European Journal of **Law and Technology** Vol 6, No 2 (2015)

The role of anonymisation and pseudonymisation under the EU data privacy rules: beyond the 'all or nothing' approach

Samson Yoseph Esayas[1]

Cite as Esayas S. Y., "The role of anonymisation and pseudonymisation under the EU data privacy rules: beyond the 'all or nothing' approach", in European Journal of Law and Technology, Vol 6, No 2, 2015.

ABSTRACT

Substantial uncertainty exists on the role of anonymised or pseudonymised data in the data privacy discourse; this is all the more so as de-anonymisation science advances and the ubiquity of information increases. Such uncertainty affects not only the wider usage of such measures but also creates the temptation, both on the part of the entities that process personal data and the individuals whose personal data is processed, to downplay privacy risks associated with anonymised or pseudonymised data. Crucial to mitigating such risks and promoting the use of anonymisation and pseudonymisation as privacy-enhancing techniques is understanding the role of such measures under data privacy rules. This article aims to contribute towards the achievement of such an objective by examining the role of anonymisation and pseudonymisation under the EU data privacy rules, particularly the Data Protection Directive, the ePrivacy Directive, Regulation 611/2013, the eIDAS Regulation, and the proposed General Data Protection Regulation. This article identifies three major roles of anonymisation and pseudonymisation under the current and *en route* rules. First, anonymisation and pseudonymisation can serve as a safe harbour from the entire application of data privacy rules provided they are used to irreversibly prevent identification, although achieving this goal seems increasingly challenging in the current state of technological advancement. Second, anonymisation and pseudonymisation can provide a safe harbour from certain data privacy obligations, such as the notification of personal data breaches, provided they are engineered appropriately and complemented by adequate organisational measures. Third, anonymisation and pseudonymisation can constitute mandated measures for compliance with data privacy obligations, such as the data security and purpose specification and limitation principles. All legal perspectives are drawn at EU level, although examples are given from member states when relevant.

---

[1] Samson Yoseph Esayas is a researcher at the Norwegian Research Center for Computers and Law (NRCCL), Department of Private Law, University of Oslo. His current research focuses on data privacy, legal aspects of cloud computing, and compliance risk management in using and developing information technology services. Other areas of research include Internet governance and policy, largely from the perspective of developing countries. This article was written as part of the Confidential and Compliant Clouds (Coco Cloud) and RASEN research projects. Both projects are funded by the European Commission (EC) via the Seventh Framework Programme, grant agreements no. 610853 and 316853, respectively. I express gratitude to my colleagues on the Coco Cloud and RASEN projects, and special gratitude to my colleagues Tobias Mahler and Kevin McGillivray at the NRCCL.

Keywords: Personal data; anonymisation and pseudonymisation; Encryption; Breach notification requirements; EU data privacy; Security of personal data

1. INTRODUCTION

Understanding the concept of 'personal data' is at the centre of data privacy discussions. This is so because the '*processing*' of '*personal data*' is the main criteria for the applicability of data privacy rules. The main objective of the European Data Protection Directive (hereafter the Directive) is the protection of 'the fundamental rights and freedoms of natural persons, in particular their right to privacy with respect to the "*processing of personal data*"' (Article 1(1)). It is evident from the article that at least two preconditions must be fulfilled for the Directive to apply: *data being processed* and this *data being personal*. According to Article 2(b), the term *'processing'* involves a wide range of activities including the 'collection, recording, organization, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, blocking, erasure or destruction.' Further, the Directive is primarily concerned with the processing of personal data *wholly or partly by automatic means*. The term *'partly'* implies that an automated operation that involves some manual use of personal data is within the realm of the Directive. In addition, the Directive is applicable to *non-automated processing which form part of a filing system or are intended to form part of a filing system*, such as specially structured paper file (Article 3(1)). Essentially, the Directive applies whenever personal data is processed, either automated or non-automated, barring certain exceptions.[2]

The second precondition is that the data being processed must be *personal*. The Directive defines personal data as '*any information relating to an identified or identifiable natural person*' (Article 2(a)). Identification involves 'describe[ing] a person in such a way that he or she is distinguishable from all other persons and recognisable as an individual' (European Union Agency for Fundamental Rights, 2013, p.40). The reference to identification includes both the term 'identified' and 'identifiable'. The former involves a situation where the identity of the person is already distinguished or manifestly clear. Such identification could happen directly from the information being processed, such as the full name of the person, or indirectly from information regarding the physical, physiological, mental, economic, cultural, or social identity of that particular individual (Article 2(a)). This implies that the person need not be identified at the level of his name; this indicates that identifying the individual at the level of, for example, his addresses, health, and financial data would also suffice. This is taken slightly further under the draft data protection Regulation in that the identification at behaviour level is also included. This seems to cover situations where individuals may be tracked and singled out at their behaviour level, for example, for making decisions about them (tailor ads according to their behavior) without their names or identities necessarily being known (Costa and Poullet, 2012, 255). However, for the Directive to apply, it is not required that the person be identified; it is sufficient that the person concerned be identifiable. 'Identifiability' implies that identification has not happened yet but is possible, for example, by combining the information being processed with other information. This implies that the mere possibility of associating certain information with a particular individual is sufficient. According to Recital 26 of the Directive, identifiability of a person should be assessed taking into account '*all the means likely reasonably* to be used either by the controller or by any other person to identify the said person'. The Recital further indicates that '*the principles of protection shall not apply to data rendered anonymous in such a way that the data subject is no longer identifiable*'.

[2] These include (1) in the course of an activity which falls outside the scope of Community law, for example, processing operations concerning public security, defence, State security, and the activities of the State in areas of criminal law; and (2) by a natural person in the course of a purely personal or household activity (Article 3(2)).

Two points within Recital 26 require particular mention, given their importance for the role of anonymisation and pseudonymisation. The first point relates to the term *'likely reasonably'*. The duo introduces two criteria for identifiability: the term 'likely' referring to 'probability' of identification and the term 'reasonably' referring to the 'difficulty' in identification—for example, in terms of costs, time required for identification, and available technology (Bygrave, 2002, p.44). The second point relates to a situation where certain information is rendered 'non-personal' through *anonymisation in such a manner that the data subject is no longer identifiable*. Data being rendered anonymous implies that information that ceases to be 'personal data' may be processed without any need of compliance to the requirements in the Directive.

The use of anonymisation can have many social and economic benefits. For example, anonymising data can be relevant in publishing data in rich and reusable formats for research and statistical purposes whilst privacy is being protected (ICO, 2012, p.9). However, in practice, it is often difficult to determine whether data has been sufficiently anonymised or is still personal data. This is partially because of the risk-based nature of anonymisation and its dependence on a variety of factors that are difficult to quantify. More particularly, this is related to the difficulty in foreseeing the available technology and information that could be used for re-identification. There are also challenges in articulating the harms on privacy (Oswald, 2014, p.260). Thus, the role of anonymisation and pseudonymisation in data privacy is considered one of the difficult areas of the law (ICO, 2012, p.9). The combination of these factors implies that individual as well as entities tend '*to downplay risks and simplify technical explanations in an attempt to reassure*' (Oswald, 2014, p.260). For example, there is a prevalent understanding in certain sections of the business and academic community that key-coded (e.g. encrypted) data may not be considered personal data so far as there are appropriate measures to exclude re-identification (i.e. strong encryption algorithm, strong encryption key, and secure key) (OPTIMIS, 2011, p.8; Hon, Millard, and Walden, 2011, p.216). Similarly, for data subjects, anonymisation gives a false sense of security to individuals, as they find it difficult to fully understand the potential risks of such measures and how such data is used (Ohm, 2009). This notwithstanding, most discussions regarding anonymisation are mainly at a technical level and focus on the risks of re-identification of different techniques (see Ohm, 2009 and Cavoukian, 2013). When legal perspectives are attached, they are only limited to whether a certain technique meets the conditions for providing a safe harbour from the entire application of data privacy rules without going into the other roles of such measures—a reference to the 'all or nothing' approach. Furthermore, the subject of rendering certain information 'non-personal' through different mechanisms in data privacy rules raises a number of legal issues. Thus, the main goal of this article is to examine and elaborate the different roles of anonymisation and related legal issues under the EU data privacy rules.

To this end, the following section examines whether the process of anonymising data constitutes the processing of personal data under the EU data privacy rules and its implications. This is followed by a discussion on the three major roles of anonymisation under the EU data privacy rules. Section 2 discusses the role of anonymisation as a safe harbour from data privacy rules in its entirety. Section 3 examines the role of anonymisation as a safe harbour from certain data privacy obligations, particularly notification of personal data breaches and further proposals under the ongoing data protection reform. This section also highlights the requirements for notifying personal data breaches under Regulation 611/2013, the eIDAS Regulation, and the proposed General Data Protection Regulation (GDPR)[3]. Section 4 analyses the role of

[3] Given that the adoption of EU legislation requires an agreement between the European Parliament and the European Council on the proposal placed by the Commission, at present there are three different drafts of the proposed GDPR that reflect the position of these organs: Commission, 'Proposal for a Regulation of the European Parliament and of the Council on the protection of individuals with regard to the processing of personal data and on the free movement of such data' [Com (2012) 11 final (**Commission draft**)]; (2) Committee on Civil Liberties, Justice, and Home Affairs, European Parliament, 'Report on the proposal for a regulation of the European Parliament and of the Council on the protection of individuals with regard to the processing of personal data and on the free movement of such data' (2013) [COM (2012)0011 – C7-0025/2012 –

anonymisation as a mandated compliance requirement under data privacy rules. The last section concludes the article with some observations.

1.1. ANONYMISATION AND PSEUDONYMISATION AS PROCESSING OF PERSONAL DATA?

Generally, the term anonymisation includes a number of techniques that aim at reducing the identifiability of individuals, and pseudonymisation can be considered as one technique of anonymisation. However, given the different legal significance attached to different anonymisation techniques under data privacy rules, in this article, a distinction is made between these two terms. Anonymisation is referred to as a process through which identifying information is manipulated (concealed or deleted) to make it difficult to identify data subjects (Ohm, 2009, p.1707). This definition mainly encompasses techniques used to produce aggregated information without any reference to information regarding a specific individual. Pseudonymisation involves replacing names or other direct identifiers with codes or numbers (Article 29 Working Party, 2007, p.18). The main purpose of such techniques is to enable the data to be associated with a particular individual without the individual being identified (ICO, 2012, p.51). In this context, given that the term 'processing' encompasses a wide range of activities on data under the Directive, a relevant question that comes into play is whether the process of anonymisation (pseudonymisation) in itself would constitute 'processing' under the Directive, thereby implying the need for compliance to perform anonymisation over certain data. This is because to generate anonymised or pseudonymised data, one has to apply a specific anonymisation or pseudonymisation technique to the personal data (Emam and Alvarez, 2014, p.8).

According to Article 29 Working Party (2014a, p.3), a group comprising national data protection authorities, '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.' This implies that the process of anonymising data by itself must comply with the test of compatibility with the original purpose. In other words, anonymising personal data for purposes not compatible with the original purpose constitutes a violation of data privacy rules, unless there are other legitimate grounds for the processing. For example, if personal data is collected to provide a certain service to the data subject, anonymising the data in order to use such data for advertising purposes would constitute a violation of data privacy rules unless, there are other legitimate grounds for processing (i.e. anonymising or marketing purposes), such as consent of the data subject.

The approach of Working Party in addressing the anonymisation process as compatible or incompatible with the original purpose represents a very narrow view of the role of anonymisation. As briefly noted above, the role of such measures is not limited to providing safe harbour from the entire application of data privacy rules. Rather, they can constitute mandated measures to comply with data privacy rules, such as purpose limitation and data security principles. For example, anonymisation could be used to comply with Article 6(1)(e) of the Directive, which requires that information should not be kept in identifiable form for a period longer than is necessary for the purposes for which the data were originally collected or for which they are further processed. Among others, Article 6(1)(e) would require the deletion of personal data, which can be achieved through anonymisation when the original legal basis is exhausted (Emam and Alvarez, 2014, p.9). Similarly, a data controller might need to anonymise personal data as part of additional security measures, even though such anonymisation does not

---

2012/0011(COD)] (link) (**LIBE draft**); (3) Council of the European Union, document 17831/13 on proposal for a regulation of the European Parliament and of the Council on the protection of individuals with regard to the processing of personal data and on the free movement of such data [(2013) (link) (**Council draft**)].

fulfil the conditions discussed in Section 2. In this sense, anonymisation is 'something different or something more than a compatible use' and might constitute mandated compliance measures (Emam and Alvarez, 2014, p.9). In addition, given that the underlying privacy interests are not threatened by such a process, the stance adopted by the Working Party would discourage the use of anonymisation and pseudonymisation as privacy-enhancing techniques (Hon, Millard, and Walden, 2011, p.214). This is the approach adopted by certain Data Protection Authorities (DPAs) like the UK Information Commissioner's Office (ICO). The ICO (2012, p.28) indicated that in the absence of any unwarranted damage or distress resulting from anonymisation, there is no need to justify the process of 'anonymisation' itself.

In this respect, a new notion in the Commission draft Regulation states that '…processing of data to the extent *strictly necessary* for the purposes of ensuring network and information security…constitutes a legitimate interest of the concerned data controller.'[4] Given that anonymisation and pseudonymisation could be considered as necessary for information security, this might give some leeway to argue that controllers might legalize the anonymisation or pseudonymisation under Article 6(f) of the draft Regulation as 'necessary for the legitimate interest of the controller', provided the controller's interests are not overridden by the interests or fundamental rights of the data subject. Similar arguments based on the 'legitimate interest of the controller' can be made under the current Directive Article 7(f) for anonymising personal data, except where such interests are overridden by the fundamental rights and freedoms of the data subject.

2. ANONYMISATION AND PSEUDONYMISATION AS A SAFE HARBOUR FROM DATA PRIVACY RULES

2.1. ANONYMISATION

According to the opinion of the Working Party, anonymisation for purposes other than the original purpose could still be considered as compatible with the original purpose as long as it fulfils certain conditions. For the anonymisation to be considered as compatible with the original purposes, the anonymisation process should produce 'reliably' anonymised information (Article 29 Working Party, 2014a, p.7). This implies that anonymising personal data for purposes that are incompatible with the original purpose would constitute a violation of the EU data privacy rules if it fails to fulfil the conditions of producing reliable anonymised data as laid down in the Working Party's document.

The document reiterates the sentiments under Recital 26 of the Directive and indicates that only when data is anonymised to the effect that it is no longer possible to associate to an individual by using '*all the means likely reasonably to be used*', either by the controller or a third party, it will not constitute personal data. Data being rendered reliably anonymous implies that such data can be processed without any need of compliance to legal requirements in the Directive.[5] However, such safe harbour would require irreversible anonymisation. In the opinion of the Working Party, the outcome of such kind of anonymisation should be, 'in the current state of technology, as permanent as erasure, i.e. making it *impossible* to process personal data' (Article 29 Working Party, 2014a, p.6). One consideration in assessing the notion of *impossibility* is the robustness of the anonymisation technique employed. In assessing the robustness of different techniques of anonymisation, the following questions should be taken into account: (1) Is it still possible to single out an individual, (2) is it still possible to link records relating to an individual, and (3) can

[4] GDPR, Commission draft, Recital 39.

[5] However, this does not deprive the data subject from the protection provided under other laws such as those protecting confidentiality of communications under Article 5(3) of the e-Privacy Directive (Article 29 Working Party, 2014a, p.11).

information regarding an individual be inferred? Using these three questions, the Working Party identified two major families of anonymisation (randomisation and generalisation) and discussed the strengths and weakness of these different techniques (Article 29 Working Party, 2014a, p.3). Further, the Working Party also underlined the dynamic nature of risks of identification as a result of developments in powerful data analysis techniques and, therefore, organizations are required to revisit the residual risks regularly. As such, even effectively anonymised data should be supported by the following follow-up measures: (1) Identify new risks and re-evaluate the residual risk(s) regularly, (2) assess whether the controls for identified risks suffice and adjust accordingly, and (3) monitor and control the risks (Article 29 Working Party, 2014a, p.24). Thus, if the result of such assessment entails '*an unacceptable risk of identification of data subjects*', the processing has to comply with data privacy rules (Article 29 Working Party, 2014a, p.10). However, the Working Party does not specify when the risk of identification is considered to be acceptable at a given time.

The lack of an acceptable risk threshold has been subject to criticism on the basis that the Working Party follows an 'absolute definition of acceptable risk in the form of zero risk' (Emam and Alvarez, 2014, p.9). First, the Directive itself does not require a zero risk approach. As noted above, the use of the term 'likely' represents the 'probability' of identification, whereas the term 'reasonably' represents the 'difficulty' in identification (Bygrave, 2002, p.44). Both are terminologies common in the field of risk management reasoning, thereby indicating the potential for a certain level of acceptable risk. Such criticism of the zero risk approach complements the claim that technologists and regulators often misunderstand the term 'anonymisation'. As Ohm (2009) puts it, '*a word that should mean, "try to achieve anonymity" is too often understood*,' just as the case with the Working Party, '*to mean achieve anonymity*'. Such an approach will significantly affect the widespread use of anonymisation measures. Some commentators suggest that '*only where risk of identification is sufficiently realistic (for example, 'more likely than not'), should information be considered 'personal data'* (Hon, Millard, and Walden 2011, p.226). In other words, where identification risk is remote or highly theoretical, given the time, expense, technology, and labour required to associate the data to a particular individual, then the data should not be considered personal. [6] Such an approach accommodates the existence low risk or very small risk in anonymised data, which might be considered acceptable. Some member states have adopted such a stance. For example, the UK ICO has held that anonymised data is not required to be completely risk free, rather it must be able to mitigate the risk of identification until it is remote (ICO, 2012, p.13).

Second, if the acceptable risk threshold is zero for any potential recipient of the data, there is no existing technique that can achieve the required degree of anonymisation. From a technical point of view, achieving full anonymisation of the data that would not allow re-identification is considered very difficult (ENISA, 2012, p.44). This implies that the processing, including the process of anonymising the data, has to be justified under one of the legitimate grounds listed in Article 7 of the Directive. Adopting such an approach represents higher risks, as it encourages the processing of data in identifiable form.

Another factor in the assessment is to examine if there is any kind of data either in the hands of the *'controller or any other person'* that could be used to identify the individual. For example, if a data controller keeps the original (identifiable) data and shares part of this dataset by removing or masking the identifiable data to another party, the resulting dataset is still personal data (Article 29 Working Party, 2014a, p.9). This is because there is still data in the hands of the controller that could be used to associate to the individual. However, this overlooks the possibility where the original data might not have any assistance in associating the anonymised

[6] Council draft Regulation, Recital 23 states that 'to ascertain whether means are reasonably likely to be used to identify the individual, account should be taken of all objective factors, such as the costs of and the amount of time required for identification, taking into consideration both available technology at the time of the processing and technological development'.

data to certain individual. Ultimately, the existence of any data—either in the hands of the controller or any other party—which could be used to associate the anonymised data to a certain individual, implies that the data is not considered to be truly anonymised and has to comply with data protection rules. The challenge with this approach is that it establishes a universal basis for assessing whether there is any 'other information' that could be combined with the anonymised data to identify the individual. According to Ohm (2009, p.1752), 'there is always some piece of information […] that could be combined with anonymised data to reveal private information about an individual'. In addition, 'determining what "other information" is available, who it is available to and whether it is likely to be used in a re-identification process' would be an extremely formidable task (ICO, 2012, p.18). A more practical, but still difficult, approach would have been to assess whether it is likely that the receiver of the anonymised data would come into the possession of such information, taking into account whether the information is available on the Internet or available only to certain organizations or public bodies. For example, the UK ICO indicates that the disclosure of anonymised data is not a disclosure of personal data—even where the data controller holds the key to enable re-identification, which is also confirmed by a decision from the UK high court (ICO, 2012, p.13).[7] Such an approach would allow recipients to process anonymised data without the need to comply with data protection rules, despite the controller retaining an identifiable form of the data.

Generally, the recent opinion of the Working Party seems to indicate that true anonymisation is unattainable in a world of 'open' datasets, thereby indicating that the current state of technology and given the increase in computational power and tools available, *'likely reasonably'* is easily attainable (Article 29 Working Party, 2014a, p.4). It is true that re-identification has become easy as a result of the technological advancement and the ubiquity of information on the Internet, but the alternative should not be a boundless and overbroad application of the Directive. This approach dislodges the appropriate balance between information flow and privacy, which hinges on restricting the application of data privacy rules to 'personal data' (Ohm, 2009, p.1763). A more practical approach would dictate establishing different layers of risks of harm that takes account the sensitivity of the data and the context of its usage. For example, the following four layers might be envisaged, where (1) the anonymisation is employed to sensitive data and would be publicly available, (2) the anonymisation is applied to sensitive data and would be available with limited access, (3) the anonymisation is applied to non-sensitive data and would be publicly available, and (4) the anonymisation is applied to non-sensitive data and would be available with limited access. At least the fourth layer could be subject to less strict requirements of anonymisation.

Overall, anonymised data that irreversibly prevents identification can be processed without the need to adhere to the legal requirements under the Directive. However, it must be noted that the Directive will still be applicable if the anonymisation techniques are engineered inappropriately; thus, any doubt in this regard should be interpreted as involving the processing of personal data.

[7] However, one also needs to mention a decision from the House of Lords contradicting such a stance - see, House Of Lords, Opinions of the Lords of Appeal for Judgment in the Cause Common Services Agency (Appellants) v Scottish Information Commissioner (Respondent) (Scotland), SESSION 2007–08 [2008] UKHL 47 on appeal from: [2006] SCOTCS CSIH 58. In this case, the court turned down an appeal from the ICO decision to share personal information in anonymised form (barnardisation). The House of Lords qualified this aspect by adding that such disclosure is in line with the Data Protection Act only if the third party cannot identify the data subject from the anonymised data even if having access to the original data. According to the decision, 'if the "other information" [for example the original data] is incapable of adding anything and "those data" [the anonymised data] by themselves cannot lead to identification, the definition [of personal data] will not be satisfied. The "other information" will have no part to play in the identification. The same result would seem to follow if "those data" have been put into a form which the individual or individuals to whom they relate cannot be identified at all, even with the assistance of the other information from which they were derived' (para24). The decision does recognize that the mere fact that the original data is kept identifiable in the hands of the controller does not make such anonymised data personal (para 27).

Similarly, member states might extend the scope of national legislation to areas not included within the scope of the Directive. For example, in France, the 'reasonableness' test does not exist, thereby implying that data remains personal data even if it is extremely difficult to re-identify the data subject and unlikely that re-identification will take place (Article 29 Working Party, 2014a, p.6). Furthermore, if a person (natural or legal) is able to associate certain data, for example by accidental matching,[8] from truly anonymised data, that person has to comply with data privacy rules. Although, at the EU level, the responsibility of the controller that released the anonymised data that is subject to re-identification is unclear, the ICO (2012, p.41) recommends that this be treated as a breach of security and that the concerned individuals be notified of the breach. Simultaneously, it is important to acknowledge the difficulties associated with such compliance, because the data might already be exported outside the EEA.

2.2. PSEUDONYMISATION[9]

Another method to reduce the likelihood of identifiability of individuals is pseudonymisation. Examples of pseudonymisation techniques include encryption and hash function. Similar arguments were made with regard to pseudonymised data in that so long as it is effective, it should not be considered 'personal' (OPTIMIS, 2011, p.8). However, the Working Party indicates that equating pseudonymised data to anonymised data is considered as one of the misconceptions among many controllers. This is because pseudonymised data continues 'to allow an individual data subject to be singled out and linkable across different datasets' and is subject to data protection rules (Article 29 Working Party, 2014a, p.10). On the one hand, individuals continue to be identified by a unique attribute that is the result of the pseudonymisation (the pseudonymised attribute) (Article 29 Working Party, 2014a, p.21). On the other, it is still possible to single-out individuals from pseudonymised data, either because someone is holding the key that could be used to re-identify the individual or the key can be bypassed by brute force attacks, or as a result of a data breach (Article 29 Working Party, 2014a, p.29). This is in line with the Working Party's suggestion that there should not be any kind of data either in the hands of the *'controller or any other person'* that could be used to identify the individual. Although such a stance seems to effectively exclude the use of pseudonymisation as a measure that provides safe harbour from data privacy rules, a combination with some anonymisation techniques such as removing and generalising attributes or deleting the original data could achieve the required level for safe harbour (Article 29 Working Party, 2014a, p.21). Some DPAs like the UK ICO have adopted a different approach. In its guide on anonymisation, the ICO indicated that although pseudonymised data may create a higher risk of re-identification, it does not mean that effective anonymisation through pseudonymisation is impossible (ICO, 2012, p.21).

2.2.1. ENCRYPTION

Encryption can serve as a pseudonymisation technique when it is used to conceal directly identifiable information. Encryption is the process of changing a plain text in to unintelligible code; in contrast cryptography, often used interchangeably with encryption, is the related science dealing with the technicalities of creating encrypted information (Perkins, 2005, p.1628). The use of encryption has been tipped as a privacy-enhancing measure, particularly in case of cloud computing services (Kuner, 1996, p.186). As noted above, there are also arguments that as far as the encryption is effective—that is, there is a strong encryption algorithm, strong encryption key, and the key is kept secure—the data may not be considered personal in the hands of a third party

[8] In an earlier opinion, the Working Party has held that the possibility of re-identification through 'unforeseeable circumstances', such as 'accidental matching' does not make an effectively anonymised data personal data (Article 29 Working party, 2007, p.20).

[9] Given that pseudonymisation could be achieved through encryption, the discussions on encryption will also be relevant.

that is not in possession of the decryption key and, thus, the Directive is inapplicable (OPTIMIS, 2011, p.8). The underlying rationale behind such suggestions is that the application of data privacy rules should be based on access to '*intelligible data*' (Hon et al. 2014, p.10). However, according to the opinion of the Working Party, given that the security of encryption or the hash function is affected by many technical and organizational measures, the focus on the robustness of the encryption, as such, is misleading. (Article 29 Working Party, 2014a, p.29). In particular, this opinion identifies essential differences in the use of encryption and anonymisation.

One major difference pertains to the goal of the techniques. The goal of anonymisation is primarily to eliminate linking attributes and avoid identification of individuals (Article 29 Working Party, 2014a, p.29). With regard to encryption and key-coding, the goal is not making a data subject unidentifiable, since, in the hands of the controller at least, the original data are still available or deducible (Article 29 Working Party, 2014a, p.29). In other words, encryption does not eliminate the *identifiability* aspect of the information, that is, the relationship of the information to the individual and the possibility of identifying the person from the encrypted information theoretically exists. This is particularly the case in an era where cryptographic attacks are being continuously improved due to increased computing power and the availability of cryptographic key cracking cloud services (Wayne and Grance, 2011, p.12). Therefore, as long as the key or the original data are available, even in the hands of a trusted third party, the possibility of identifying a data subject is intact (Article 29 Working Party, 2014a, p.29). This implies that the use of two-way cryptography algorithms (rendering personal data unintelligible with the possibility of backtracking the individual under predefined circumstances, for example, by entering the correct key/password) is still subject to data privacy rules. In an earlier Opinion, the Working Party suggested that data containing one-way cryptography identifiers (irreversibly encrypted data) would not be 'personal data' provided that the cryptography is effective (Article 29 Working Party, 2007, p.18; Hon, Millard, and Walden 2011, p.217). In other words, one-way encrypted data or keyed-hash function with deletion of the key can be processed independent of the Directive. Although the latest Opinion does not provide a clear indication to this effect, it acknowledges that when deterministic encryption or keyed-hash function with deletion of the key is used, 'it becomes computationally hard for an attacker to decrypt or replay the function, as it would imply testing every possible key' (Article 29 Working Party, 2014a, p.21). It is the opinion of this author that such data should benefit from safe harbour. However, the safest approach is that the use of encryption or the hash function does not provide safe harbour for the application of data privacy rules under the EU data privacy framework, unless complemented by anonymisation techniques. This is not to mention that such measures can still be mandated as a compliance requirement, as discussed in Section 4.1.

In summary, the recent opinion of the Working Party does not seem encouraging for businesses to use anonymisation and pseudonymisation in processing personal data. Furthermore, the Opinion does not provide any guidance for data controllers or data processors to follow to anonymise their data (Emam and Alvarez, 2014, p.3). One important contribution could be that the Opinion has identified the different risks associated with different techniques. As the Working Party has indicated, different combinations could be used to reach the required level of 'anonymisation', in which case the Directive does not apply. In certain cases, it is advised to apply a double anonymisation, that is, the application of a second anonymisation on the result of a first anonymisation. However, it might be useful to come up with different layers of risks that take account the kind of data, the anonymisation or pseudonymisation technique employed, the context of their use, and then lay down different conditions accordingly. A further consideration could be to mitigate some obligations with respect to the use of a specific anonymisation technique if certain risks no longer exist. As discussed below, recent developments in the data privacy discourse seem to heed such suggestion by introducing a safe harbour from certain obligations when such measures are implemented.

3. ANONYMISATION AND PSEUDONYMISATION AS A SAFE HARBOUR FROM CERTAIN DATA PRIVACY OBLIGATIONS

As noted above, the current data privacy framework has been criticized for its 'all or nothing approach' where if there is slight possibility to associate certain information to an individual, the data privacy rules apply entirely regardless of whether the data is anonymised or pseudonymised. However, in the past few years, there have been some developments that exempt controllers from certain obligation if data is anonymised or pseudonymised and fulfil certain conditions. Currently, such initiatives are essentially focused on personal data breach notification requirements. At the EU level, the amendment to the 2002 ePrivacy Directive, through Directive 2009/136/EC, introduces mandatory personal data breach notification obligation under its Article 4(3). However, the inconsistent implementation of the breach notification requirements within the ePrivacy Directive is believed to create 'significant legal uncertainty, complexity and considerable administrative costs for providers operating cross-border' (Recital 4, Regulation 611/2013). Therefore, Regulation 611/2013 is adopted to harmonize the notification of personal data breaches by public electronic communications service providers, which include both traditional telecom providers such as telephony companies and Internet Service Providers (ISPs).[10] Similarly, the Regulation on electronic identification and trust services (eIDAS), which will replace the eSignature Directive 1999/93/EC, introduces personal data breach notification requirement for trust service providers, which could range from telecom service providers to banks and other financial institutions to universities (Regulation No 910/2014). Furthermore, the proposed General Data Protection Regulation (GDPR) contains personal data notification obligations for controllers and, to a certain extent, processors. In the following paragraphs, the personal breach notification requirements under these laws are briefly highlighted, followed by the role of anonymisation and pseudonymisation in providing safe harbour from such obligations.

3.1. NOTIFICATION OF PERSONAL DATA BREACHES UNDER THE EU LEGAL FRAMEWORK

It is not the aim of this article to examine the details of personal data breaches under the EU legal framework. Such requirements are only discussed to the extent they are relevant to the discussion on the role of anonymisation and pseudonymisation. In this regard, it is important to note that the breach notification requirements under Regulation 611/2013 are essentially similar to the requirements under the proposed GDPR and the eIDAS Regulation. This is not incidental. It originates from the legislators intent of harmonizing notification requirements regarding personal data across sectors.[11] Thus, unless mentioned otherwise, the discussions regarding Regulation 611/2013 on notification of personal data breaches are generally relevant to the proposed GDPR and the eIDAS Regulation.

Under Regulation 611/2013, providers of electronic communication services are required to notify personal data breaches to the relevant authorities within 24 hours after detection of the breach (Article 2(2)).[12] Moreover, Article 3(1) of the Regulation provides that *'when the personal*

[10] In the European legal framework, a Directive has to be transposed to national law for its application, whereas a Regulation becomes binding on Member States without the need to transpose it into national law.

[11] See Recital 19 of 611/2013 referring that the Regulation is fully consistent with the proposed measure under the draft Regulation.

[12] Personal data breach is defined as '…*a breach of security leading to the accidental or unlawful destruction, loss, alteration, unauthorized disclosure of, or access to, personal data transmitted, stored or otherwise processed in connection with the provision of a publicly available electronic communications service in the Union*' (Recital 2 of Regulation 611/2013).

data breach is likely to adversely affect the personal data or privacy of a subscriber or individual, the provider shall, in addition to the notification referred to in Article 2, also notify the subscriber or individual of the breach.'

An important aspect of Regulation 611/2013 is that it covers breaches affecting not only natural persons but also legal persons. Thus, it is important to distinguish between the 'subscriber' and 'individual' user because, the subscriber, which can be either a legal person or natural person, may not always be the same person as the user. For example, some parents may subscribe to a service that locates the mobile phone of their children. In such instances, the parents are the subscribers and their children are the individual users. Further, according to Article 3(1) of the Regulation, both the subscriber and the individual user are to be notified when a personal data breach is likely to affect the privacy of the subscriber or the individual. Providers are required to notify the affected individual or subscribers when the breach *'is likely to adversely affect'* their personal data or privacy rights. Article 3(2) of the Regulation lists three elements as essential in determining the adverse effect of a breach. These are '(a) the nature and content of data concerned, (b) the likely consequences of a personal data breach for the subscriber or individual concerned, and (c) the circumstances of the personal data breach.'

Breaches affecting certain categories of personal data are considered to fulfil such a requirement. Examples are breaches affecting financial information, like credit card data, and special categories of data,[13] 'e-mail data, location data, internet log files, web browsing histories and itemized [*sic*] call lists' (Regulation 611/2013, Recital 61). This is because such breaches might lead to 'identity theft [,] fraud, physical harm,… [significant] humiliation or damage to reputation' (Regulation 611/2013, Art 3(2(b))). This implies that the assessment is not limited only to 'breaches that result in economic loss, but also breaches that may cause immaterial damages, such as any moral and reputational damages' (EUR. CONSUMER ORG, 2011, p.4). The Working Party also underlines the need to consider secondary effects of the breach such as the time spent in attempts to rectify the breach and the extent of distress suffered (Article 29 Working Party, 2014b, p.13). In addition, the reference to the term '*likely*' implies that the mere likelihood that the breach will adversely affect the individual is sufficient, meaning that an actual adverse effect is not necessary.

The notification to the subscriber or individual shall be made without *undue delay* after the detection of the personal data breach (Regulation 611/2013, Art 3(3)). Furthermore, the notification shall not be dependent on the notification to national authorities (Regulation 611/2013, Art 3(3)). This implies, for example, that an organization should not attempt to prioritize notification to the authorities over the subscribers or individuals. Given that the aim of such notification is to avoid or mitigate the consequences of the breach, the notification should be given immediately or in such time as to enable the subscriber or individual to mitigate the adverse effects of the breach. In light of such rationale, Barcelo and Traung (2010, p.96) argue that the term *without undue delay* involves a shorter interval than notification to the authority.

Similar provisions are to be found in the eIDAS Regulation. Article 19(2) of the Regulation requires trust service providers to notify personal data breaches to the relevant authorities *without undue delay but in any event within 24 hours after having become aware of it.* Moreover, paragraph 2 of Article 19(2) states that '*Where the breach of security or loss of integrity is likely to adversely affect a natural or legal person to whom the trusted service has been provided, the trust service provider shall also notify the natural or legal person of the breach of security or loss of integrity without undue delay.'* This is essentially similar to the requirements under Article 3(1) of Regulation 611/2013.

[13] Defined as personal data '*revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, trade-union membership, and the processing of data concerning health or sex life and data relating to offences, criminal convictions or security measures, or to administrative sanctions or judgments in civil cases*' (Data Protection Directive, Article 8)

Furthermore, the different drafts—that is, the initial Commission draft, LIBE draft, and the Council draft of the proposed GDPR—contain provisions for notification of personal data breaches both for the DPAs and individual data subjects. There are slight differences among the three drafts and the main differences with respect to the breach notification will be highlighted when relevant. All the three drafts of the Regulation require notification to the data subjects to happen '*without undue delay*' after the controller becomes aware of the breach.

3.2. ANONYMISATION OR PSEUDONYMISATION AS A SAFE HARBOUR FROM BREACH NOTIFICATION OBLIGATIONS

According to Article 4(1) of the Regulation 611/2013 '*notification of a personal data breach to a subscriber or individual concerned shall not be required if the provider has demonstrated to the satisfaction of the competent national authority that it has implemented appropriate technological protection measures, and that those measures were applied to the data concerned by the security breach.*' More particularly, the Regulation indicates that individuals or subscribers do not need to be notified when providers manage to demonstrate that the data affected by the breach was rendered *unintelligible* (Regulation 611/2013, Article 4(1)). According to Article 4(2) of the Regulation, data is considered to be *unintelligible* where (a) it has been securely encrypted with a standardized algorithm or replaced by its hashed value, calculated with a standardized cryptographic keyed hash function, (b) the key used to decrypt or to hash the data has not been compromised in any security breach, and (c) it has been demonstrated that the key used to decrypt or hash the data cannot be ascertained by available technological means by any person not authorized to access the key.

Data can be rendered unintelligible in a number of techniques, but the reference under Article 4(1) to encryption or hashing, in particular, indicates that the provision focuses on pseudonymisation techniques that enable the re-identification of the individual subsequently. However, this does not imply that the use of anonymisation would not exempt the controller from notifying data subjects. As noted above, this flows from the general principle that if data is rendered 'non-personal' through anonymisation, data protection rules do not apply. In fact, for the exemptions under Article 4(1) to apply, it is not necessary that the outcome of such anonymisation should be as permanent as erasure, thereby making it *impossible* to process personal data (Article 29 Working Party, 2014a, p.6). For example, as noted above, the Working Party has held that if a data controller keeps the original (identifiable) data and hands over part of this dataset by removing or masking the identifiable data to another party, the resulting dataset is still personal data and has to comply with data protection rules (Article 29 Working Party, 2014a, p.9). However, the controller might still benefit from safe harbour for notifying breaches, even if it keeps the original (identifiable) data, unless the security of the latter is compromised.

A number of reasons exist for such a safe harbour. First, if certain data was made initially unintelligible, the residual privacy risks of the breach are considered to be minimal—not likely *'to adversely affect'* the personal data or privacy rights of individuals (Article 29 Working Party, 2014b, p.1). Second, such exemption aims to reduce the regulatory compliance burden on organisations and the negative impact of over-notification, referred to as notification fatigue, for users. Schwartz and Janger (1999, p.916) argue that if consumers are flooded by frequent caution messages with merely putative threats, it is likely that they will fail to act when important warnings finally arrive. Furthermore, it could serve as an incentive for the wider adoption of technological measures. A survey shows that such safe harbour from notifying breaches have increased the use of encryption (Ponemon Institute 2009). Meanwhile, the exception related to technological protection measures under EU rules is not an automatic safe harbour and must be approved by the competent regulatory authority. This is derived from the following statement: '*the provider has demonstrated to the satisfaction of the competent national authority*' in Article 4(1). Three approaches to providing safe harbour to such obligation could be identified: an *exemption*,

a *rebuttable presumption*, and *factor-based analysis* (Burdon, Reid, and Low, 2010, p.529).[14] The EU legislator seems to prefer the factor-based analysis, where the implementation of such measures is one factor in demonstrating to the regulatory authorities that the rights of the data subject are not affected.

In addition, organizations are still required to notify the relevant national regulatory authorities regardless of such measures, thereby implying that the technological measures under Article 4(1) serve as a safe harbour only from the notification to individuals and not from regulatory authorities. In some cases, the organizations might even be required to notify the breach to individuals even if the data is sufficiently encrypted. This is because in the absence of adequate backups, a loss or alteration of encrypted data can still negatively affect data subjects (Article 29 Working Party, 2014b, p.1). This is important because encryptions cannot prevent loss of data. Thus, for the purposes of the exemptions, it is important to make a distinction among the three kinds of personal data breaches: '"*availability breach*" —which refers to the "accidental or unlawful destruction of data[;]" "*integrity breach*"—which refers to "alteration of personal data," and "*confidentiality breach*"—which relates to "unauthorized disclosure of, or access to, personal data" (Article 29 Working Party, 2014b, p.2). This implies that the safe harbour regarding unintelligible data does not prevent '*availability breach*' and might not exempt the entity from notifying the individual. Although one could argue that breaches affecting *availability* might not, in the strict legal sense, affect the *privacy rights* of individuals, Article 3(1) of the Regulation refers to '*personal data* or privacy of a subscriber or individual.' This implies that breaches affecting availability might still adversely affect the personal data of subscribers or individuals.

There are similar provisions for safe harbour to the notification of data subjects under the proposed GDPR. However, unlike Regulation 611/2013 that provides a detailed description regarding the exemptions for notifying individual data subjects, the initial Commission draft and the LIBE draft adopt a very general approach without any reference to specific technological measures in rendering data unintelligible. This might be because of the issues of technology neutrality in making specific reference to encryption or hashing within Regulation 611/2013. However, the Council draft refers to encryption or pseudonymisation as mechanisms that can be employed to render personal data unintelligible. Given that encryption can be considered as one technique of pseudonymisation, the reference to both terms seems to be redundant. Moreover, the Council draft does not employ the term '*…has demonstrated to the satisfaction of the competent national authority*', thereby implying that the Council prefers an automatic safe harbour (an exemption as opposed to the factor-based analysis under Regulation 611/2013, the Commission draft, and the LIBE draft). However, for this to apply, the encryption or pseudonymisation has to fulfil two conditions that can be derived from the definition of 'pseudonymous data' under Article 4(2a). These are (1) the data cannot be attributed to a specific data subject without the use of additional information, and (2) such additional information is kept separately and subject to technical and organisational measures to ensure non-attribution. However, the reference to encryption specifically might create a misunderstanding that the implementation of such measures can relieve an organization from notifying and overlooks the possibility of breaches affecting the availability of data.

Furthermore, one important deviation of the Council draft from the other drafts and Regulation 611/2013 is that the safe harbour also applies to the notification of the regulatory authorities. This implies that when notification to individuals is not required because of technological measures including encryption or pseudonymisation (Article 31(1a)), controllers do not need to notify the DPAs. In addition, the Council draft requires notification to the DPAs only when the breach is

[14] *Exemptions* provide a general safe harbour for notification if personal data has been acquired in unintelligible form. *Rebuttable presumptions* create a presumption that no risk exists if unintelligible data is acquired, which can be rebutted if evidence is found to the contrary. In *factor-based analysis*, unintelligibility is merely a factor accounted for in determining whether harm will reasonably result from the breach (Burdon, Reid, and Low, 2010, pp.528-530).

likely to affect the rights and freedom of data subjects. This implies that, apart from the safe harbour through encryption or pseudonymisation, if the controller is able to demonstrate to the regulatory authorities that the breach does not affect the rights of the data subjects, it will benefit from a safe harbour. This could be achieved through measures that do not necessarily fall under Article 4(2a). Some scholars argue that the 'notification of personal data breaches to the DPAs may not be necessary when the rights and freedoms of data subjects are not likely to be affected' (Hon et al., 2014). However, the rationale behind the notification to the DPAs goes beyond the protection of rights and freedoms of individuals. On the one hand, such notification enables the authorities to identify patterns of breaches and learn where policy interventions and cooperation might be required. On the other hand, the notification to the authorities enables them to assess whether notification to individuals should take place. This ensures that the assessment of whether the rights of the data subjects are likely to be affected does not solely rest in the hands of the entities. Moreover, such requirement would introduce unjustified discrimination among entities that process personal data in the electronic communications sector that have to notify *any breaches* to the authorities under Regulation 611/2013 and others that have to notify only data breaches that are likely to 'severely' affect the rights and freedom of data subjects. Thus, the notification to DPAs should not be dependent on the 'severity' of the breach to the rights of data subjects.

Overall the approaches for the safe harbour from notifying breaches as a result of the technological measures under the Commission draft and LIBE draft are essentially similar to Regulation 611/2013. However, the Council draft deviates in terms of the following aspects: (1) employs an automatic safe harbour from notifying individuals when personal data is encrypted or pseudonymised, (2) gives safe harbour from notifying the DPAs when personal data is encrypted or pseudonymised, (3) requires notification to the DPAs only where the breach is likely to affect the rights and freedom of data subjects.

Quite surprisingly, the eIDAS Regulation does not provide similar safe harbour for notifying breaches to the individuals, as in Regulation 611/2013 and the proposed GDPR. However, this might not necessarily imply that trust providers are not exempted when they have implemented appropriate technological measures. This is because the concept of *'adversely affect'* under Article 19(2) of the Regulation apparently encompasses the exemption of technological measures. As noted above, the main rationale behind such exception is that if certain data was made initially unintelligible, the residual privacy risks of the breach are considered to be negligible (Article 29 Working Party, 2014b, p.1). This implies that the requirement of adverse effect would not be fulfilled and, thus, notification is not required. However, this has to be approved by the competent regulatory authorities.

Apart from the data breach notification exemptions, there is a rare case where the European Commission has held, in its Frequently Asked Questions (FAQs), that transfer of key-coded data outside the EU without transferring or revealing the key does not involve transfer of personal data (see in Hon, Millard, and Walden, 2011, p.216).[15] This implies that the use of such measures might exempt controllers from complying with the requirements under Articles 25 and 26 of the Directive. Although, the Commission might have adopted a different stance following the Snowden revelation, such an approach mitigates the application of certain provisions to pseudonymised data. Similarly, Austria's implementation of the Directive recognizes information which is 'indirectly personal data' if the controller, processor, or recipient cannot identify individuals using legally permissible means (Hon, Millard, and Walden, 2011, p.216). This implies that such data can, for example, be exported without regulatory approval (Hon, Millard, and Walden, 2011, p.216). Furthermore, the ICO (2012, p.12) indicates that the disclosure of anonymised data is subject to fewer legal restrictions. More particularly, if data is anonymised, the Data Protection Act's purpose-limitation rules do not apply to it (ICO, 2012, p.12). This

[15] See also Commission Decision 2000/520/EC [2000] OJ L 215/7.

implies that the use of anonymised data for incompatible purposes would not constitute a violation of data privacy rules. Such an approach involves a move from the 'all or nothing approach' regarding personal data and introduces room for 'more or less personal' data and accordingly 'more or less protection' (Robinson et al, 2009, p.26-27). This would encourage the wider use of such techniques.

3.3. OTHER PROPOSED SAFE HARBOURS

The different drafts for the proposed GDPR contain additional safe harbours from other obligations. One such obligation that could be exempt through the use of pseudonymisation is the prohibition against measures, which are based on profiling by means of automated processing. In this regard, Article 20 of the Commission draft prohibits the use of automatic processing intended to evaluate, analyse, or predict a natural person's performance at work, economic situation, location, health etc. and *which can significantly affect this natural person* except when carried out in the course of entering or fulfilment of a contract, or when the data subject has given his consent. However, the LIBE draft indicates that profiling based solely on the processing of *pseudonymous data should be presumed not to significantly affect the interests, rights, or freedoms of the data subject (Recital 58a)*.[16] This creates a rebuttable presumption that implies that unless proven otherwise, the controller may use such measures if the data is *pseudonymised and does not* permit the controller to attribute pseudonymous data to a specific data subject. Furthermore, one of the recommendations of the Committee on the Internal Market and Consumer Protection to the LIBE draft contains an exemption from the obligation to rectify inaccurate or incomplete personal data under Article 16 of the proposed Regulation related to the processing of pseudonymous data.

4. ANONYMISATION AND PSEUDONYMISATION AS AN INTEGRAL PART OF DATA PRIVACY COMPLIANCE

In this section, the role of anonymisation and pseudonymisation as mandated compliance measures with data privacy rules are examined. More particularly, the role of anonymisation and pseudonymisation as measures to comply with the data security obligation, purpose specification and limitation principle is provided.

4.1. ANONYMISATION AND PSEUDONYMISATION FOR FULFILLING DATA SECURITY OBLIGATIONS

Data security is one of the fundamental principles of data privacy under the EU legal framework. Article 17(1) of the Directive requires the controller to take '*appropriate technical and organisational measures to protect personal data against accidental or unlawful destruction or accidental loss, alteration, unauthorised disclosure or access, in particular where the processing involves the transmission of data over a network and against all other unlawful forms of processing*'. According to Recital 46 of the Directive, the 'appropriate technical and organizational measures be taken, both at the time of the design of the processing system and at the time of the processing itself.' Article 17(2) of the Directive also requires the controller to choose a processor that has the appropriate technical and organizational measures in place and that ensures its compliance. Furthermore, Article 17(3) the Directive requires data processors to implement appropriate security measures as defined by the law of the Member State in which the processor is established. Similarly, Article 4(1) of the ePrivacy Directive contains provision for protection of personal data stored or transmitted against

[16] Article 4(2a) defines 'pseudonymous data' as '*personal data that cannot be attributed to a specific data subject without the use of additional information, as long as such additional information is kept separately and subject to technical and organisational measures to ensure non-attribution*.'

accidental or unlawful destruction, accidental loss or alteration, and unauthorised or unlawful storage, processing, access or disclosure. Therefore, taking appropriate security measures is an integral part of data privacy compliance.

More generally, Article 17(1), in particular, refers to two kinds of data security measures, that is, technical and organization.[17] The Directive does not stipulate which specific technical and organisational measures have to be taken by the controller or processor. In this section, organizational measures are only discussed so far as they are relevant for elaborating the technological measures. Overall, the reference to security includes '*availability*', which includes measures against the accidental or unlawful destruction or loss of data; '*integrity*', which includes measures against alteration of personal data; and '*confidentiality*', which includes measures against unauthorized disclosure of, or access to, personal data. Technical measures such as anonymisation and pseudonymisation could help to comply with the above obligation, particularly the *confidentiality* and the *integrity* aspects. The specific requirements are left to the member states. Studies show that there is a considerable disparity in the security requirements of member states (Hon, Hörnle, & Millard, 2012, p.151). For example,

> In the UK the requirement is simply to take 'appropriate technical and organizational measures', whereas Italy has set out in detail what those security measures should be, e.g. for reuse of storage media, access to sensitive passwords, etc.; Denmark requires internet transmissions of personal data to be encrypted, and Austria, as well as defining detailed minimum security measures, requires documentary records of those measures (Hon, Hörnle, & Millard, 2012, p.151).

Thus, some member states like Denmark explicitly require the use of encryption. Similarly, the Spanish Royal Decree identifies three levels of security for processing of personal data: *basic*, *medium*, and *high* security levels (Royal Decree, Article 80). Further, in a high-level security environment, which is applicable where sensitive data as defined under Article 8 of the Directive is processed, the transfer of personal data through public or wireless electronic communications networks shall guarantee that the information shall not be intelligible or manipulated by third parties (Royal Decree, Article 104). One way of fulfilling such a requirement is by encrypting communications through such networks. Such a requirement also applies to backup copies. Furthermore, according to Article 93, the data controller has to guarantee the correct identification and authentication of users. When the authentication mechanism is based on the use of passwords, passwords shall be stored in an unintelligible way, for example, in encrypted form.

In Italy, the Privacy Code requires the implementation of encryption techniques for specific processing operations with respect to data disclosing health and sex life. The integrity of the backup files should also be ensured, possibly through anonymisation of the data. Referring to the information stored in back-up systems, the Italian Data Protection Authority (2014) specified that it 'must be protected against unauthorized access by means of suitable encryption techniques or, where necessary, by *anonymising* the data in question'. In Germany, the obligation for disclosure control aims to ensure that personal data cannot be read, copied, altered, or removed without authorization during electronic transfer or transport or while being recorded onto data storage media (Section 9 and Annex of the Federal Data Protection Act (Bundesdatenschutzgesetz)). This requirement might be fulfilled by encrypting data during transmission or during recording onto data storage media.

Moreover, the obligation regarding Data protection by design and by default under the proposed Regulation has similar requirements for the use of technical measures such as anonymisation and pseudonymisation. More particularly, Article 23 of the Council draft stipulates that 'having

[17] Although it is not easy to distinguish between the two, technical measures include the use of encryption, anonymisation, secure connections, and firewalls; on the other hand, organizational measures include access policies and identity management for the IT system processing the data (OPTIMIS, 2011, p.19).

regard to available technology, the cost of implementation and risks presented, the controller shall implement technical and organisational measures appropriate to the processing activity being carried on and its objectives, *including the use of pseudonymous* data, in such a way that the processing will meet the requirements of this Regulation'. Similarly, the Council draft added a specific reference to the use of pseudonymous data into Article 30(1) of the proposed Regulation, which specifies data security obligations for both controllers and processors. This shows that the use of anonymisation and pseudonymisation could constitute an integral part of compliance with data security obligations. However, in the event that the anonymisation or pseudonymisation is reversible, for example, for the purpose of establishing correlations between various data, organisational measures should complement the technical measures, for example, ensuring the confidentiality of the secret key and tracking accesses to such a key.

State of the art and context of the processing

It has to be noted that the mere use of anonymisation and pseudonymisation might not be sufficient for compliance. According to Article 17, the technical and organizational measures should be *appropriate*. The term 'appropriate' aims at highlighting 'the impact that different "security related threats and vulnerabilities might have on an organisation's data processing"' (OPTIMIS, 2011, p.28). According to Recital 46, the appropriateness of a security measure should be assessed taking into account *the state of the art* and the costs of their implementation in relation to the *risks inherent in the processing and the nature of the data to be protected*. Overall, the Directive requires neither the implementation of the most sophisticated security measures available in the market nor does the use of obsolete security technology meet the requirements of the Directive (OPTIMIS, 2011, p.27).

The Directive does not refer to any technology as 'state of the art' and borrows a similar concept from the fields of computer science and information security (OPTIMIS, 2011, p.27). Furthermore, the Directive does not refer to existing standards, thereby enabling more flexibility—both at national and organizational levels—in adopting appropriate measures for compliance with such requirements. However, there are attempts to map existing standards as 'state of the art' technologies. In this regard, Meints (2009) makes a distinction among the different international standards according to their target. Accordingly, international certification standards aiming at 'best practices' are considered as exceeding the *state of the art* whereas standards aiming at 'good practice' are considered to meet the *state of the art* requirements. This notwithstanding, member states might have different requirements. For example, where encryption is the appropriate security measure, the French Data Protection Authority (CNIL) recommends the use of state of the art algorithms, such as AES (Advanced Encryption Standard) or triple Data Encryption Standard (Triple-DES) for symmetric encryption and the use of RSA (Rivest, Shamir, and Adelman) or Elliptic Curve Cryptography (ECC) for asymmetric encryption (CNIL, 2010, p.31).

In addition, the context or circumstance of the processing should also be taken into account. This involves the type of data under processing (whether it is sensitive personal data or not), and other circumstances like whether they involve a transmission through a public network such as the Internet, and the costs of implementing the measure. This implies that, for example, an encryption technique that might be relevant for protecting certain kind of data may be considered insufficient when employed to secure the integrity of sensitive personal data that involves transmission over, for example, public Wi-Fi. Similarly, whether the encrypted data is stored in another jurisdiction (democratic or non-democratic) also plays a role. This is because most government authorities have sufficient resources to decrypt an encrypted data backed by statutory restrictions on the use of certain encryption methods.

4.2. ANONYMISATION AND PSEUDONYMISATION FOR FULFILLING THE PURPOSE SPECIFICATION AND LIMITATION PRINCIPLE

Pursuant to Article 17(1) of the Directive, the technical and organizational measures to be implemented should also protect personal data '[...] *against all other unlawful forms of processing*'. This aspect attempts to establish compliance with all the principles under the Directive. One such principle, apart from data security, that can be complied with through the use of anonymisation or pseudonymisation is the purpose specification and limitation principle. The purpose specification principle, which is laid down under Article 6(1)(b) of the Directive, sets the boundaries within which personal data may be processed by providing that the controller is only allowed to process the personal data under specific and legitimate purposes and that the data may not be processed further in a manner that is not compatible with those purposes that were originally specified. Hence, it is required to set clear purposes for each processing activity prior to the collection of such data, and not use the collected personal data for other incompatible purposes. Furthermore, Article 6(1) (e) provides that personal data must not be '*kept in a form that permits identification of data subjects for no longer than is necessary for the purposes for which the data were collected or for which they are further processed*.'

This implies that when the data are no longer necessary for the purposes for which they were originally collected or for which they are further processed, they should be deleted or anonymised. In fact, the Article 29 Working Party underlines that Article 6(1)(e) of the Directive makes a strong point that personal data should be anonymised 'by default' (Article 29 Working Party, 2014a, p.7).

Moreover, Article 6(1) of the ePrivacy Directive states that *'Traffic data relating to subscribers and users processed and stored by the provider of a public communications network or publicly available electronic communications service must be erased or made anonymous when it is no longer needed for the purpose of the transmission of a communication …'*[18] In addition, Article 9(1) of the ePrivacy Directive provides that *'where location data other than traffic data, relating to users or subscribers of public communications networks or publicly available electronic communications services, can be processed, such data may only be processed when they are made anonymous, or with the consent of the users or subscribers to the extent and for the duration necessary for the provision of a value added service.'*[19]

Both Articles 6(1) and 9(1) of the ePrivacy Directive require the use of anonymisation. The former is essentially similar to Article 6(1)(e) of the Data Protection Directive in that traffic data must be deleted or made anonymous when such data is no longer necessary for the purpose (i.e. *transmission of a communication)*. On the other hand, the use of anonymisation under Article 9(1) provides a legitimate purpose to process location data. This implies anonymising traffic data provides a legitimate ground for processing such data. For such compliance under Article 6(1)(e) of the Data Protection Directive and Articles 6(1) and 9(1) of the ePrivacy Directive to be achieved, the data should be anonymised *in such a way that the data subject is no longer identifiable.* As elaborated above, the anonymisation should be used so as to irreversibly prevent identification (Article 29 Working Party, 2014, p.7). In other words, pseudonymisation of data that would enable the re-identification of the person would not be sufficient. Therefore, deleting

[18] Traffic data is defined as any data processed for the purpose of the conveyance of a communication on an electronic communications network or for the billing thereof (Article 2(b)).

[19] Location data constitutes any data processed in an electronic communications network, indicating the geographic position of the terminal equipment of a user of a publicly available electronic communications service (Article 2(c)).

the data is a more privacy-friendly solution and preferable, although this may entail a trade-off with the utility of the data (ENISA, 2012, p.44).

In this regard, some member states explicitly require the legitimate duration of certain kinds of processing. For example, the French DPA (CNIL) issued a simplified notification norm, which defines the kind of data that can be processed by the employer and the maximum duration of storage, which in general is two months (ENISA, 2012, p.47). This implies that employers have to anonymise or delete the data after the expiration of two months. Similarly, the Italian Data Protection Authority (2012) indicates that data subjects have the right to object, in whole or in part, on legitimate grounds, to the processing of personal data concerning him/her, even though they are relevant to the purpose of the collection. More particularly, in case of a breach of the law, data subjects may have their data blocked, erased, or anonymised. This implies that in Italy, anonymisation could be mandated subsequent to a breach of the data privacy rules.

The different drafts for the proposed GDPR contain additional provisions on the role of pseudonymisation as an integral part of compliance with the 'legitimate purpose' principle. Article 6(2) of the Commission draft permits the processing of personal data, which is necessary for the purposes of historical, statistical, or scientific research subject to the conditions and safeguards referred to in Article 83. Accordingly, Article 83(1) allows such processing only if (a) these purposes cannot be otherwise fulfilled by processing data which does not permit or not any longer permit the identification of the data subject; (b) data enabling the attribution of information to an identified or identifiable data subject is kept separately from other information, as long as these purposes can be fulfilled in this manner. Although Article 83 requires the fulfilment of both conditions in paragraphs 'a' and 'b', the latter, in particular, is a reference to the use of pseudonymised data and securely managing the key for such data. There are also some suggestions to enable bodies conducting historical, statistical, or scientific research to publish or otherwise publicly disclose personal data if the personal data is processed for the purpose of generating aggregate data reports, wholly composed of anonymous data, pseudonymous data, or both.[20] Furthermore, the recommendation from the Committee on the Internal Market and Consumer Protection to the LIBE draft proposed the inclusion of a legitimate ground for processing under Article 6 if only pseudonymous data is processed. This implies that processing pseudonymous data as defined in Article 4(2(a)) of the LIBE draft Regulation serves as a legitimate basis for processing personal data.

5. CONCLUSION

To date, discussions on anonymisation and pseudonymisation have been more focused on the technical aspects of these measures and the associated risks of re-identification. When legal perspectives are attached, they are only limited to whether a certain technique meets the conditions for providing the safe harbour from the entire application of data privacy rules without going to other roles of such measures—a reference to 'all or nothing'. This article takes the next step and provides an in-depth analysis of the role of anonymisation and pseudonymisation (including encryption) under the current and *en route* European Data Privacy rules—a reference to the 'beyond the all or nothing approach'. To this end, the article identifies three major roles of anonymisation and pseudonymisation. First, anonymisation and pseudonymisation can serve as a safe harbour from the application of data privacy rules entirely, provided they are used to irreversibly prevent identification. Second, they can provide a safe harbour from certain data privacy obligations such as the notification of personal data breaches provided they are engineered appropriately and complemented by adequate organisational

[20] See the suggestions from the Committee on Industry, Research, and Energy to the LIBE draft.

measures. Third, they can constitute mandated requirements for compliance with data privacy obligations such as the data security and purpose specification and limitation principles.

This notwithstanding, it is worth noting that the above three roles are interrelated. This implies that if a data controller is compliant with the data security requirements under data privacy rules and implements all the appropriate measures in advance, it is likely that it will benefit from safe harbour for notifying breaches under the same rules. Similarly, implementation of the appropriate technological protection and organisational measures is relevant in ascertaining whether a personal data breach has taken place and thereby comply with the notification requirements. Furthermore, an anonymisation that aims at compliance with data privacy rules—for example, with the purpose limitation principle—leads to providing the safe harbour from compliance with the rules in entirety so far as it is implemented in a manner that prevents reverse identification.

REFERENCES

Article 29 Working Party (2014a), Opinion 05/2014 on Anonymisation Techniques, WP216.

Article 29 Working Party (2014b), Opinion 03/2014 on Personal Data Breach Notification, WP213.

Article 29 Working Party (2007), Opinion 4/2007 on the Concept of Personal Data, WP136.

Barcelo, R, & Traung, P (2010), The Emerging European Union Security Breach Legal Framework: The 2002/58 ePrivacy Directive and Beyond. In S. Gutwirth et al. (eds.), Data Protection in a Profiled World (Springer), pp. 77–104.

Burdon, M, Reid, J.F, and Low, R (2010), 'Encryption safe harbours and data breach notification laws', Computer Law and Security Review 26(5).

Bygrave, L (2002), Data Protection Law: Approaching Its Rationale, Logic and Limit (The Hague: Kluwer Law International)

Cavoukian, A (2013), Looking Forward: De-identification Developments—New Tools, New Challenges (Canada Information & Privacy Commissioner, May 2013)

CNIL (2010), Guide: Security of Personal Data, available (link), accessed 28 September 2014

Commission, 'Proposal for a Regulation of the European Parliament and of the Council on the Protection of Individuals with regard to the Processing of Personal Data and on the Free Movement of Such Data (Draft General Data Protection Regulation)' Com (2012) 11 final (**Commission draft**)

Commission Regulation (EC) **611/2013** of 24 June 2013 on the Measures Applicable to the Notification of Personal Data Breaches under Directive 2002/58/EC of the European Parliament and of the Council on Privacy and Electronic Communications OJ L173/2

Council Directive 95/46/EC of 23 November 1995 on the Protection of Individuals with regard to Processing of Personal Data and the Free Movement of such Data OJ L281/31 (Data Protection Directive).

Committee on Civil Liberties, Justice and Home Affairs, European Parliament, 'Report on the proposal for a regulation of the European Parliament and of the Council on the protection of individuals with regard to the processing of personal data and on the free movement of such data' (2013) (COM(2012)0011 – C7-0025/2012 – 2012/0011(COD)) (link) (**LIBE draft**).

Council of the European Union, document 17831/13 on proposal for a regulation of the European Parliament and of the Council on the protection of individuals with regard to the processing of personal data and on the free movement of such data (2013) (link) (**Council draft**).

Costa, L and Poullet, Y (2012), 'Privacy and the regulation of 2012', Computer Law and Security Review (28) 254-262.

Directive **2009/136/EC** of the European Parliament and of the Council of 25 November 2009 amending Directive 2002/22/EC on universal service and users' rights relating to electronic communications networks and services, Directive 2002/58/EC concerning the processing of personal data and the protection of privacy in the electronic communications sector, and Regulation (EC) No 2006/2004 on cooperation between national authorities responsible for the enforcement of consumer protection laws [2009] OJ L 337/11.

Directive **2002/58/EC** of the European Parliament and of the Council of 12 July 2002 concerning the processing of personal data and the protection of privacy in the electronic communications sector OJ201/37 (ePrivacy Directive).

ENISA (2012), Study on data collection and storage in the EU (Deliverable: 2012-02-08).

Emam, K.E., and Alvarez, C (2014), A Critical Appraisal of the Article 29 Working Party Opinion 05/2014 on Data Anonymisation Techniques, Draft paper for a web conference.

Eur. Consumer Org., E-Privacy Directive: Personal Data Breach Notification (2011), *available at* http://www.beuc.org/publications/2011-09742-01-e.pdf

European Union Agency for Fundamental Rights (2013), Handbook on European data protection law (Belgium: Council of Europe)

German Federal Data Protection Act (Bundesdatenschutzgesetz).

Hon, W.K., Kosta, E., Millard, C., and Stefanatou, D (2014), Cloud Accountability: The Likely Impact of the Proposed EU Data Protection Regulation, Tilburg Law School Legal Studies Research Paper Series No. 07/2014

Hon, W.K., Millard, K., and Walden, I (2011), 'The Problem of "Personal Data" in Cloud Computing—What Information is Regulated? The Cloud of Unknowing', International Data Privacy Law 1(4).

Hon W. Kuan, Hörnle, J., & Millard, K (2012), 'Data protection jurisdiction and cloud computing – when are cloud users and providers subject to EU data protection law? The cloud of unknowing', International Review of Law, Computers & Technology 26(2-3).

House Of Lords, Opinions of the Lords of Appeal for Judgment in the Cause Common Services Agency (Appellants) v Scottish Information Commissioner (Respondent) (Scotland), SESSION 2007–08 [2008] UKHL 47 on appeal from: [2006] SCOTCS CSIH 58

ICO (2012), Anonymisation: Managing data protection risk code of practice.

Italian Data Protection Authority (2014), Decision Setting forth Measures Google Inc. Is Required to Take to Bring the Processing of Personal Data under Google's New Privacy Policy into Line with the Italian Data Protection Code, (link).

Italian Data Protection Authority (2012), Cloud Computing: How to Protect Your Data Without Falling From A Cloud, (link).

Kuner, C (1996), 'Legal aspects of encryption on the internet', International Business Lawyer 24.

Meints, M (2009), 'The Relationship between Data Protection Legislation and Information Security Related Standards' In Matyáš, V et al. (ed) The Future of Identity, IFIP AICT 298, IFIP International Federation for Information Processing.

Ohm, P (2009), 'Broken promises of privacy: Responding to the surprising failure of anonymisation', UCLA Review 57

OPTIMIS Project D7.2.1.2. (2011), Cloud Legal Guidelines: Data Security, Ownership Rights and Domestic Green Legislation (Part II).

Oswald, M (2014), 'Share And Share Alike? An Examination of Trust, Anonymisation and Data Sharing with Particular Reference to an Exploratory Research Project Investigating Attitudes to Sharing Personal Data with the Public Sector', SCERIPTed 11(3).

Perkins, A (2005), 'Encryption Use: Law and Anarchy on the Digital Frontier [comments]', Houston Law Review 41(5)

Ponemon Institute (2009), US Enterprise Encryption Trends

Regulation (EU) No 910/2014 of the European Parliament and of the Council of 23 July 2014 on electronic identification and trust services for electronic transactions in the internal market and repealing Directive 1999/93/EC OJ L 257/73.

Robinson, N., Graux, H., Botterman, M and Valeri, L (2009), Review of the European Data Protection Directive, RAND Europe Technical Report, (link) accessed 28 August 2014.

Royal Decree (REAL DECRETO 1720/2007, de 21 de diciembre, por el que se aprueba el Reglamento de desarrollo de la Ley Orgánica 15/1999, de 13 de diciembre, de protección de datos de carácter personal

Wayne, J, and Grance, T (2011), Guidelines on Security and Privacy in Public Cloud Computing (NIST publications)