Draft
4 May 2018

Privacy through Biometric De-identification: Bridging the Gap Between Legal and Technological Perspectives

Catherine Jasserand* and Arun Ross**

Introduction:

Biometrics refers to the automated process of recognizing individuals based on their physical and behavioral attributes such as face, fingerprints, iris, gait, and voice. A face biometric system, for example, compares two face images and computes a match score indicating the degree of similarity or dissimilarity between them. This automated comparison process can be used to determine the identity of an unknown face image.

Recent research in biometrics has investigated the development of algorithms to perturb digital face images such that the identity of an individual cannot be deduced (see Figure 1). While such an approach can enhance data privacy, the perturbed images cannot be reliably used by a biometric system for person recognition thereby preempting their utility in the context of biometrics.

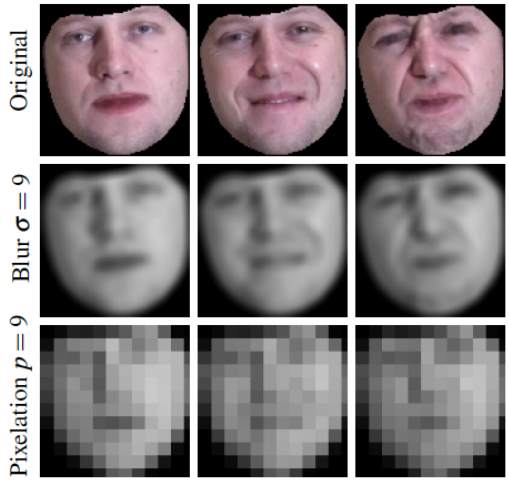

Figure 1. This image from Gross et al. (2009) illustrates the process of "de-identifying" an original face image through the application of image processing applications such as blurring and pixelation.

The new data protection regulation ('GDPR') adopted by the European Union (EU) offers an excellent opportunity to assess facial images from a data privacy perspective, and define the framework applicable to de-identified data.[1] Adopted in April 2016, the General Data Protection Regulation ('GDPR') replaces the previous Data Protection Directive (Directive 95/46/EC). The instrument regulates the processing of personal data, except for law enforcement purposes and purposes falling outside the scope of EU

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

law.[2] Operations such as collection, transformation, storage, dissemination or re-use of personal data are understood as processing. The GDPR is the first EU instrument on data protection that has introduced a statutory definition of biometric data. Under the new regulation, facial images belong to the category of sensitive data if they are used to 'uniquely identify' an individual.[3] As such, they are subject to stringent data protection rules prior to their processing. One way to secure the data and 'reduce' their sensibility would be to 'de-identify' the facial images. But is it possible to separate the identifying elements of a facial image to preserve privacy while still ensuring their usability for biometric recognition? The purpose of the paper is thus to analyze the level of de-identification that is applicable to facial images to ensure data usability and data privacy at the same time.

In this work, we perform the following tasks:

1. Investigate the different types of de-identification schemes that have been proposed in the face recognition literature and explore the development of a set of criteria to establish the "degree of identifiability" of the resulting de-identified data.
2. Analyze the requirements specified in the legal text of reference to establish whether we can define a level of de-identification that will ensure data usability. As the GDPR is still a new untested instrument, we will study the literature on data protection as it relates to technological issues (such as privacy enhancing technologies, privacy by design, data pseudonymization, etc.), as well as pre-GDPR reports by EU bodies (among others, opinions of the Article 29 Data Protection Working Party on anonymization and pseudonymization (Opinion 05/2014) and on biometric technologies (working document on biometrics, Opinion 02/2012 and Opinion 3/2012)).
3. Investigate the differences and similarities between de-identification from a machine vision perspective (i.e., face recognition using an automated face matching algorithm) and a human vision perspective (i.e., face recognition using the human brain).
4. Attempt to define a 'level of de-identification' applicable to facial images from both legal and technological perspectives.

Keywords:
Identification; test of identifiability; reversibility; pseudonymization; risk-based approach, personal identifiers; biometric accuracy; biometric recognition

I. Conflicting interests

As established by several scholars, [4] data privacy and data utility may pursue different aims. Their relationship is often described in terms of trade-off. According to Wu, defining the terms and the context of the relationship between privacy and data utility is critical in understanding their relationship.[5] In the field of biometrics, some authors have shown that privacy seeks to obscure as much identifying information as possible; whereas for the purpose of matching, biometric recognition needs to preserve as much identifying information as possible. There is thus an inherent conflict between the privacy of an individual's biometric data and the processing of their data for biometric recognition purposes.

[2] Article 1(1) GDPR.
[3] Article 9(1) GDPR.
[4] Rubinstein (2016), 2.
[5] Wu (2013).

In the context of this paper, we use the term 'privacy' in the sense of 'data privacy', 'data protection' or 'informational privacy.'[6] As for 'data utility', it refers to the threshold of utility of biometric data processed to perform biometric recognition. In the context of biometric template protection, some researchers have suggested a curve to illustrate the tension between privacy protection and biometric accuracy.[7] We use a similar curve to illustrate the tension existing between privacy and biometric accuracy. This curve, comparing privacy protection to data utility, is known in other fields.[8]

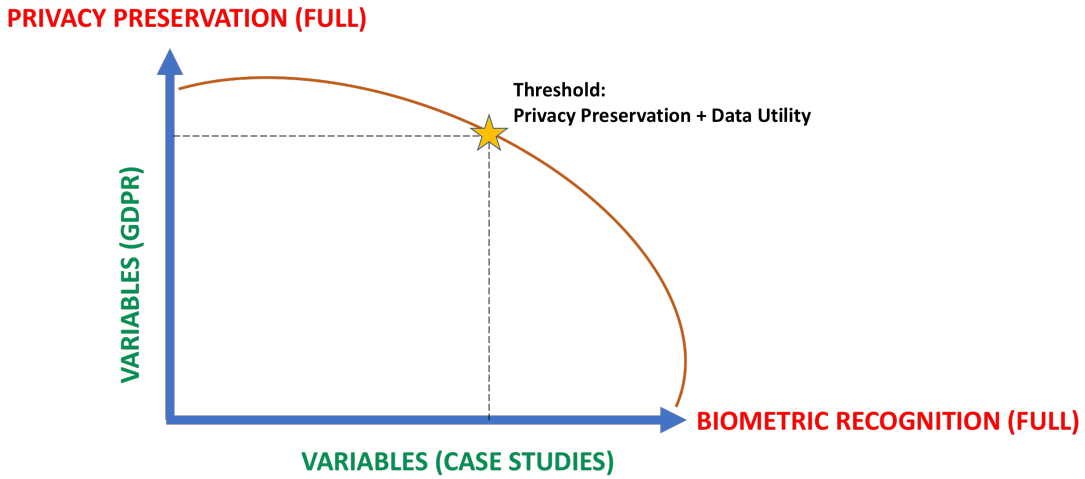

Figure 2. The interplay between personal privacy and data utility in the context of biometric recognition.

According to the curve, while the privacy preservation increases, the biometric accuracy decreases. The challenge is therefore to determine a level of identifiability that allows for both the privacy of biometric data (as personal data) as well as preserve their utility for biometric recognition. This paper attempts to define the variables of this opposition from a legal and a technological perspective. It takes the example of facial images, which constitute a specific type of personal data, and questions the possibility to de-identify them while retaining their utility.

The purpose of the paper is not, therefore, to find a mathematical formula that could define the level of identifiability from both legal and technological perspectives. We do not believe that such a formula can be found, in particular, because the concepts of 'identification' and 'identifiability' are not static concepts. Their meaning is subject to interpretation, in particular the test of 'reasonable means' used to determine whether data relates to an identifiable individual.[9] From a biometrics perspective, the identification of an individual is a probabilistic event that is impacted by the rate of false match (e.g., wrong identification) and the rate of false non-match (e.g., failure to identify).

---

[6] The term 'privacy' has several meanings and can be approached from different perspectives ; there are also differences in meanings between the European approach - where privacy can encompass data protection (Council of Europe's level) or being distinct from data protection (EU level) – and the US approach – where privacy can be interpreted as encompassing the protection of personally identifiable information, designated as 'data privacy' or 'informational privacy'

[7] See Rathbeg and Busch (2018).

[8] Such as in the field of clinical trial; see for instance the background report written by El Emam and Malin on 'Concepts and Methods for De-identifying Clinical Trial Data', Annex B to Report « Sharing Clinical Trial Data : Maximizing Benefits, Minimizing Risks', available at https://www.nap.edu/read/18998/chapter/10

[9] See Recital 26 GDPR and the analysis of the "all the means likely reasonably" test in the next section.

II. Variables

One of the challenges of this topic is linked to the terminology. First of all, the terms are defined differently from one discipline (data protection) to the other (biometric recognition). Second, the statutory meanings of *biometric data* as well as *identification* are subject to interpretation.

The second challenge is related to the definition of biometric accuracy, which is linked to measurements in terms of comparison errors (false acceptance and false rejection).

Finally, the last issue revolves around the nature of facial images and their qualification as personal and/or biometric data from a data protection perspective.

1. Terminology

Defining the terms from a legal and a technical perspective is key in the debate on the de-identification of biometric data. It is impossible to discuss how to de-identify biometric data without first defining biometric data and identification from a data protection and a technological perspective.

a) Biometric data

Data Protection perspective:
A face contains distinctive characteristics, which allows the recognition of an individual. Once the physical or behavioral attributes of an individual have been captured (via a sensor or a camera), they are transformed into a format that will enable the recognition process. The data generated during the process of feature extraction is usually called 'biometric data'. Thus, biometric data does not designate the characteristics themselves but their representations as images, templates or any other forms generated during the process. A template designates the mathematical representation of the biometric characteristics. From a biometrics perspective, facial images and facial templates are both considered to be biometric data.

From an EU data protection perspective, biometric data is a particular type of personal data. As observed by the A29WP, an EU advisory body to the European Commission, biometric data is both *content* and *link* to an individual.[10] For instance, a facial image can link to an individual's identity, and is also sufficient to identify an individual (as it contains identifying facial attributes). The distinction made by the A29WP has not, however, been introduced in the GDPR. In Article 4(14), the new legal instrument defines 'biometric data' as:

> "Personal data resulting from specific technical processing relating to the physical, physiological or behavioural characteristics, which allow or confirm the unique identification of that natural person, such as facial images or dactyloscopic data."

And concerning facial images, Recital 51 clarifies the following:

[10] A29WP, Opinion 4/2007 on the concept of personal data, 8.

> "The processing of photographs should not systematically be considered to be processing of special categories of personal data as they are covered by the definition of biometric data only when processed through a specific technical means allowing the unique identification or authentication of a natural person."

The early interpretations of the definition diverge on the meaning of 'unique identification', which is crucial in the context of biometric data and their possible use. [11] It shows that the statutory meaning of 'biometric data' is not settled yet. Having said this, the definition refers to 4 components:

First, biometric data is *a type of personal data*. As a consequence, if the data at stake does not reach the threshold of personal data (Article 4(1) GDPR), it cannot qualify as biometric data. This is a pre-requisite condition. The threshold is linked to the identifiability to whom the data refers, as discussed later in this paper.

Second, biometric data results from *technical processing*. The definition does not mention it, but it is fair to assume that this specific technical processing operation encompasses the different formats generated during the biometric recognition process: from the capture of the biometric characteristics to their transformation into images and templates, and their storage for comparison at a later stage.

Third, biometric characteristics that are transformed during the process are described in the statutory definition as *physical, physiological and behavioral characteristics*. This description seems consistent with the technical definition.[12]

Last, biometric data '*allows or confirms the unique identification*' of an individual. This last criterion is not crystal-clear. First, it is not clear whether the phrase refers to the quality of biometric characteristics, the quality of personal data that are biometric data, or the result of the technical processing. The structure of the sentence does not shed light, especially if we consider that, in the GDPR, personal data as well as biometric data is a plural noun instead of a singular one. Scholars who have interpreted this definition do not agree on the meaning of this key phrase.[13] Recital 51 GDPR, which supplements Article 4(14), refers to photographs that fall within the category of biometric data when *processed* for one of these two purposes. The recital does not mention the quality of biometric characteristics themselves. It seems that qualification of biometric data is given to personal data that undergoes a technical processing for biometric recognition and not to personal data that has the ability to recognize individuals. However, referring to Recital 51 to interpret Article 4(14) GDPR may be problematic. When the recital describes the two functions for which biometric data is processed, the wording of Recital 51 is not in line with that of Article 4(14).

According to Recital 51, photographs –understood here as photographs portraying individuals – become biometric data when "processed through a specific means allowing the unique identification or authentication of a natural person." From that wording, one can deduce that 'unique identification' refers to biometric identification; whereas authentication is used instead of verification. It is quite common to use these two terms

[11] Kindt considers that 'unique identification' refers to the processing of biometric data for (biometric) identification purposes; whereas Jasserand suggests that the expression be understood from a data protection perspective and cover both verification and identification functions; see respectively Kindt (2018), 'Having yes, saying no? About the new legal regime for biometric data,' in Computer Law & Security Review and Jasserand (2016), 'Legal nature of Biometric Data: From 'Generic' Personal Data to Sensitive Data', in EDPL, vol. 2.
[12] see e.g. ISO/IEC 2382-37, term 3.1.3
[13] Ibid.

as synonyms, even if the biometric community prefers avoiding the use of 'authentication'. Biometric identification and verification are two of the purposes for which biometric data are usually processed.[14] They are part of 'biometric recognition.'

But this reading is inconsistent with the wording of Article 4(14) GDPR. In that provision, the two functions are defined as 'allowing the unique identification' and 'confirming the unique identification.' Thus, the phrase 'unique identification' cannot refer to the biometric identification function. Instead, we argue that the expression 'unique identification' should be understood from a data protection perspective. As such, it would not replace the biometric identification function but, instead, refer to the highest threshold of identification, where individuals are identified thanks to their 'unique' attributes, i.e. their biometric characteristics. Under the previous Data Protection regime, Kotschy made a similar analysis. She considered that 'biometric data (photos, fingerprints, DNA) or personal identification numbers (PINs)…would permit to achieve truly *unique identification*."[15] To confirm this position, one could observe that the phrase 'unique identification' is not only used in relation to the definition of 'biometric data.' It is also found in relation to health data in Recital 35 GDPR, which includes "a number, symbol or particular assigned to a natural person to *uniquely identify* the natural person for health purposes."[16] Besides, in Opinion 4/2007 on the concept of personal data, A29WP analyzed the particular nature of biometric data. It noted their 'unique link' to an individual to allow his or her identification. Following this analysis, 'unique identification' might not refer to biometric identification but to an individual's identity instead. As such, the processing of biometric data would allow either the establishment of an individual's identity (described as 'allowing the unique identification') or the confirmation of his or her identity (described as 'confirming the unique identification'). The discussion on the scope of the notion of 'unique identification' is crucial to determine which data benefits from the protection granted to sensitive data. According to Article 9(1) GDPR, only "biometric data processed for the purpose of uniquely identifying" an individual is sensitive data.

On the basis of the interpretation made in this section, biometric data processed for either biometric identification or verification but that can be linked to an identified individual is sensitive biometric data. This identification –in the sense of being identified- would result from the positive matching of data sets (for either purpose). Biometric data relating to identifiable individuals, i.e. the link to a specific individual has not been established or cannot be verified, would be excluded from the scope of sensitive data unless they reveal other sensitive information.[17]

Technological perspective

Before the adoption of the GDPR, many legal reports made reference to biometrics and use the term as a synonym of 'biometric data.' The term has a very precise meaning in the field of biometric recognition. It refers to the automated process through which biometric samples are recognized. The paper uses the term 'biometrics' as defined by the biometrics community. The paper mainly focuses on *automated* biometric recognition and thus on the *automated* processing of biometric data.

[14] see Jain et al. (2011)
[15] Kotschy (2010), 35.
[16] Emphasis added.
[17] Such as ethnicity, health condition, etc., see Article 9(1) GDPR.

According to the ISO/IEC standard developed on the harmonization of biometric vocabulary,[18] biometric data refers to the different formats resulting from the different stages of biometric recognition. However, contrary to the legal concept, biometric data from a technological perspective does not need to relate to an individual.[19] This is a major distinction. The focus of the technical definition is on the generated formats.

b) Identification /Identifiability

Data Protection perspective:

The notion of identification is not defined in the GDPR, neither was it in the previous Data Protection Directive. Article 4(1) GDPR describes the concept of personal data as relating to an individual, who is identified or identifiable. As observed by Bygrave, the two conditions are cumulative but cannot be dealt separately because "data will normally relate to, or concern a person if it enable that person's identification." [20] They key criterion in the definition of personal data is the term 'identification.' Following Article 4(1) GDPR, data that relates to an identified or identifiable individual is personal data. Article 4(1) specifies what an 'identifiable' individual is:

> "One who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person."

The GDPR refers to direct and indirect identification and provides various examples. To qualify as personal data, the data only needs to relate to identifiable individuals (and not identified ones). As observed by the A29WP in Opinion 4/2007 on the previous regime, identification only relates to the ability to identify, i.e. to single out or distinguish, an individual.[21] Bygrave defines this 'identifiability' as the "ability to distinguish [a person] from others by linking him or her to pre-collected information of some kind. As such, identification does not require knowledge of a person's name but it does require knowledge of some unique characteristics of the person relative to a set of other persons." [22] The threshold of identification is thus rather low: the GDPR, like the Data Protection Directive, does not require the actual identification of an individual but the ability to identify him or her.[23] However, as interpreted in the previous sub-section, biometric data falls within the remit of sensitive data if it relates to an *identified* individual (*processed for the purpose of uniquely identifying*).

Identified and identifiable:

The GDPR does not define 'being identified'. But the A29WP did it in Opinion 4/2007. Being identified is being *singled out* or being *distinguished* from a group of individuals. Thus, being identified from a data protection perspective does not mean to establish an individual's (civil) identity but only to be able to distinguish him or her from others. The identification of an individual might depend on the context of a situation. Sometimes a very common name is not sufficient to identify an individual. But other times, the name will be sufficient if the person is a well-known individual (such as a politician, a celebrity).

---

[18] IEC stands for the International Electrotechnical Commission; ISO/IEC JTC 1/SC 37, ISO/IEC 2382-37 : 2017, Information Technology-Vocabulary- Part 37 : Biometrics

[19] See note below term 37.03.06, ISO/IEC 2382-37:2017(E)

[20] Bygrave (2014), 129-130.

[21] A29WP (2007), Opinion 4/2007, 13-15.

[22] Bygrave (2014), 130.

[23] On the interpretation of 'identification' under the Data Protection Directive, see Kotschy.

Concerning 'being identifiable', the GDPR is more detailed. Recital 26 GDPR provides, in particular, a test to determine the identifiability of an individual.

Test of identifiability: "all the means reasonably likely to be used":
Recital 26 GDPR has introduced a test of "reasonable means" of identification applicable to a data controller or to a third party. The recital provides the following guidance on the application of the test:

> " 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 amount of time required for identification, taking into consideration the available technology at the time of the processing and technological developments."

The test is new but was suggested by the A29WP in Opinion 4/2007. As observed by the A29WP, the 'mere' likelihood to identify an individual at some point of time is not sufficient.[24] Thus, a test of 'identifiability' – known as "all the means reasonably likely"- has to be carried out. The test has also been interpreted by the European Court of Justice (ECJ) in the *Breyer* case.[25] The Court found, in particular" that "if the identification of the data subject was prohibited by law or practically impossible on account of the fact that it requires a disproportionate effort in terms of time, cost and man-power, so that the risk of identification appears in reality to be insignificant."[26] One should note that the state-of-the art of technologies used for identification purposes is mentioned as one of the factors. Thus, the state-of-the art in biometric recognition, and thus in facial recognition, should be taken into account to assess the identifiability of an individual. This issue will be further discussed.

Last, the test of 'identifiability' is specified in relation to personal data and pseudonymized data, as this will be explained in the next section.

Technological perspective
Identification from a technological (i.e., biometrics) perspective has a very different meaning. It is usually different from verification. Together, identification and verification constitute biometric recognition. Given a face image, *identification* from a biometric recognition perspective entails comparing this image with several other face images in a database in order to determine a corresponding match. This is what is called the one-to-many comparison. Police use this operation to determine the origin of a latent fingerprint found at a crime scene for instance. By contrast, *verification* entails confirming or refuting a match between the given face image with another one stored on a device or database. Verification does not aim at establishing who an individual is. It is often called authentication, like in Recital 51 GDPR, but the biometric industry should instead refer to verification, as specified in ISO/IEC 2382-37:2017 on harmonized biometric vocabulary. [27]Verification is performed by custom guards when they check the photograph displayed in a passport with the face of the individual holding the passport.

[24] A29WP, Opinion 4/2007 on the concept of personal data, 20 June 2007, WP 136, 15.
[25] Case C-582/14, Patrick Breyer v Bundesrepublik Deutschland
[26]Ibid, para. 25.
[27] See ISO/IEC JTC 1/SC 37, ISO/IEC 2382-37 : 2017, Information Technology-Vocabulary- Part 37 : Biometrics.

Identification from a biometrics perspective has thus a narrower meaning than identification from a data protection perspective. However, from both perspectives, identifying an individual does not mean establishing his or her civil identity. In the first case, it means singling out an individual; in the second case finding a matching set of biometric data. The identity of an individual will thus be established based on metadata (or demographic data) associated with the biometric data.

From this analysis made in this section, it appears essential to define the terms taking into account their context of use. In the context of data protection, the terms 'biometric data' and 'identification' have different meanings and scope than in the context of biometric recognition.

2. Data utility and accuracy

Biometric accuracy can be expressed in terms of errors (false acceptance and false rejection). The accuracy of facial recognition depends on many factors, including the resolution of the image, head pose, external illumination and facial expressions. Methods to determine face recognition accuracy are multiple, and the results of face recognition are based on matching probabilities.

3. Facial images

From a data protection perspective, facial images are a particular type of personal data as previously explained. Contrary to other types of personal data, a facial image does not need additional information to single out an individual. The additional information is only necessary to establish the civil identity of an individual.

Following Recital 51 GDPR, photographs are not necessarily classified as biometric data: they need to be processed for biometric recognition purposes to be classified as such.

Due to its nature, can a facial image be de-identified? In particular is it possible to separate the identifying information contained in a picture from the data itself? If so, is it still a facial image? And more importantly can it still be used for facial recognition?

III- De-identification of Facial Images

To protect individuals' data privacy, one of the measures envisaged is to de-identify their personal data. But what does it mean in the context of GDPR? And how can biometric data, such as facial images, be de-identified? These are some of the questions raised in this section.

The debate on the meaning of de-identification is not settled. The term does not appear in the GDPR. And according to some scholars,[28] depending on the reversibility of the process, de-identification might either refer to pseudonymization (reversible process) or to anonymization (irreversible process). For others, the term is used as an umbrella term to cover both pseudonymization and anonymization.[29] As observed by Polonetsky et al, "[d]espite a broad consensus around the need for and value of de-identification, the debate as to whether and when data can be said to be truly de-identified has

[28] See definition in Ribaric, 293.
[29] Hintze and El Emam (2017), 3.

appeared interminable. Although academics, regulators, and other stakeholders have sought for years to establish common standards for de-identification, they have failed so far to adopt even a common terminology." [30] There is also no consensus on the meaning of anonymization, which can, indeed, refer to the use of pseudonyms.

1. De-identification from a data protection perspective

The GDPR does not define or refer to the concept of de-identification. In the field of biometrics, some scholars describe de-identification as "…the process of removing or concealing personal identifiers or replacing them with surrogate personal identifiers to prevent direct or indirect identification of a person." From this definition, the question stems whether it is possible to define a threshold of de-identification, which would preserve the privacy of an individuals' biometric data while still allowing for their identification.

Pseudonymization

The GDPR has introduced a new category of personal data, pseudonymized data, to ensure data security[31] and data minimization.[32] In Article 4(5) GDPR, pseudonymization is defines as:

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

Pseudonymization is thus composed of three elements: first, it ensures the protection of individuals' personal data by 'delinking' the data from a specific individual; second the data can be only re-attributed to an individual with additional information; and third, the 'additional information is kept separately' and protected through privacy by design measures. From this definition, one understands that pseudonymization is a *reversible* process, which might, however require some effort to trace back the individual to whom the data belongs. Pseudonymization makes it more difficult to identify the individual by minimizing the risk of identification.[33] However, the individual remains identifiable. This is why pseudomymized data remains personal data under the GDPR.[34] To determine an individual's identifiability, a test of "all the means reasonably likely to be used" needs to be carried out. As presented in the previous section, the test is based on different factors defined in Recital 26 GDPR.

Anonymization
By contrast, the GDPR also refers to anonymous or anonymized data, which is defined in Recital 26 GDPR as:

> "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."

[30] Polonetsky et al, 596.
[31] Art. 32 GDPR
[32] Art. 25 GDPR
[33] Recital 28 provides that « the application of pseudonymisation to personal data can reduce the risks to the data subjects concerned and help controllers and processors to meet their data-protection obligations. »
[34] Recital 26 GDPR.

Anonymized or anonymous data are not personal data and are thus excluded from the scope of the GDPR. One could deduce that the process is irreversible but this position is disputable. The GDPR does not, indeed, refer to the irreversibility of the process. Should the data be linked to the individuals again, they would fall under the scope of the GDPR. But since the paper is more interested in reversible processes than in irreversible ones – which might be useless for biometric recognition purposes, it will focus on the pseudonymization of biometric data. And in particular, it will address the issue of de-identification (i.e. pseudonymization) of facial images and facial templates.

2. De-identification from a technological perspective

A number of de-identification schemes have been proposed in the biometrics literature. The ad-hoc solution for face de-identification is to naively distort images by blurring or pixelation;[35] however, this technique obscures facial details which reduces the utility of the biometric data. Other deidentifying techniques that preserve much of the facial details have been proposed in the literature. Newton et al. (2005) proposed the k-same technique in which the facial color and texture of k faces are averaged, thereby reducing the chance of uniquely identifying a face image. Later, Gross et al. (2006) improved this technique and proposed the k-same-M approach by incorporating the Active Appearance Model (AAM). Face swapping was proposed by Bitouk et al. (2008), where a source face is seamlessly blended with candidate images similar in appearance and pose, resulting in new de-identified faces. Jourabloo et al. (2015) adopted the k-same algorithm and proposed an optimization scheme to find the optimal set of weights for the k face images to preserve facial attributes such as gender, age-group, ethnicity and other details like eyewear.

Application of three conditions defined in Article 4(5) GDPR:
- no link to a specific individual
- need for additional information to re-identify the data
- separation between data and individual ensured by Privacy by Design measures (e.g. encryption of the data; identity kept separately)

Key issue: is it possible to pseudonymize facial images since the images are or constitute the link to an individual from a data protection perspective?

Another issue: a template is different from an image as the template cannot allow the direct identification of an individual. However, the distinction between human-based recognition versus computer-based recognition is very relevant here: the human brain can only identify a facial image, it cannot identify a facial template. By comparison, an automated recognition can be performed equally on a facial image and on a facial template.

IV Suggested Solutions

According to some scholars, "true de-identification is not possible, or at least is not sustainable. Rather than focus on *how* to de-identify personal information, the discussion has increasingly shifted to *whether* personal information can be (or can be said to be) "de-identified" and thus not personally identifiable." [36]

[35] Boyle et al. (2000)
[36] Polonetsky et al, 599.

- Some of the solutions suggested to blur or pixelate facial images to protect individuals' privacy. But, as rightly observed by Gross et al. "In previous studies, it has been claimed that *ad hoc* de-identification methods, such as pixilation and blurring prevent humans from reliably recognizing the identity of de-identified images. However, as our experiments demonstrate, these methods can not prevent a computer from reliably performing recognition."[37]

- Reversibility of the de-identification process to allow biometric recognition.

- Solution proposed by Gross et al. in 'Integrating Utility into Face De-identification' is to keep data utility for *classification* purposes (gender, emotion) but no solution proposed for biometric recognition.

- Thus, pseudonymization of facial images seems to be a good option, as long as it is feasible to pseudonymize them. This would comply with the obligation of Data Protection by Design obligation that entities processing facial images must comply with. See Art. 25 GDPR. One of the measures suggested to implement the obligation is data pseudonymization.

- Some authors have found ways to preserve some 'face clues' to use the de-identified images for 'behaviour and emotions analysis.' However, these images cannot be used for biometric identification or identity verification.[38]

- Some researchers have attempted to modify a face image such that its biometric utility is retained, but information about other attributes such as gender are suppressed.[39]

- Currently, there are no metrics that can be found to define the relationship between privacy preservation and data utility.

[37] Gross et al. (2005), Section 6.1.
[38] Letournel et al. , as cited by Ribaric and Paveslic, 301.
[39] See Mirjalili and Ross (2017), Mirjalili et al. (2018), Othman and Ross (2014)

Selected literature:

Acquisti, A., Gross, R., Stutzman, F. (2014), *Face Recognition and Privacy in the Age of Augmented Reality*, Journal of Privacy and Confidentiality, Vol. 6, Issue 2.

Adler, A. and Schuckers, M. (2007), *Comparing Human and Automatic Face Recognition Performance',* IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics, vol. 7, no. 5, pp. 1248-1255.

Bitouk, D., Kumar, N., Dhillon, S., Belhumeur, P. and Nayar, S. K. (2008), Face swapping: automatically replacing faces in photographs. ACM Transactions on Graphics (TOG), 27(3):39.

Boyle, M., Edwards, C. and Greenberg, S. (2000), The effects of filtered video on awareness and privacy. In Proceedings of the ACM Conference on Computer Supported Cooperative Work, pages 1–10.

Dantcheva, A., Elia, P. and Ross, A. (2016), "What Else Does Your Biometric Data Reveal? A Survey on Soft Biometrics," IEEE Transactions on Information Forensics And Security (TIFS), Vol. 11, No. 3, pp. 441 - 467.

El Emam, K. and Alvarez, C. (2015), *A critical appraisal of the Article 29 Working Party Opinion 05/2014 on data anonymization techniques,* International Data Privacy Law, Vol. 5, Issue 1, pp. 73-87.

Gross, R., Sweeney, L., De la Torre, F. and Baker, S. (2006), Model-based face de-identification. In Computer Vision and Pattern Recognition Workshop (CVPRW).

Gross, R., Sweeney, L. et al (2009), 'Face de-identification', in A. Senior (ed.) *Protecting Privacy in Video Surveillance*, pp. 129-146.

Gross, R., Airoli, E., Bradley, M. and Sweeney, L. (2005), *Integrating Utility into Face De-Identification*

Hintze, M. (2016), *Viewing the GDPR Through a De-Identification Lens: A Tool for Clarification and Compliance*, paper available at https://fpf.org/wp-content/uploads/2016/11/M-Hintze-GDPR-Through-the-De-Identification-Lens-31-Oct-2016-002.pdf

Hintze, M. and El Emam, K. (2017), *Comparing the Benefits of Pseudonymization and Anonymization Under the GDPR,* Privacy Analytics White Paper.

Introna, L. and Nissenbaum, H. (2009), Facial Recognition Technology: A Survey of Policy and Implemention Issues, Center for Catastrophe Preparedness and Response, New York University, available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1437730

Jain, A. K., Ross, A., and Nandakumar, K. (2011), "Introduction to Biometrics: A Textbook", Springer Publishers.

Jasserand, C. (2016), *Legal Nature of Biometric Data: from 'Generic' Personal Data to Sensitive Data,* in European Data Protection Law Review, 3/2016, pp. 297-311.

Jourabloo, A., Yin, A. and Liu, X. (2015), Attribute preserved face de-identification. In International Conference on Biometrics (ICB), pages 278–285.

Kindt, E. (2014), Privacy and Data Protection Issues of Biometric Applications, A Comparative Legal Analysis, ed. Springer (Dordrecht).

Letournel, G., Bugeau, A., Ta, V-T, and Domenger, J.P. (2015), *Face De-Identification with Expressions Preservation,* Proc. 2015 IEEE Int. Conf. on Image Processing, pp. 4366-4370.

Li, S. Z. and Jain, A. K. (eds.) (2011), *Handbook of Face Recognition*, Second Edition, Springer, ISBN 978-0-85729-931-4.

Mann, M. and Smith, M. (2017), *Automated Facial Recognition Technology: Recent Developments and Approaches to Oversight*, in University New South Wales Law Journal, pp.121-145.

Mirjalili, V., Raschka, S., Namboodiri, A., Ross, A. (2018), *Semi-Adversarial Networks: Convolutional Autoencoders for Imparting Privacy to Face Images*, Proc. of International Conference on Biometrics (ICB), (Gold Coast, Australia).

Mirjalili, V. and Ross, A. (2017), "Soft Biometric Privacy: Retaining Biometric Utility of Face Images while Perturbing Gender," *Proc. of International Joint Conference on Biometrics (IJCB)*, (Denver, USA).

Newton, E., Sweeney, L. and Bradley, M. (2005) *Preserving Privacy by De-Identifying Face Images,* IEEE Transactions on Knowledge and Data Engineering, vol. 17, no. 2, pp. 232-243.

Othman, A. and Ross, A. (2014), "Privacy of Facial Soft Biometrics: Suppressing Gender But Retaining Identity," *Proc. of ECCV Workshop on Soft Biometrics*, (Zurich, Switzerland).

Polonetsky, J., Tene, O. and Kelsey, F. (2016), *Shades of Gray: Seeing the Full Spectrum of Practical Data De-Identification,* Santa Clara Law Review, vol. 56, no. 3, pp. 593-629.

Rathgeb, C. and Busch, C. (2018), 'Biometric Template Protection: State-of-the-art, Issues and Challenges', Chap. 8 in Vielhauer et al. (eds) *User-centric privacy and security in biometrics.*

Ribaric, S. and Paveslic, N. (2018), 'De-identification for Privacy Protection in Biometrics', Chap. 13. in Vielhauer et al. (eds) *User-centric privacy and security in biometrics.*

Ross, A. and Jain, A. (2007), *Human Recognition using Biometrics: An Overview,* in Annales des Telecommunications, vo. 62, no 1-2, 11-35.

Ross, A. and Othman, A. (2011), *Visual Cryptography for Biometric Privacy*, IEEE Transactions on Information Forensics and Security (TIFS), Vol. 6, Issue 1, pp. 70 - 81, March.

Rubinstein, I. (2016), 'Identifiability: Policy and Practical Solutions for Anonymisation and Pseudonymisation", Brussels Privacy Sumposium.

Spiekermann, S. and Cranor, L. (2009), *Engineering privacy*, IEEE transactions on Software Engineering, vol. 35, no. 1, January/February 2009, pp. 67-82.

Stalla-Bourdillon, S. and Knight, A. (2017), *Anonymous Data v. Personal Data- A False Debate: an EU Perspective on Anonymization, Pseudonymization and Personal Data,* Wisconsin International Law Journal, vol. 34, issue 2, pp. 284-322.

Zhao, W., Chellappa, R., Phillips, P.J. and Rosenfeld, A. (2003), *Face Recognition: A Literature Survey*, vol. 35, issue 4, pp.399-458.

*Catherine Jasserand is PhD researcher, University of Groningen, European Technology Law and Human Rights Department STeP (Security, Technology and ePrivacy) Research Group

**Arun Ross is Professor, Department of Computer Science and Engineering, Michigan State University, USA