:559

Commission Decision 2000/520/EC of 26 July 2000 pursuant to Directive 95/46/EC of the European Parliament and of the Council on the adequacy of the protection provided by the safe harbour privacy principles and related frequently asked questions issued by the US Department of Commerce [2000] OJ L215/7

D'Souza R et al, "Publicly verifiable secret sharing for cloud-based key management" (2011) Progress in Cryptology– INDOCRYPT Berlin, Springer

Dang Q, "Recommendation for Applications Using Approved Hash Algorithms" (2012), NIST Special Publication 800-107

Durante M, "Safety and Security in the Digital Age. Trust, Algorithms, Standards, and Risks" (2019) in Don Berkich and Matteo Vincenzo D'Alfonso (eds.) On the Cognitive, Ethical, and Scientific Dimensions of Artificial Intelligence, Springer Nature

EDPS and AEPD, Introduction to the Hash Function as a Personal Data Pseudonymisation Technique (2019)

Egorov M, MacLane W and Nuñez D, "NuCypher KMS: Decentralized key management system" (2017) arXiv preprint arXiv:1707.06140

Ermakova T and Fabian B, "Secret sharing for health data in multiprovider clouds" (2012) International Conference on Business and Informatics

Esayas S, "The Role of Anonymisation and Pseudonymisation Under the EU Data Privacy Rules: Beyond the 'All or Nothing' Approach" (2015) European Journal of Law and Technology, 6(2)

Finck M and Pallas F, "They who must not be identified—distinguishing personal from non-personal data under the GDPR" (2020) International Data Privacy Law, 10(1)

Floridi L, "The Method of Levels of Abstraction" (2008) Minds and Machines 18

Gentry C, "Fully homomorphic encryption using ideal lattices" (2009) STOC '09: Proceedings of the forty-first annual ACM symposium on Theory of computing

Graef I, Gellert R and Husovec M, "Towards a holistic regulatory approach for the European data economy: Why the illusive notion of non-personal data is counterproductive to data innovation" (2019) European Law Review, 44(5)

Hildebrandt M, "Digital security and human rights: a plea for counter-infringement" (2019) in Mart Susi (ed) Human Rights, Digital Society and the Law: A Research Companion, Taylor & Francis

Hon KW, Christopher Millard and Ian Walden, "The problem of 'personal data' in cloud computing: what information is regulated?—the cloud of unknowing" (2011) International Data Privacy Law, 1(4)

Huang Y, Evans D and Katz J, "Private Set Intersection: Are Garbled Circuits Better than Custom Protocols?" (2012) NDSS Symposium

Ismail T et al, "Hybrid and Secure E-Health Data Sharing Architecture in Multi-Clouds Environment" (2020) International Conference on Smart Homes and Health Telematics, Springer, Cham

Jacobs B and Popma J, "Medical research, Big Data and the need for privacy by design" (2019) Big Data & Society 6(1)

Kahney L, "The FBI Wanted a Back Door to the iPhone. Tim Cook Said No" Wired (16 Apr 2019)

Kuner C, *European Data Protection Law: Corporate Compliance and Regulation* (2007) 2nd edition, Oxford University Press

Laud P and Kamm L (eds), *Applications of Secure Multiparty Computation* (2015) Cryptology and information security series volume 13, IOS Press

Limniotis K and Hansen M, "Recommendations on Shaping Technology According to GDPR Provisions - An Overview on Data Pseudonymisation" (2018) ENISA

Malina L et al, "A Privacy-Enhancing Framework for Internet of Thing Services" (2019) International Conference on Network and System Security, Springer

Marwedel P, *Embedded System Design: Embedded Systems Foundations of Cyber-Physical Systems, and the Internet of Things* (2018), 3rd edition, Springer

Mourby M et al, "Are 'pseudonymised' data always personal data? Implications of the GDPR for administrative data research in the UK" (2018) Computer Law and Security Review, 34(2)

Narayanan A and Shmatikov V, "Robust De-anonymization of Large Sparse Dataset" (2008) IEEE Symposium on Security and Privacy

Ohm P, "Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization" (2010) UCLA Law Review, 57(2)

Pagallo U, "The Legal Challenges of Big Data: Putting Secondary Rules First in the Field of EU Data Protection" (2017) European Data Protection Law Review, 3(1)

Pagallo U, Durante M and Monteleone S, "What Is New with the Internet of Things in Privacy and Data Protection? Four Legal Challenges on Sharing and Control in IoT" in Leenes R, van Brakel R, Gutwirth S, De Hert P (eds.), Data Protection and Privacy: (In)visibilities and Infrastructures (2017) Law, Governance and Technology Series, Springer

Pentland A, "The data-driven society" (2013) Scientific American, 309(4)

Preneel B, "The First 30 Years of Cryptographic Hash Functions and the NIST SHA-3 Competition" (2010) CT-RSA 2010: Topics in Cryptology, Springer

Purtova N, "The law of everything. Broad concept of personal data and future of EU data protection law" (2018) Law, Innovation and Technology 10(1)

Rayes A and Salam S, *Internet of Things: from Hype to Reality* (2019) 2nd edition, Springer

Roth A and Work C, "The Algorithmic Foundations of Differential Privacy, Foundation and Trends" (2014) Theoretical Computer Science, 9(3-4)

Scarfone K, Souppaya M and Sexton M, "Guide to Storage Encryption Technologies for End User Devices" (2007), NIST Special Publication, 800-111

Schneier B, *Data and Goliath: The Hidden Battles to Collect Your Data and Control Your World* (2015) WW Norton & Company

Schneier B, *Applied Cryptography: Protocols, Algorithms, and Source Code in C* (1995), Wiley Inc.

Serpanos D and Wolf M, I*nternet-of-Things (IoT) Systems: Architectures, Algorithms, Methodologies* (2018), Springer

Shamir A, "How to Share a Secret" (1979) Communications of the ACM 22(11)

Shi W, Cao J, Zhang Q, Li Y and Xu L, "Edge computing: vision and challenges" (2016) IEEE Internet of Things Journal, 3(5)

Spindler G and Schmechel P, "Personal Data and Encryption in the European General Data Protection Regulation" (2016) JIPITEC 163(7)

Stallings W and Brown L, *Computer Security: Principles and Practice* (2018) 4th edition, Pearson

Sweeney L, "Simple Demographics Often Identify People Uniquely" (2000) Health, 671

Unger L and Tranvik T, "IP Addresses – Just a Number?" (2011) International Journal of Law and Information Technology, 19(1)

Veale M, Binns R and Edwards L, "Algorithms that remember: model inversion attacks and data protection law" (2018) Philosophical transactions of a royal society

Veale M, "Knowing without seeing: informational power, cryptosystems and the law" (2019) available at: https://suri.epfl.ch/slides/2019/michael-veale.pdf

Verheul E and Jacobs B, "Polymorphic encryption and pseudonymisation in identity management and medical research" (2017) Nieuw Archief voor Wiskunde, 5(18)

Verheul E et al, "Polymorphic Encryption and Pseudonymisation for Personalised Healthcare" (2016) White Paper, Institute for Computing and Information Sciences Radboud University Nijmegen

Wiewiórowski W, Keynote: Data protection needs encryption, EDPS, 1st Online IPEN Workshop, 3 June 2020

Zhang H et al, "Cloud Storage for Electronic Health Records Based on Secret Sharing With Verifiable Reconstruction Outsourcing" (2018) 6 IEEE Access