Research, the GDPR, and Mega Biometric Training Datasets: Opening the Pandora Box[1]

Catherine Jasserand
KU Leuven
CiTiP
Leuven, Belgium
catherine.jasserand@kuleuven.be

2022 International Conference of the Biometrics Special Interest Group (BIOSIG) | 978-1-6654-7666-9/22/$31.00 ©2022 IEEE | DOI: 10.1109/BIOSIG55365.2022.9897040

***Abstract*—This paper sheds light on the challenges that the EU data protection pose to the constitution and the use of large-scale biometric datasets for research purposes. The paper analyses the GDPR (General Data Protection Regulation) rules applicable to research and explains their inadequacy for the constitution of large biometric datasets developed to train and test biometric recognition models. Researchers need a considerable amount of data to train and test their models. These data cannot be realistically obtained with the consent of individuals to whom the data relate. But can researchers collect them from publicly available sources, such as social media or open platforms, to create large biometric datasets? Under the current GDPR rules, it seems rather difficult for researchers located in the EU to do so. A thorough analysis of the provisions will explain the challenges that researchers face. In that context, one could have hoped that the Data Governance Act (DGA) would have brought some opportunities for research when it aimed at promoting *data altruism*. But it seems that the DGA is a missed opportunity to open up personal data for research purposes. As acknowledged by some companies, these regulatory impediments might force researchers to train their models outside the EU with datasets not subject to the EU data protection rules.**

Keywords—biometric data, research, mega training datasets, GDPR.

I. INTRODUCTION

Several large-scale face datasets constituted outside the EU for research purposes made the headlines a few years ago. The public discovered that pictures that they had published or that were published by third parties (friends, colleagues) on various social media had been re-used by researchers to train facial recognition algorithms. In 2019, the New York Times revealed the existence of MegaFace, a massive dataset composed of face photographs scraped from a photo platform by a team of researchers from the University of Washington [1]. The article explained that children's pictures were among those scrapped and that commercial applications had been developed using that dataset, although MegaFace was reserved to research only. Soon after the article's publication, the team of researchers deactivated the dataset. MegaFace had been constituted with photographs crawled from the photo platform Flickr. Only pictures released under a Creative Commons license had been collected. These licenses allow certain re-uses of protected materials. But even if they can be very permissible – such as allowing commercial re-use, they are not meant to allow the use of personal data contained in these images.

The scope of Creative Commons licenses is limited to copyright issues [2]. After the attention that MegaFace received, another dataset composed of celebrities' images, MS-Celeb-1M, was *quietly removed* from the Internet [3]. Although this dataset was mainly composed of US and UK actors, the term 'celebrity' was loosely used to scrape images from journalists and other public figures (such as policymakers or academics). Despite the deletion of the dataset, it was reported that Chinese authorities had used it to track foreign journalists [4]. Since 2017, two researchers, Adam Harvey and Jules Laplace, have done investigative journalism on the use of Flickr photographs to constitute face training datasets [5]. These datasets are not limited to MegaFace and MS-Celeb-1M. They have identified, at least, 13 datasets [5].

These datasets have in common to have been constituted with images released under a Creative Commons license or publicly accessible on social media or other platforms. But as this paper explains, the scope of Creative Commons licenses does not extend to personal data (they only cover copyrightable content). Images that are publicly accessible on the Internet are not necessarily re-usable. From an EU data protection perspective, there is no such thing as freely re-usable personal data (even for research purposes). As explained, photographs showing a face of an identifiable individual are personal data. As such, their collection and processing are subject to data protection rules governed by the General Data Protection Regulation (GDPR). Yet, the GDPR does not provide a research regime for collecting and processing personal data for research purposes. But it provides some specific rules (including the possibility for Member States to allow the processing of sensitive data for research purposes). However, as it will be demonstrated, the rules of the GDPR are complex and do not allow scraping the Internet (i.e. existing platforms) to constitute large-scale face datasets.

II. THE COMPLEX NATURE OF PHOTOGRAPHS

Photographs displaying faces of human beings can be both the object of copyright protection (under the condition that they reach the threshold of originality as defined in national copyright law) and be protected as personal data under data protection legislation.

[1] The title is inspired by one of the issues raised by Richard Van Noorden in 'the ethical questions that haunt facial-recognition research', published in Nature on 18 November 2020.

A. Photographs as copyrightable works

In many jurisdictions, photographs are considered copyrightable works, i.e. they are protected by copyright law if they reach the threshold of originality. What originality means varies from country to country. In civil law countries (such as France), a work is original if it is the author's own intellectual creation. Whereas in common law countries (such as the USA), a work must show a minimum of creativity to be protected by copyright. As a consequence, depending on the country of reference, a photograph might or might not reach the threshold of copyrightability to be protected as a copyrighted work.

The discussion about the threshold of copyrightability is not important for the scope of the paper. However, it matters to understand that if protected, photographs can be licensed to allow others to re-use them. Only the copyright holder (i.e. the person or company that holds the exclusive rights) can license the work to a third party. The copyright holder of a photograph published on social media or photo platforms, such as Flickr, will most likely be the image's photographer. He or she will be able to authorize their re-use. However, in most cases, he or she will not be the person portrayed in the photographs. Thus, the authorization that a photographer gives to re-use his or her works should not be interpreted as the consent given by the individuals represented in these images to re-use these works from a data protection perspective.

Yet there seems to be a general confusion among researchers who consider that pictures released under a Creative Commons license as freely re-usable. Even though these images are free from copyright restrictions, they are not freely re-usable from a data protection perspective. This is a significant misconception about the scope of Creative Commons licenses. It is not uncommon to read that large-scale face (or voice) datasets have been constituted with data released under permissible Creative Commons licenses or other permissive licenses [6]. As the former CEO of Creative Commons acknowledged, "CC licences were designed to address a specific constraint, which they do very well: unlocking restrictive copyright. But copyright is not a good tool to protect individual privacy" [7].

B. Photographs as personal and biometric data

From a data protection perspective, facial images (photographs) and audiovisual excerpts from which facial images can be extracted are personal data, provided they relate to an identifiable or identified individual (Article 4(1) GDPR)[2]. Identifying someone in a data protection context does not mean establishing that person's identity or knowing his or her name. It only means being able to single out that person, i.e. distinguishing him or her from a group. Thus, an individual will be identifiable in a photograph depending on the quality of the image and other technical factors (such as light, exposure, or resolution). Not every image might reach the threshold of identifiability and be considered personal data [8].

Besides, the GDPR has introduced the legal notion of biometric data in the data protection landscape (Article 4(14) GDPR). To qualify as biometric data, the data at stake must first meet the criteria applicable to personal data (i.e. to reach the threshold of identifiability, as described above). Then, they must result from the *technical processing* of biometric characteristics (described as 'biological, physiological, and behavioral characteristics'). Finally, they are processed to either 'allow or confirm the unique identification' of an individual. Legal scholars have debated the meanings of 'technical processing' (whether it includes biometric samples) and 'unique identification' (whether it refers to the identification modality only or the ability to single out individuals thanks to the processing of their 'unique' biometric characteristics in the context of an identification or verification modality) [e.g. 9, 10, 11]. In its draft Guidelines on facial recognition use in law enforcement, the European Data Protection Board, an EU advisory body, seems to consider that 'unique identification' refers to both the identification and verification modalities [12]. Biometric data processed for classification purposes, such as age, ethnicity, or gender, would be excluded from the regulatory definition of biometric data. The purpose of that processing is not to single out someone using his or her biometric characteristics but to find common attributes to a group of individuals.

According to Article 9(1) GDPR, biometric data processed 'to uniquely identify' someone are not only personal data but also sensitive data. As such, they cannot be processed unless an exception listed in Article 9(2) GDPR applies. These exceptions are discussed in the next section. Concerning the status of photographs, the GDPR further specifies in Recital 51 (a non-binding provision) that photographs should only be considered biometric data when they result from technical processing for either identification or verification purposes. The recital does not refer to verification but to authentication, which is often used as a synonym for verification. Based on this recital, some scholars have rightly pointed out that photographs collected to create a face dataset should not be considered biometric data as the pictures do not undergo any technical processing [10]. By contrast, photographs processed to create biometric templates should be regarded as biometric and, thus, sensitive data. One could infer that creating a face dataset with 'mere' photographs would process ordinary personal data, whereas using a face dataset to generate biometric templates would result in processing biometric and sensitive data.

III. Legal Grounds for Processing

There is no such thing as freely re-usable personal data. Images that can be found on the Internet, either released under a Creative Commons license (or another permissive copyright license) or voluntarily disclosed by the individuals portrayed in the images, are not freely re-usable. But the GDPR acknowledges the existence of publicly accessible sources (such as information collected by the public sector, still subject to data protection rules, though) and data manifestly made public by the data subjects themselves. The legal grounds for processing are discussed below.

A. Non-suitability of consent as a legal ground

Ordinary personal data can be processed based on the consent of the individual to whom they relate (Article 6(1)(a) GDPR). Consent must be freely given, unambiguous and informed to be valid. It must also be explicit, i.e. it results from a clear statement (such as a written statement) when it serves as a legal basis to process sensitive data (Article 9(2)(a) GDPR). To constitute large-scale datasets of facial images, whether these images are ordinary personal data or sensitive

[2] Regulation (EU) 2016/679 or General Data Protection Regulation, 2016.

data, it does not seem realistic and feasible to obtain consent from tens of thousands, hundreds of thousands (or even more) individuals to use their images for research purposes.

B. Other legal grounds to process non-sensitive data

As mentioned, mere photographs will not be considered biometric data and will be processed as ordinary personal data if they do not reveal sensitive information. As a consequence, they can only be processed following one of the six legal grounds provided by Article 6 GDPR. Besides consent, one could try to rely on either the performance of a task carried out in the public interest (Article 6(1)(e) GDPR) or the legitimate interests of the controller (Article 6(1)(f) GDPR).

1) Performance of a task in the public interest

This legal ground applies to controllers performing a task in the public interest as laid down by law. These controllers will mainly be public authorities, but they can also be private entities vested with the same official authority or performing a task in the public interest. Research should first be identified in (national) legislation as a task carried out in the public interest. The former advisory body to the EU Commission, the Article 29 Working Party (Article 29 WP), called for "a strict interpretation and a clear identification, on a case-by-case basis, of the public interest at stake and the official authority justifying the processing" [13]. As a result, in some countries, universities might be included in the list of official authorities and research be identified as a task in the public interest. However, there is no indication or guidance that scraping images from social media or other platforms to create mega training datasets of face images would be considered necessary for the performance of a task in the public interest.

2) Legitimate interests of the controller

Controllers often invoke this legal ground, but it is difficult to apply it due to the requirements it must satisfy. First, this legal ground is not available to public authorities performing a task in the public interest (as explained in Article 6(1) GDPR). Second, applying this legal basis requires a case-by-case analysis based on three elements. First, the controller must prove the existence of a legitimate interest. There is no list of what constitutes a legitimate interest (although the GDPR mentions fraud prevention as an example in Recital 47 GDPR). Under the previous data protection regime, the Article 29 WP identified scientific research as a possible legitimate interest due to its benefits to society [13].

However, there is currently no indication that the constitution of a training dataset of face images scraped from online sources would constitute such a legitimate interest. It could also be argued that some European bodies seem opposed to this idea as they suggest using synthetic data (instead of real data) to train AI models [14] or develop facial recognition algorithms [15]. But even if such a purpose would be considered a legitimate interest, researchers should assess whether such a purpose is necessary, i.e. they should demonstrate that there are no less intrusive alternative solutions to reach the same goal. Besides, they should also assess the impact of the processing on data subjects, which includes the individuals' reasonable expectations concerning the use of their images. If one could argue that researchers need millions of images to train their facial recognition algorithms, one would object that individuals publishing their images on social media could expect these images to be re-used for research purposes, including to train biometric recognition algorithms. This legal ground seems very difficult to be relied upon.

C. Other legal grounds to process sensitive data

As explained, photographs scraped from the Internet will not be considered biometric data per Recital 51 GDPR. But they could still be sensitive data if they reveal sensitive information, such as religious beliefs or political opinions (someone wearing a religious symbol or present at a political meeting) or health condition (someone showing signs of illness on their face). Article 9(1) GDPR provides an exhaustive list of categories of personal data that fall in the category of sensitive data (officially named 'special categories of personal data'). It could be that only a minority of images would be considered sensitive data. At the same time, researchers are advised to train their algorithms on more diverse data (including data representing various ethnicities) [15]. This recommendation should have an impact on the number of images in a dataset that could be considered sensitive data.

Besides explicit consent, Article 9(2) GDPR allows the processing of sensitive data if the data have been manifestly made public by the data subjects themselves (Article 9(2)(e) GDPR) or if based on a research exception allowed by national law (Article 9(2)(j) GDPR).

1) Data manifestly made public by the data subjects themselves

When sensitive data are voluntarily disclosed by the data subjects themselves, they should no longer benefit from the extra protection granted to sensitive data. One could say that the data at stake are retrograded to the status of ordinary personal data. Their processing should, therefore, still comply with one of the legal grounds of Article 6 (1) GDPR (applicable to non-sensitive personal data). Yet, as described in the previous sub-section, finding the appropriate legal basis to harvest the photographs from social media or other platforms might be challenging.

As interpreted by the Article 29 WP, the data subjects' intention should be narrowly interpreted as they should be aware that their data will be made available to everyone [16]. Interestingly, the EDPB, which has replaced the Article 29 WP, considers that if facial images have been manifestly made public by the data subjects themselves, this does not extend to their processing as biometric data. In other words, the biometric templates generated from these images will not be considered as being manifestly made public by the data subjects themselves [12]. A different legal ground will be necessary to process them.

2) Research exception

In application of Article 9(1)(j) GDPR, sensitive data can be processed based on law as long as their processing is necessary for research purposes and that safeguards are in place (as defined in Article 89(1) GDPR). If the term research is not defined in the GDPR, it includes commercial and non-commercial research whether publicly or privately funded (based on the examples provided by Recital 159 GDPR). In case such an exception is set out by law, the controller (i.e. the researcher) will also have to demonstrate there is no other alternative than collecting such a huge amount of images to train the facial recognition algorithms. It might be difficult to make such an argument as several European bodies seem to

believe that synthetic data could be an alternative to the constitution of massive biometric training datasets [14, 15].

IV. OTHER LEGAL CONSIDERATIONS BEYOND THE LEGAL GROUNDS FOR PROCESSING

Besides finding the suitable legal basis to allow researchers to constitute these large-scale datasets, other legal considerations should be taken into consideration. First, independently of the applicable legal basis, researchers will have to implement safeguards while processing personal (and a fortiori sensitive) data for research purposes. Second, data subjects have specific rights, which might hamper researchers from collecting personal data at large-scale.

A. The difficult implementation of safeguards

Besides, personal and sensitive data processed for research purposes must comply with Article 89(1) GDPR. Following that article, processing for research purposes should be subject to safeguards. Without specifying what constitutes an appropriate safeguard or providing criteria to assess such a safeguard, the GDPR refers to pseudonymization and data encryption (i.e. a security measure) as adequate safeguards in the context of research.

As defined in Article 4(5) GDPR, pseudonymized data are personal data that cannot be attributed (anymore) to a data subject without using additional information, which is kept separate from the data. This implies the possibility of separating identifying information from the rest of the data. But as discussed in [17], facial images do not need extra information (such as the individual's name or identity) to single out an individual. Thus, it cannot be claimed that pseudonymizing a facial image is a matter of "delinking an image from its associated data". Instead, pseudonymizing means being able to alter or remove identifying information from the content of an image. The identifying information should be understood as the physical characteristics that make a face recognizable or that allow the identifiability of a person (i.e. to single out him or her). One could argue that such a transformation of the data might render them unusable for biometric recognition purposes. So, it might be challenging to preserve the utility of the data for biometric recognition purposes while exploring the feasibility of pseudonymizing facial images.

B. Data subjects' rights

1) Right to information

The purpose of the right to information is to enable individuals to know that personal data relating to them have been collected. It is an essential right as it triggers the application of all data subjects' rights (including the right to remedy). But there are exceptions to this right, including in the context of processing for research purposes. Article 13 GDPR details information to be provided when data have been directly collected from individuals. Article 14 GDPR covers the situation when personal data were indirectly obtained, which is the case at stake.

Following Article 14 GDPR, individuals whose personal data have been collected through third parties should be informed about it. The article provides the list of information to be provided, together with exceptions that lift the obligation of information. Interestingly, Article 14 does not apply in case "such information proves impossible or would involve a disproportionate effort, in particular for processing for…scientific or historical research purposes" if the safeguards of Article 89(1) GDPR are in place (Article 14(5)(b) GDPR). Thus, the exception applies when it is impossible to deliver the information or doing so would require a disproportionate effort. [18] According to the Article 29 WP, the controller should prove the impossibility of providing the information, whereas it should assess the disproportionate effort to deliver the information based on identified factors (the number of subjects involved, their age, and the existing safeguards) [19]. Interestingly, the French Council of State considered in 2014 that informing 25 million persons that their personal data had been collected did not constitute a disproportionate effort as long as the controller (Business Telephone Directories in the case) had their contact details. The application of this exception depends on the context of the case [20].

Concerning the massive collection of images, the researchers could try to argue that if the images relate to individuals who can be singled out (i.e. identifiable), in most cases, they do not have any means to contact these individuals as the photographs are not necessarily associated with a name or an email address. They might find another argument based on Article 11 GDPR, which does not oblige a controller to maintain, acquire or process additional information to identify the data subject for the sole purpose of complying with the GDPR. But data subjects themselves may provide additional information to allow their identification. Although there is not much literature on the interpretation of Article 11 GDPR [21], the article seems to relate to personal data about *identifiable* (but not identified) individuals [22] and thus lifts certain obligations for this type of data.

2) Right to object

Data subjects have the right to object to the collection of their personal data in specific circumstances: when their data are collected based on a task performed in the public interest (Article 6(1)(e) GDPR) or following the legitimate interests ground (Article 6(1)(f) GDPR). As explained in section 3.2, these two legal grounds are the ones that seem the most appropriate for the constitution of large-scale datasets (even if their application is challenging).

Article 21 (6) GDPR further limits the right to object when personal data are processed for research purposes in compliance with Article 89(1) GDPR. In that case, data subjects cannot object to the processing based on the performance of a task in the public interest (Article 6(1) (e) GDPR). They can only object the processing of it is based on the legitimate interests ground unless the controller demonstrates a compelling legitimate interest that would override the data subject's right (Article 21(1) GDPR, Recital 69 GDPR). The GDPR does not specify what a compelling legitimate interest is, but according to the Article 29 WP, scientific research could be such an interest [13], as cited in [22]. But there is no guidance from EU advisory bodies that the constitution of large-scale datasets for research purposes would still constitute a *compelling* legitimate interest (see discussions in section III.B).

C. Beyond data protection issues, the narrow concept of data altruism

In May 2022, the EU legislators adopted the Data Governance Act (DGA), a new regulatory framework aimed at facilitating data sharing and ensuring better data access (including for the data held by the public sector). The DGA is

part of the Data Strategy published by the European Commission in 2020 where the Commission highlighted the necessity of available data "for training artificial intelligence systems, with products and services rapidly moving from pattern recognition and insight generation to more sophisticated forecasting techniques and, thus, better decisions" [24].

The DGA covers personal and non-personal data. It has introduced the concept of 'data altruism'. According to Recital 45 DGA, the objective of data altruism is, among others, to "contribute to the emergence of sufficiently-sized data pools made available on the basis of data altruism in order to enable data analytics and machine learning, across the Union." In application of Article 2(6) DGA, personal data can be voluntarily shared based on individuals' consent whereas non-personal data can be shared with the permission of data holders without compensation. As criticized by certain scholars [25], the DGA does not enable individuals to make their personal data available for the public good. It also adds extra burdens, on top of the existing GDPR rules, for organizations interested in these data donations. It is thus a missed opportunity to create exceptions for research and allow individuals to make their data available for research purposes.

V. Synthesis

This paper has addressed a recurring issue, whether the current GDPR rules allow researchers to get access to the vast amount of personal data they need to train and test their biometric models and constitute large-scale datasets. It clarified a common misunderstanding concerning publicly available data, which are not freely re-usable from a data protection perspective. It has also explained the complex nature of photographs and their legal status (making a distinction between 'mere' photographs and photographs technically processed for biometric recognition).

Reviewing the legal grounds applicable to personal data, the paper identified *the performance of a task in the public interest* and *legitimate interests* as possible legal bases to process the images. But both legal bases are challenging: the first one needs to be laid down in law and research must be identified as a task in the public interest; the second one requires a balance of interests (which includes the reasonable expectations of the individual concerning the use of his or her data). For the reminder of images that would qualify as sensitive data, the paper analyzed the suitable of two legal exceptions: data manifestly made public by the data subject and the research exception. The first legal ground revolves around the intention and expectation of the data subject. Yet, it seems difficult to argue that he or she has disclosed his or her data with the expectation they would be part of a large-scale dataset for research purposes. The research exception could be an option (provided it complies with different requirements). But researchers will have to demonstrate there is no less intrusive alternative to train their models. Yet, several European bodies seem to favor the use of synthetic data as an alternative to the constitution of large-scale biometric datasets (with real data). Besides the legal grounds for processing, the researchers will have to consider other issues that might constitute further obstacles (such as adopting appropriate safeguards). In short, the GDPR does not offer a ready-to-use solution. What is missing is the opportunity to 'donate' personal data for research purposes. This was one of the ideas behind the proposal for the Data Governance Act (DGA)[3] through data altruism. But the adopted text seems to have missed the objective of opening up personal data for research purposes.

VI. Conclusions

In front of the legal complexity of the GDPR, researchers might be forced to train their models outside the EU with local datasets not subject to the GDPR rules. In a report published by the French Senate in May 2022 on facial recognition, some companies reported already using their subsidiaries outside the EU to train their models with local datasets subject to fewer constraints [26]. Researchers who would use, in the EU, existing datasets constituted abroad would still be subject to the GDPR as the rules apply to controllers established or present in the EU (Article 3 GDPR). So, the solution could be to get guidance from authoritative authorities (such as the EDPB) to explain how researchers could use the potential that data represent to train and develop AI biometric systems within the boundaries of the GDPR. Besides 'donating' own data for research, the solution could be to design a special regime (through derogations from the GDPR) that would consider the collection and use of a massive amount of personal data to train and test biometric models for research purposes as a necessity in the public interest.

Acknowledgment

This paper was made possible thanks to the funding received from the European Union's Horizon 2020 research and innovation programme in the context of the individual fellowship Marie Skłodowska Curie for the DATAFACE project (under grant agreement no. 895978). The author would like to thank the three anonymous reviewers, Lydia Belkadi and Marcel Grimmer, for insightful discussions on the topic. The views expressed are solely those of the author, who retains full responsibility for any errors or omissions.

References

[1] K. Hill, and A. Krolik, "How Photos of Your Kids Are Powering Surveillance Technology", New York Times, November 10, 2019.

[2] C. Jasserand, "Free to re-use? The case of facial image scrapped from the Internet and compiled in mega research datasets", CiTiP Blog, November 17, 2020.

[3] M. Murgia, "Microsoft quietly deletes largest public face recognition data set," Financial Times, June 6, 2018.

[4] C.Healy, and D. Maye, "Punishing Journalists PRC Province's Latest Mass Surveillance Project," Won by Neusoft Powered by Huawei, IPVM, November 29, 2021.

[5] A. Harvey, and J. LaPlace, "Exposing AI", IPVM, 2021.

[6] ML Commons, "Introducing the People's Speech dataset – 30,000 + hours of diverse speech data to drive ML innovation," ML Commons, December 14, 2021.

[7] M. Murgia, "Who is using your Face? The ugly truth about facial recognition," Financial Times, September 18, 2019.

[8] Article 29 WP, "Opinion 02/2012 on Facial Recognition in Online and Mobile Services," WP192, p.4.

[9] C. Jasserand, "Legal Nature of Biometric Data: From 'Generic' Personal Data to Sensitive Data," 2(3) EDPL, 2016, pp. 297-311.

[10] E. Kindt, "Having Yes, Using No? About the New Legal Regime for Biometric Data," 34(3) Computer Law & Security Review, 2018, pp. 523-538.

[11] D. Clifford, "The Legal Limits to the Monetisation of Online Emotions," PhD thesis, 2019.

[12] EDPB, "Guidelines 05/2022 on the Use of Facial Recognition Technology in the Area of Law Enforcement," 2022.

[3] Regulation (EU) 2022/868 or Data Governance Act, OJ L152, June 2022.

[13] Article 29 WP, "Opinion 06/2014 on the Notion of Legitimate Interests of the Data Controller under Article 7 of Directive 95/46/EC," WP 217, 2014, pp. 22-26.

[14] EDPS, "Synthetic Data," 2021.

[15] Council of Europe, "Guidelines on Facial Recognition," 2021.

[16] Article 29 WP, "Opinion on Some Key Issues of the Law Enforcement Directive (EU 2016/680)," WP258, 2017, p.1.

[17] C. Jasserand, "Massive Facial Databases and the GDPR: the New Data Protection Rules applicable to Research," in Data Protection and Privacy: The Internet of Bodies, R. Leenes et al., Eds, Hart Publishing 2018, pp.169-188.

[18] G. Zanfir-Fortuna, "Article 14" in The EU General Data Protection Regulation (GDPR), C. Kuner, C. et al., Eds. Oxford University Press 2020, pp. 434-448.

[19] Article 29 WP, "Guidelines on Transparency under Regulation 2016/679," WP260rev.01, 2018, pp. 29-30.

[20] Conseil d'État (French Council of State), "Décision 353193, Pages Jaune Groupe," ECLI: FR: CESSR: 2014: 353193.20140312, March, 12 2014.

[21] L. Georgieva, "Article 11. Processing Which Does not Require Identification" in in The EU General Data Protection Regulation (GDPR), C. Kuner, C. et al., Eds. Oxford University Press 2020, pp. 391-397.

[22] D. Erdos, "Identification in Personal Dara: Authenticating the Meaning and Reach of Another Broad Concept in EU Data Protection Law" 46 Computer Law & Security Review, 2022, 105721.

[23] Article 29 WP, "Guidelines on Automated Decision-Making and Profiling for the Purposes of Regulation 2016/679," WP251rev.01, 2018, p.18.

[24] European Commission, 'Communication "A European Strategy for Data," COM (2020) 66 final, February 2020.

[25] W. Veil, "Data Altruism: How the EU is Screwing Up a Good Idea," 2021.

[26] French Senate, "Information Report on Facial Recognition," no. 627. May 10, 2022, p.68.