A. Brömme, N. Damer, M. Gomez-Barrero, K. Raja, C. Rathgeb,
A. Sequeira, M. Todisco and A. Uhl (Eds.): BIOSIG 2022,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2022

When do the images of biometric characteristics qualify as special categories of data under the GDPR?: a systemic approach to biometric data processing

Bilgesu Sumer [1]

Abstract: As legal scholars observed, the definition of biometric data in the GDPR is not aligned with the technical definition in the international standards, which define biometric data as biometric samples, or aggregation of them at any stage of processing. The legal definition excludes the images of biometric characteristics from the sensitive data regime under Article 9 (special categories of data). If not considered sensitive, the images of characteristics can be processed based on a broad category of legal grounds, such as legitimate interests, including marketing purposes under Article 6. This article looks beyond the technical and legal definitions of biometric data and interprets the GDPR in two ways to address the confusing *status quo*. First, it dissects the objective nature of the data under the sensitive data regime of the GDPR. Second, it will systemically inquire about the meaning of purpose and how it can help us understand the margins of the sensitive data regime.

Keywords: biometric samples, Clearview AI, biometric data, facial recognition, GDPR.

1 Introduction

The New York Times reported in January 2020 that a US-based facial recognition software firm, Clearview AI (CW), created a biometric database owing to a collection of images gathered from online sources, such as Facebook, YouTube, and Twitter, along with related data, such as the image's URL source, geolocation, and occasionally the names of the subjects. CW has not been only scrutinized under US laws; it has been expelled and heavily fined across the EU as the GDPR has been violated by CW on many levels, including the unlawful processing, use, and processing of biometric data, pictures, and related information [Ja22].

The decisions of the European Data Protection Authorities (DPAs) on the matter once more demonstrated that the images initially processed do not have biometric data status under Recital 51 of the GDPR and Art. 4(14) of the GDPR and that such interpretation has become a general practice. [IC22], [GD22], [CN21].The decisions generally consider

[1] Doctoral Researcher at KU Leuven, Center for IT and IP Law (CiTiP), Biometric Law Lab (BLL), Sint-Michielsstraat 6 box 3443, 3000 Leuven, bilgesu.sumer@kuleuven.be. This research has received funding by the European Union's Horizon 2020 research and innovation program under the Marie Sklodowska-Curie Grant agreement No 860315.

three categories of personal data in the CW database:

(i) images of identifiable individuals;
(ii) metadata and URLs linked to the images;
(iii) the database (hashed) vectors originated from the images [IC22], [Ja22].

While all these three categories are used to build the biometric reference database, only the third category is considered sensitive under the GDPR and is subject to the regime of special categories of data under Article 9. However, according to the international technical standards, biometric data means biometric samples, i.e., images, or aggregation of such at any stage of processing [IS22]. This indicates a clear difference between biometric data's legal and technical understanding. The wording of the legislator creates several categories of biometric data, from regular personal data to sensitive data (special categories), and the protection granted to them differs accordingly [Ki18], [Ja16]. Moreover, the legislation might not be sufficient to protect the rights and freedoms of natural persons due to its over-reliance on the purpose of the processing [Ki18].

This study aims to contribute to an enhanced understanding of the relevant provisions of the GDPR by proposing an interpretation to protect the rights and freedoms of natural persons. It first demonstrates the controversial reading of the relevant provisions of the GDPR, i.e., Article 4(14) and Recital 51. Next, the paper asks whether prioritizing the data's sensitivity or the controller's purpose to interpret the theoretical boundaries of biometric data can better protect the rights and freedoms of natural persons. The main approach taken to interpret the relevant notions and provisions of the GDPR is systemic. The systemic interpretation presupposes that legal provisions shall be interpreted in a manner that is coherent with the "system," with the principles, regulations, and concepts defining the same area of the legal system to which the provision belongs. With a systemic interpretation, the relevant GDPR provisions, namely, the definition of the controller and the principle of accountability, are reviewed. Finally, I argue that the nature and, thus, the sensitivity of the data should prevail over the controller's purpose to provide better protection to data subjects.

2 Blurry boundaries of the biometric formats

The GDPR 4(4) defines biometric data as the data "resulting from **specific technical processing** relating to the physical, physiological or behavioral characteristics." It, however, does not define what this specific technical processing refers to. Typically the technical processing results in three formats:

(i) biometric sample: analog or digital representation of biometric characteristics, e.g., an image of a face. The result of the initial processing that converts signals from biometric characteristics [Ja16];
(ii) biometric feature: numbers or labels extracted from the sample [IS22], e.g., datapoints of a face;

(iii) biometric template: the mathematical construction derived from the biometric sample [Ja16]; or set of stored biometric features [IS22].[2]

Recital 51 states, "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 Recital explicitly excludes biometric samples from the scope of the definition of biometric data. However, according to the technical definitions, the processed images would be logically considered biometric data since this processing constitutes the very first phase of biometric data processing. The wording of the Recital means that biometric samples are not considered as sensitive as the features or templates, and they might not be protected under Article 9, which provides the legal basis for the processing of the special categories of personal data (sensitive data regime). Article 9 states that:

"The processing of personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership, and the processing of genetic data, **biometric data for the purpose of uniquely identifying a natural person**, data concerning health or data concerning a natural person's sex life or sexual orientation shall be prohibited."

European Data Protection Board (EDPB) also states that if a system does not generate biometric templates to identify persons uniquely but instead detects only the physical characteristics and consequently only classifies the person, the processing does not fall under Article 9 [ED20]. Regarding the same provision, legal scholars observe that the limitation of the prohibition to the unique identification purpose allows for collecting and storing biometric data for other purposes based on a less strict data processing regime in Article 6 of the GDPR [Ki18]. Article 6 provides the general legal basis for processing regular personal data, such as the necessity for the performance of a contract or the purposes of the legitimate interest pursued by the controller or a third party. The possibility to rely only on Article 6 for processing the images of characteristics paves the way for weaker protection for natural persons. The inconsistent language used in the EDPB guidance makes the issue more complicated.

The EDPB further states that ,"raw data will be the building block of any template"

[2] After, the feature extraction, biometric features are stored in a biometric reference database. At this point they are called biometric templates. In other words, biometric features are not considered a biometric template unless they are stored for reference [IS22].

[ED20].[3] The Board refers to "the biometric samples" by raw data, which are considered biometric data components; but are not afforded the same protection. More confusingly, at the end of the same paragraph, the Board states that "the controller must also delete biometric data and templates" [ED20, 21].[4] Until here, we have seen that biometric data do not systematically cover the images of biometric characteristics or samples. However, the Board here also distinguishes between biometric data and templates. Does this mean that only the biometric features can be considered biometric data?

All in all, the legal guidance's take on the issue seems to be rather simplistic or reductionist; it views biometric data processing as only a two-step process: the collection and its transformation into a template [ED22,7]. However, a closer examination in the next section indicates that biometric data processing can be more perplexing than such reductionist explanations in light of the provisions and the overall rationale of the GDPR.

3 The margins of the sensitive data regime under Article 9

Under the EDPB Guidelines, three criteria must be considered to ascertain whether the processing constitutes biometric data processing: (i) nature of data; (ii) means and way of processing; (iii) purpose of processing [ED20]. However, the means and ways of the processing, i.e., specific technical processing, have created confusion as to what falls under the sensitive data regime. Can we still provide legal clarity and a more efficient data protection framework by interpreting the other two criteria? Perhaps, one should look beyond the current technical and legal definitions of biometric data and interpret Article 9 and Recital 51 in the broader context of the GDPR to address the confusing *status quo*. This paper first scrutinizes the importance of the nature of the data under the sensitive data regime. Second, it inquires about the meaning of purpose in the GDPR and how it can help us understand the margins of the sensitive data regime.

3.1 The nature of the data

Article 9 has two consecutive components to invoke the sensitive data regime: (i) the existence of personal data and (ii) revealing sensitive aspects. For the first component, Recital 26 of the GDPR provides the identifiability threshold for any information to be deemed as personal data:

"*To determine whether a natural person is identifiable, account should be taken of all the means reasonably likely to be used, such as singling out, either by the controller or by another person to identify the natural person directly or indirectly.*

[3] "[…]it may not always be easy to read a biometric template without the knowledge of how it was programmed, whereas raw data will be the building blocks of any template. „

[4] "The controller must also delete biometric data and tempeltest in the event of unauthorized access to the read-comparison terminal or storage server […]"

To ascertain whether means are reasonably likely to be used to identify the natural person, account should be taken of ***all objective factors****, such as the costs of and the amount of time required for identification,* ***taking into consideration the available technology at the time of the processing and technological developments.***"

The case law of the CJEU confirms that identifiability refers to direct or indirect links between the identifier data, e.g., IP address, and one or more factors relating to a person's physical, physiological, mental, economic, cultural, and social identity.[5] In *the Breyer* case concerning dynamic IP addresses, the CJEU emphasized that the data's objective nature is considered the primary element when determining whether the data are considered personal data[CJ16]. Internet service providers (ISPs) can link Internet Protocol (IP) addresses to the names of their subscribers and the stored information related to their use of certain websites and files of particular dates and times. Although such stored data do not directly identify the individuals, if other personal data, e.g., name, is provided by any other means, e.g., by ISPs, the operators of the websites can also identify their visitors.[6]

Under the current system, the threshold for identifiability for biometric data can be invoked only if there is an already identified individual under the GDPR [Ja16]. Therefore, the images of biometric characteristics alone might not always directly identify individuals; however, if additional personal data are matched with them, they may become identifiable.[7]

As a matter of fact, for a biometric verification system to operate, there is often no need for additional personal data such as one's name or address; instead, the systems typically match the images with the biometric templates. Nevertheless, in such systems, verification usually occurs after the identification of the persons, e,g., the creation of customer profiles or digital identity. Thus, there is often an already identified individual before the biometric recognition system operates.

For identification purposes, there is always a need for additional personal data, as in the CW case. What should be noticed here is that in both cases, i.e., verification and identification, the digital images of biometric characteristics can qualify as personal data, not automatically by their nature but almost as if they are pseudonymous data that are intrinsically sensitive. Recital 26 of the GDPR states:

"*Personal data which have undergone pseudonymisation, which could be attributed to a natural person by the use of additional information should be considered to be information on an identifiable natural person.*"

[5] Judgement of 24 November 2011. Scarlet Extended (C-70/10, EU:C:2011:771).

[6] For a more detailed discussion of the issue of identifiability see: Purtova N. (2022) Purtova, N: From knowing by name to targeting: the meaning of identification under the GDPR. International Data Privacy Law. 2022 Jun 21.

[7] It should be noted that the characteristics themselves are not personal data that are the body parts, the natural source of the biometric data.

The sensitivity and value of the unique identifiers in a system typically depend on other personal data to be linked, e.g., names or metadata. However, concerning the first component of Article 9, the images of biometric characteristics can themselves generate personal data depending on the circumstances, which will be further explained below.

As per the second component of Article 9 regime, biometric data are considered highly sensitive due to their inherent characteristics enabling the controllers to identify individuals uniquely. However, their sensitivity is contingent upon the available technologies. For instance, in the *S Marper & UK case* (2008), the ECHtR ruled on the sensitive nature of DNA information but did not regard fingerprints as sensitive as DNA [EC08]. This ruling today would not be relevant as technological advances have demonstrated that it is possible to derive information related to ethnicity or illnesses from fingerprints [ZZ17]. This points to the tremendous pace of developments in biometrics and thus the change in the sensitivity of the characteristics over time.

As explained per the Recital 26 GDPR, all possible means should be taken into account when evaluating whether the data in question is personal data. I tend to argue that a similar approach can also be used to interpret the images of biometric characteristics sensitivity [Ja16], [ED20,18]. Taking into consideration the available technologies of today, to conduct a reverse image search, uploading an image or pasting a URL would be enough to find an image online and associate it with other metadata, meaning that additional information regarding a person can be reached easily. Facial images of persons can reveal highly sensitive information about people, such as their ethnicity, sexual orientation [Wa18], political views [Ko21], or health. Hence, in the CW case, facial image processing should fall under the sensitive data regime, considering that they enable the identification of a person. This would also align with the primary purpose of the data protection legislation: protecting the rights and freedoms of natural persons.[8]

For the other biometric characteristics, the answer is more complex. Most biometric characteristics reveal additional sensitive data, such as ethnicity and health information. For example, voice recordings can reveal health data and fall under special categories [Hi20]. Palmprints may reveal a person's ethnicity, gender, or age [Da19]. Similarly, vein patterns might be affected by temperature, physical activity, aging, and diseases, which implies that personal data relating to one's age, lifestyle, and health can also be inferred from veins. Iris [BRS15] and even soft biometric characteristics, behavioral patterns such as gait reveal similar personal data [Ha19]. These personal data may still be subject to the sensitive data regime [ED22], [EC08] as long as they meet the low threshold of identifiability in the GDPR, which provides that all the means reasonably likely to be used should be taken into account to ascertain whether a natural person is identifiable (Recital 26). Nevertheless, vein patterns or iris of an individual usually do not reveal a person's identity unless there are additional personal data is available. Therefore, most biometric characteristics and their digital representations alone may not be considered personal data in most cases. Yet, this should be evaluated considering all objective factors. For instance,

[8] Article 1(2) GDPR.

unique traits such as skin color, tattoos, or size can identify a person in a small sample set. As a corollary, determining whether the sensitive data regime applies to the digital representation of biometric characteristics requires the controller to take into account the nature and sensitivity of the data in question along with the specific circumstances applicable.

3.2 Systemic interpretation of the GDPR: the controllers' purpose and accountability

Pursuant to Article 9, the criterion of purpose requires that biometric data be processed to identify a natural person uniquely. Since there is confusion regarding what biometric data legally refer to, I prefer defining the purpose as 'biometric recognition' [Ja16]. This provides a broad space to interpret the purpose to the extent that the sensitive data regime covers the images of biometric characteristics. However, it is unknown at what point the purpose begins and thus triggers the material scope of the sensitive data regime. Answering this question can help us understand the boundaries of the sensitive data regime.

Where there is confusion as to the interpretation of a legal provision, one should consider the logic of the entire legislative framework, taking into consideration the normative context and other related norms within the same framework [It09]. The GDPR is a risk-based legal accountability framework, which stipulates controllers to comply with the Regulation and demonstrate that they are so. It requires them to self-regulate prior to the processing and carefully contemplate the consequences of the processing on natural persons. It means that a controller has to be aware of the consequences of any action that might lead to biometric data processing.

The definition of the controller reads as the entity which *decides the purposes and means* of the processing. The purpose here grammatically refers cumulatively to (i) 'interest' and (ii)'finality' [Va16].

Interest is subjective, referring to the personal interest of the controller. One should ask 'why' the processing of images takes place. [9] Let us imagine that CW was interested in scraping the images for biometric recognition purposes. In light of the meaning of the purpose above, this very processing should be considered biometric data processing because the interest starts before the processing when the CW determines to process the images for biometric recognition purposes. Therefore, this processing would fall under Article 9.

Finality, on the other hand, points to an objective understanding of anticipation of a final result. The question here should be, what is the expectation in scraping the images? While the anticipation cannot be easily assessed, the finality of the processing can be assessed by the DPAs retroactively based on the evidence they gather regarding the biometric data

[9] The answer of this question cannot be however general, e.g., economic interests because the purpose should be defined as specific as possible in accordance with the purpose limitation principle.

processing. Unfortunately, we do not see such reasoning in the DPA decisions; they seem to consider only the templates, e.g., database vectors, as biometric data. For instance, Information Commisioner's Office (ICO) decision states:

"*The images, metadata and URLs that are held in the Clearview Database constitute personal data. In particular: (a) an image of an identifiable individual, held in the Clearview Database, would constitute personal data about that individual; and (b) any metadata and URLs associated with such an image would likewise constitute personal data about the individual in question. Further, the Database Vectors derived from any such images would constitute special category data within the meaning of Article 9(1) GDPR and UK GDPR (since the Database Vectors would constitute biometric data falling within Article 9(1))*" [IC22].

French CNIL reaches a similar conclusion: "*The image of the individual photographed or filmed constitutes personal data as soon as the individual is identifiable, i.e., they can be recognized. In addition, this image can be compared (by automated or non-automatic means) with an image held elsewhere and attached to an identified individual and the identity of that individual can be inferred. The company* ***also*** *processes biometric data associated with such images*" [CN21]. Unfortunately, the decisions ignore the controller's initial purpose to process the photographs.

As found by the Swedish DPA in the CW case, another party can still unlawfully use this database for biometric recognition purposes [Sw21]. This begs another question: what would happen if CW was only interested in scraping the images but not for biometric recognition? A possible answer can be that because CW does not have an interest and finality aiming at biometric recognition, CW would be responsible only for the unlawful processing of the images, not for the biometric data processing, [10] unless CW enabled this unlawful processing.

Let us now imagine that while initially not having interest or finality in biometric recognition, CW decided to use the database for biometric recognition purposes ten years after the images were scraped. When the decision is made, the facial images should be considered biometric data since the interest and finality emerge with the decision. Where the interest emerges subjectively, the controllers' accountability starts; however, without the processing, no liability occurs at this stage. By this stage, the controller shall contemplate how to remain accountable for the processing of biometric data.

This interpretation can provide the first step to determining whether biometric data processing has taken place and if it falls under the sensitive data regime. Other relevant provisions forming the accountability framework of the GDPR, i,e., the principle of data protection by design and default (Article 25) and data protection impact assessment requirement (Article 35) points to a similar conclusion: controllers should clarify their interest and finality in processing biometric characteristics. This assessment should be done for the whole cycle, starting from the emergence of interest to the virtual deployment

[10] Without prejudice to other obligations, such as security of the processing.

of biometric technologies.

4 Synthesis of the discussion: the nature over the purpose

As explained, controllers' accountability for biometric data processing, chronologically, starts with the emergence of the purpose of the controller deploying a biometric recognition system, not with the execution of the means, e.g., feature extraction. Nevertheless, the over-reliance on purpose in biometric data protection is still precarious since the data stored in centralized databases can be later used for other purposes [Ki13]. Morevover, the purpose may change at any stage of the processing. Therefore, it is challenging, if not impossible, to know the interest before the processing for anyone except for the controllers themselves.

Although the facial images, and the images of other characteristics, are systematically not considered biometric data under the GDPR, they are still capable of revealing sensitive aspects of natural persons, as discussed above. Therefore, in most cases the sensitive data regime can be invoked by the nature of the personal data, regardless of whether biometric data processing is the purpose of the initial processing. When assessing whether the data in question is sensitive, the focus should be on objectively what kind of information the characteristics could reveal. I argue that the same objective perspective of Recital 26 and *the Breyer* should be valid for the threshold for the images of biometric characteristics. Following the objective approach, such sensitivity highly depends on the relevant technological developments.

5 Conclusion

The GDPR and EDPB Guidelines (and their interpretation in CW decisions) afford biometric templates additional protection compared to the biometric samples, i.e., the images of biometric characteristics. The paper suggested that the first step to understanding whether biometric images fall under the sensitive data regime, and thus are afforded the same protection as biometric data should be clarifying the nature and sensitivity of the data in question, considering the technological realities.

The second step analyzes the controllers' purpose with a systemic approach, concluding that biometric data processing can be in question depending on the controller's interest and finality, which might change at any time after the initial processing. While the purpose is a significant determinant in the accountability framework, proving the controller's purpose is not easy unless a risk has already been materialized. The paper considers that when considering whether processing of the images of biometric characteristics falls under the sensitive data regime, the nature and sensitivity of the personal data should prevail over the controller's purpose as it provides a more objective criterion for the sensitive data regime.

Acknowledgments. I would like to thank Prof. Els Kindt for her valuable opinions on the draft of this study. Also, I sincerely thank Steven Leenaerts for his editing skills.

References

[BRS15] Bansal, A, Ravinder, and Sharma R. K.: Determining diabetes using iris recognition system. International journal of diabetes in developing countries 35.4, 2015.

[CN21] CNIL: Decision n° MED 2021-134 of 1st November 2021 issuing an order to comply to the company CLEARVIEW AI (No. MDMM211166) 2021.

[CJ16] Court of Justice of the European Union: Judgment of 19 October 2016. Breyer, C-582/14, ECLI:EU:C:2017:994, 2016.

[Da19] Damak W et al. Palm Vein Age and Gender Estimation Using Center Symmetric-Local Binary Pattern. In: Martínez Álvarez F., Troncoso Lora A., Sáez Muñoz J., Quintián H., Corchado E. (eds) International Joint Conference: 12th International Conference on Computational Intelligence in Security for Information Systems, 2019.

[EC08] S Marper v United Kingdom 2008 European Court of Human Rights 1581 2008.

[ED20] European Data Protection Board: Guidelines 3/2019 on processing of personal data through video devices, Version 2.0, 2020.

[ED22] European Data Protection Board: Guidelines 05/2022 on the use of facial recognition technology in the area of law enforcement, 2022.

[Ha19] Hamme T. et al. : A systematic comparison of age and gender prediction on imu sensor-based gait traces. Sensors 19.13, 2945, 2019.

[GD22] GDPD: Ordinanza ingiunzione nei confronti di Clearview AI - 10 febbraio 2022 [9751362] 2022.

[Hi20] Higuchi M et al. Effectiveness of a Voice-Based Mental Health Evaluation System for Mobile Devices:Prospective Study JMIR Form Res;4(7):e16455, 2020.

[It09] Itzcovich, G: The interpretation of community law by the European Court of Justice. German Law Journal 10 (5) 537-560, 2009.

[Ja16] Jasserand, C: Legal nature of biometric data: From generic personal data to sensitive data. Eur. Data Prot. L. Rev. 2, 297, 2016.

[Ja22] Jasserand, C.: Cleraview AI: Illegally collecting and selling our faces in total impunity? (Part I) https://www.law.kuleuven.be/citip/blog/clearview-ai-illegally-collecting-and-selling-our-faces-in-total-impunity-part-i/ 2022 (20.07.2022).

[Ki13] Kindt, E J. Privacy and data protection issues of biometric applications. Vol. 1. Springer, 2013.

[Ki18] Kindt, E J. :Having yes, using no? About the new legal regime for biometric data. Computer law & security review 34.3, 523-538, 2018.

[Ko21] Kosinski, M: Facial recognition technology can expose political orientation from

naturalistic facial images, Scientific Reports (open access), www.nature.com/scientific reports 11:100, 2021.

[IC22] ICO: Clearview AI inc. Enforcement Notice, 2022 <https://ico.org.uk/media/action-weve-taken/enforcement-notices/4020437/clearview-ai-inc-en-20220518.pdf> 2022.

[IS22] ISO/IEC 2382-37: 2022 Information technology-Vocabulary- Part 37: Biometrics.

[Va16] Van Alsenoy, B: Regulating data protection: the allocation of responsibility and risk among actors involved in personal data processing. KU Leuven, CiTiP 2016.

[Sw21] Swedish DPA: Police unlawfully used facial recognition app, 12 February 2021 <https://edpb.europa.eu/news/national-news/2021/swedish-dpa-police-unlawfully-used-facial-recognition-app_en> (20.07.2022).

[Pu22] Purtova, N: From knowing by name to targeting: the meaning of identification under the GDPR. International Data Privacy Law. 2022 Jun 21.

[WK18] Wang, Y & Kosinski, M: Deep Neural Networks Are More Accurate Than Humans at Detecting Sexual Orientation from Facial Images, 114 Journal of personality and Soc. PSYCHOL. 246, 254-56, 2018.

[ZZ17] Zhou Z and Zare N. R: Personal information from Latent Fingerprints using Desorption Electrospray Ionization Mass Spectometry and Machine Learning, Anal.Chem. 89,2, 1369-1372, 2017.