Is it Personal data? Solving the gordian knot of anonymisation

Alexandre Lodie, Cedric Lauradoux

Is it Personal data?
Solving the gordian knot of anonymisation

Alexandre Lodie and Cedric Lauradoux

Grenoble-Alpes University, INRIA
alexandre.lodie@inria.fr

Abstract. The concept of personal data is pivotal to understand the scope of the General Data Protection Regulation (GDPR). Since data protection regulations and directives were adopted, national courts and the Court of Justice of the European Union (CJEU) had to determine whether some data like IP addresses are personal. Courts' rulings are often based on the possibility to re-identify individuals from the datasets under dispute. The different views adopted by Courts over the years do not always reach the same conclusions, which is source of legal uncertainties. This is especially the case when data controllers are using data protection techniques like pseudonymisation and anonymisation. Recently, the ruling of the CJEU in the SRB vs EDPS case challenged the stance adopted by data protection authorities concerning the distinction between pseudonymisation and anonymisation. Data protection watchdogs consider that pseudonymized data are always personal data. The dictum of the court in the SRB vs EDPS case is that pseudonymised data can be considered as anonymised and thus non-personal data depending on the re-identification capability of the data holder. This creates legal uncertainties as the legal qualification of data that have been subject to data protection techniques. In this paper, we question the extent of the definition of personal data and how it applies to data protection techniques such as pseudonymisation and anonymisation. Eventully, we emphasise that this issue is challenging with regard to the protection of data within the EU borders and beyond.

Keywords: Personal data · Anonymisation · Pseudonymisation · Re-identification

1 Introduction

The General Data Protection Regulation (GDPR) is entered into force in 2018, and has been a real milestone for the protection of personal data in the EU. Although its material scope and guiding principles are quite similar to those of the former directive 95/46/EC - that the GDPR replaces – its nature is different since a European Regulation is directly applicable in Member States' legal systems, as provided by article 288 of the Treaty on the functioning of the European Union [4] and recalled by the Court of Justice of the European Union[1] (CJEU) [1].

The main purpose of this regulation was to create a single legal framework to regulate personal data processing within the EU borders and to avoid fragmentation with regard to the application of data protection law [5]. The principles enshrined by the text are derived from the fundamental right to protection of personal data provided by article 8 of the Charter of Fundamental rights of the EU [3]. From this background a rather inclusive definition of the notion of personal data was adopted by the GDPR. In this context personal data are defined in Article 4 of the GDPR [28] as:

> any information relating to an identified or identifiable natural person ('data subject'); an identifiable natural person 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.

[1] We refer to both the ECJ and the General court under the acronym CJEU.

To summarise, personal data are any piece of information related to an individual, be it in a direct or in an indirect manner. Data subjects do not need to be identified directly by their name to consider data related to them as personal.

The main aim of such a regulation is thus to protect the fundamental right to data protection of EU citizens. On the other hand, this human-rights-centered approach is challenged by the increasing need of sharing data for trade or research purposes and to preserve the competitiveness of industries and companies based in the EU. This materialises by the adoption of regulations aiming to promote and regulate the free-flow of non-personal data [30] or to ensure the availability and sharing of data generated by the use of a product [34]. Therefore, there are two trends concerning the regulation of data in the EU: on the one hand the will to encourage the free-flow of non-personal data and on the other, the protection of personal data, in the interest of data subjects.

A major issue is that some non-personal data are de-identified personal data. In other terms, they are data which used to be personal, but no longer are. This possibility lies in the development of de-identification techniques and in particular anonymisation. De-identification can be defined as "the process of removing personal information from a record or data set" [37]. De-identification techniques aim to improve data security and data subjects' privacy by avoiding the direct or indirect identification of data subjects from their data. De-identification may lead to anonymisation, a term used "to cover the techniques and approaches you can use in the pursuit [...] of preventing the identification of the individuals the data relates to, taking into account all relevant factors" [37].

Following this approach, anonymisation is the technical notion that bridges the gap between personal and non-personal data. Understanding anonymisation is thus critical to understand what personal data are, since anonymisation leads to the exclusion of data processing from the scope of the GDPR. The reliability of anonymisation technique is an important criterion to qualify data as personal or non-personal. The development of de-identification techniques, and more specifically anonymisation, question in a broader manner the definition of personal data under EU law.

In a first part we will try to illustrate the current debate on the definition of personal data through the prism of recent case law related to the qualification of IP addresses. Indeed, such a discussion demonstrates that the distinction between personal and non-personal data is not that clear-cut. In the second part, we emphasise that the CJEU seems to adopt a contextual approach of the definition personal data, according to which the qualification of data as personal depends on the additional information accessible by the person who holds data. This approach blurs the line between personal and non-personal (anonymised) data. This raises the risk of undermining the level of data protection within the EU as this definition conditions the application of data protection law.

2 The debate on IP addresses

The purpose of the GDPR is to regulate the processing of personal data. Despite the definition of personal data provided by Article 4 of the GDPR, it is still unclear what this concretely means from a technical perspective. DPAs and courts within the EU have been repeatedly asked to consider whether IP addresses could be regarded as personal data. To provide some background on this matter, it is needed to understand what IP addresses actually are, before assessing their legal qualification.

From a technical perspective, IP addresses are identifiers used to route information. They are assigned to terminals connected to the Internet. IP addresses allocation is globally supervised by ICANN[2] which delegates the allocation and management of pools of IP addresses to regional authorities. The regional authorities create sub-pools of IP addresses which are distributed to private and public organizations like Internet Service Providers (ISPs). When an individual purchases a subscription to an ISP, a public IP address is allocated to the access point (a set top-box for instance) used by the individual. Each time a terminal (smartphones, tablets, etc.) is connected to the access point, it receives a local IP address. For somebody outside the local network like a website, there is only the public IP address of the access point.

[2] Internet Corporation for Assigned Names and Numbers https://www.icann.org/ (previously IANA).

Is it critical to know if IP addresses are personal data? It is, because websites and ISPs can learn a lot of information that go beyond the normal function of Internet. An IP address can leak information on the user's location: IP addresses are to some extend geolocated [39]. They are used by many websites to adapt the language of a web content to the user for instance. This possibility of personalisation creates a risk of IP address-based discrimination. It was also demonstrated [48] that IP addresses can be used effectively to track users online despite their dynamic nature (possibility of re-allocation). Furthermore, IP addresses are also used to create cookies [36]. At the level of ISPs, there is also the risk of zero-rating discrimination [56]: an ISP favours some services over others. If IP addresses are personal data, the GDPR regulates the use and sharing of IP addresses and provide means (rights) to investigate how they are used by data controllers.

From a legal perspective, the debate on whether data are personal data is largely based on what one considers as an "*identifiable*" person, which is one of the criteria of the definition of personal data as provided for by recital 26 of the GDPR [29].

Some elements of answer have been put forward by data protection authorities, by the CJEU and even by some domestic courts in the context of the qualification of IP addresses under EU data protection law.

Three solutions can be underlined by these different stakeholders, which may be summarised as follows:

- IP addresses are not personal data.
- IP addresses are personal data when enabling identification of data subjects.
- IP addresses are personal data depending who holds additional information on an individual.

We have classified the opinions of DPAs, case law into these different categories, although this classification is subjective. In particular, Section 2.2 and 2.3 analyse opinions which added specific criteria to consider IP addresses as personal data, although many DPAs' opinions only state in a laconic manner that IP addresses are personal data without any explanation such as the Belgian "autorité de protection des données"[3] or the Norwegian DPA[4]. The authors do not claim to have found the perfect typology, and any comment on this would be welcome.

2.1 IP addresses are not personal data

This view is not widely shared; however, it is interesting to mention it for many reasons. First, from a logical point of view it would seem pretty consistent that IP addresses are not considered personal data for a reason: IP addresses are not allocated to an individual but to electronic devices which can be used by several persons and, in some circumstances, by hundreds of them or even thousands if the device is a router. For instance, it is reported that in Qatar, the traffic of all Qatari citizens is routed via only few IP addresses [61].

More importantly, there seems to be a certain confusion as regards what must be considered as enabling the identification of an individual. In other terms, does data need to be related to a specific individual, with his name, or does an IP address can be considered as a personal data considering that the individual is, if not identified, at least identifiable?

First, it must be recalled that "*such an address does not directly reveal the identity of the natural person who owns the computer from which a website was accessed, or that of another person who might use that computer*" [24]. However, there are other means to indirectly relate data to an individual, this is why, in our opinion, an IP address can be considered as personal data.

Despite this, French Judges and even the French '*Cour de Cassation*' which is the supreme judicial Court have considered for a long time that an IP address could not be considered as a personal data since it was not a nominative information. As a matter of fact, the Paris Court of Appeal claimed that "*this series of figures in no way constitutes indirectly nominative data relating to a person insofar as it relates only to a machine, and not to the individual who uses the computer to engage in counterfeiting*" [11]. In other words, since the IP address does not relate to the name of an individual - even indirectly - it means, according to the French Judge, that it does not constitute personal data. In another context the Cour de Cassation concluded that the processing of IP addresses to investigate and find people responsible for unlawful downloading of

[3] APD, Avis n 62/2016 du 23 novembre 2016 and Avis n 117/2019 du 5 juin 2019

[4] Datatilsynet, Offl. § 13 jf. fvl. § 13 (1) nr. 2, 13/12/2021

protected music or files did not result in personal data being processed [14]. We must be careful with such assertions because they are very context-dependent. It does not mean that the French Judge considered that IP addresses could not be considered as personal data in whatever circumstances.

2.2 IP addresses are personal data when enabling identification of data subjects

On the other side of the spectrum, many commentators and academics tend to consider that IP addresses are, or at least may be, personal data. Such a conclusion is derived from the fact that actual persons behind an IP address are identifiable even when not identified. In a common setup, the router used is one of a household, which means that it involves only a few persons who can be easily identified. It is worth mentioning that "*[t]o 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*" [29]. From this background one could conclude that IP address can be personal data since it is possible, at least for internet service providers, to bridge the gap between the IP address and the person who holds the contract with the ISP.

Many DPAs across the EU have acknowledged that IP addresses could be considered personal data as they enable the re-identification of data subjects, even indirectly. For further details, one can consult the documents provided by the Belgian DPA[5], the Norwegian DPA[6], the Irish DPA [12] and the opinion of Article 29 Working party [8].

To prove their point, some of these DPAs have argued, pretty interestingly, that it is irrelevant to determine precisely who is behind an IP address as long as you can single out such IP address. If an individual called "X" owns an IP address (`XXX.YYY.ZZZ.WWW`) and that one is able to single out this IP address among the internet traffic, the individual(s) behind this address are singled out, no matter whether the service provider is able to identify precisely "X" by his name.

The French DPA has acknowledged such a view in its opinion[7] about Google Analytics by claiming that "*[i]t is not required to know the actual visitor's name or (physical) address since, in accordance with recital 26 of the GDPR, such singling out of individuals is sufficient to make the visitor identifiable*". The EDPS concluded the same way in [26], considering that "*[a]ll records containing identifiers, including IP addresses, which can be used to single-out users, are considered as personal data and must be managed and protected as such*".

Other stakeholders, in this debate on the extent of the definition of personal data, have claimed that IP addresses would constitute personal data only when combined with other data.

According to such a view, IP addresses are not, by nature, personal data but they can become so if they are combined with other information. For instance, in the previously mentioned decision[7], the CNIL acknowledged that "*online identifiers, such as IP addresses or information stored in cookies can commonly be used to identify a user, particularly when combined with other similar types of information*".

In other words, IP addresses are a type of information which is part of a broader cluster of indicators which, put together, enable the unique identification of an individual. From this background, the European Data Protection Supervisor (EDPS) underlines in [26] that "*[i]n practice, IP addresses are not processed in isolation, but in combination with other attributes which provide additional possibilities for the identification of the individual*". Such a view seems to be – at least implicitly – shared by the Irish data protection commissioner (see [12]).

Such a debate even transcends the European borders. For instance, the New Zealand Office of the Privacy Commissioner[8] asserts that "*[b]y itself, an IP address is unlikely to be personal information but it could be*

[5] APD, Avis n° 62/2016 du 23 novembre 2016 (in french), https://www.stradalex.com/en/sl_src_publ_jur_be/document/cpvp_F-20161123-7 and APD, Avis n° 117/2019 du 5 juin 2019 (in french), https://www.stradalex.com/fr/sl_src_publ_jur_be/document/cpvp_F-20190605-21.

[6] Datatilsynet, Offl. § 13 jf. fvl. § 13 (1) nr. 2, 13/12/2021.

[7] https://www.cnil.fr/sites/cnil/files/atoms/files/med_google_analytics_anonymisee.pdf

[8] Office of the Privacy Commissioner, Are IP addresses personal information?,https://privacy.org.nz/tools/knowledge-base/view/338.

when combined with other information or when used to build a profile of an individual, even if that individual's name is unknown".

As a conclusion it must be said that, beyond the debate on the precise nature of IP addresses, the context is important to understand whether personal data are being processed. If cookies collected can reveal certain sensitive characteristics and that these cookies are related to a unique IP address one can deduce pretty easily the identity of the person behind these unique identifiers.

2.3 IP addresses are personal data depending on who holds the 'additional information'

This view largely bears on European Court of Justice case law which can be described as acknowledging a relative approach towards what must be considered as personal data.

The starting point of this reflection was set with the Promusicae judgment [22]. In this case an association of music producers asked an internet service provider to communicate the name of persons to whom IP addresses suspected of breaching copyrights law belong. The judgment did not tackle the issue of IP address as personal data, however the Opinion of the Advocate General provided some insights on this issue. The Advocate General [44] stated in particular that "*[t]he indication of which users were assigned particular IP addresses at particular times consists of personal data under Article 2(a) of Directive 95/46, namely information relating to identified or identifiable natural persons. With the aid of those data, the actions performed using the IP address concerned are linked to the subscriber*". However, it is unclear whether this assertion bears directly on the nature of IP addresses as personal data or whether it implies that it is the fact of linking an IP address to an individual which can be considered as processing personal data. The Advocate General assesses in [44] whether there is a legal basis in this case to process traffic data, which suggests that traffic data (including IP addresses) are personal data protected under the data protection directive.

The solution given is not very surprising since it is clear that when an Internet Service Provider associates an IP address with a contract subscriber it processes personal data. In any case this ruling was a milestone as regards the CJEU's view on IP addresses as personal data.

In the same line of thoughts, the CJEU had to settle another similar case a few years later involving an Internet Service Provider and an association of authors and editors. The ISP was asked by SABAM (the association) to implement a filtering system on IP addresses to prevent customers from downloading files belonging to its catalogue in violation of copyright law [25]. In this case, the Court explicitly stated that "*the injunction requiring installation of the contested filtering system would involve a systematic analysis of all content and the collection and identification of users' IP addresses from which unlawful content on the network is sent. Those addresses are protected personal data because they allow those users to be precisely identified*". From this background it would have been logical to conclude that the CJEU embraces the fact that IP addresses are personal data but actually CJEU's view is much more complex.

However, the Advocate General in this case [59] paved the way for a more relative approach regarding the qualification of IP addresses as personal data by considering that "*[t]he question is, therefore, to determine not so much the legal status of IP addresses as the circumstances in which and the purposes for which they may be collected, the circumstances in which the resulting personal data may be resolved and processed, or even the conditions under which their collection and resolution may be requested*". In other words, IP addresses are not, per se, personal data. On the contrary, their qualification as personal data is pretty much context-related. This relative approach has been further explained in another famous ruling of the CJEU, in the Breyer case. This case is instructive since neither the plaintiff, nor the defendant were internet service providers, i.e persons in capacity to identify directly an individual via the IP address.

The CJEU's stated in this case [23] that "*the fact that the additional data necessary to identify the user of a website are held not by the online media services provider, but by that user's internet service provider does not appear to be such as to exclude that dynamic IP addresses registered by the online media services provider constitute personal data*" but that "*it must be determined whether the possibility to combine a dynamic IP address with the additional data held by the internet service provider constitutes a means likely reasonably to be used to identify the data subject*". Put differently, it must be assessed whether the web service provider is in capacity to obtain the additional information from the third party (the ISP) to consider whether the IP address is personal data. The Advocate General explained in [53] the "reasonable means" likely to be

undertaken. To apply this criterion to the facts, the Advocate General considered that "*[a] dynamic IP address must be classified, for the provider of Internet services, as personal data in view of the existence of a third party (the Internet service provider) which may reasonably be approached in order to obtain other additional data that, combined with a dynamic IP address, can facilitate the identification of a user*". In *Breyer* [24], the Court considered that German federal institutions, acting as a web media service provider, had legal means to access information held by the ISP as German law provides that, in case of cyberattacks, the web media service provider "*is able to contact the competent authority, so that the latter can take the steps necessary to obtain that information from the internet service provider and to bring criminal proceedings*".

What is interesting with the CJEU case law is that it tends to assess whether an individual is identifiable from the data controller's perspective and not the nature of data processed in itself. This debate extends to another critical issue, which is the issue of the qualification of pseudonymised data as personal data. Indeed, IP addresses can be compared to pseudonymised data, as in both cases re-identification is possible through the use of additional information.

3 The debate on the extent of the definition of personal data

The purpose of these following subsections is to demonstrate that the distinction between pseudonymised data and anonymised data is clearly set by the GDPR even though the application of such a distinction can be difficult in practice. In particular, the Court of Justice of the European Union retains a relative approach of what constitutes personal data under EU law, which leads to great uncertainties for data controllers, data subjects, and more generally for the legal framework concerning data transfers.

3.1 The theoretical distinction between pseudonymisation and anonymisation

The premise of this work is related to the difference between pseudonymised data and anonymised data. In this paper we will use on purpose the expression de-identified data [50] to designate both categories since these two techniques aim to hide the relations between data and people they belong to, for privacy purposes. However, the implementation of these two techniques do not lead to the same conclusions from a legal perspective.

We will not dig into too much detail when it comes to pseudonymisation or anonymisation as processes, but we will analyse thoroughly the outputs of such processes, i.e pseudonymised data on the one hand and anonymised data on the other.

Anonymised data are defined by the NIST [37] as "*data from which the [data subject] cannot be identified by the recipient of the information*". On this matter, data protection authorities consider in [6] that "*[d]ata can be considered effectively and sufficiently anonymised if it does not relate to an identified or identifiable natural person or where it has been rendered anonymous in such a manner that the data subject is not or no longer identifiable*" (See also Recital 26 of the GDPR [29]). This process must be irreversible [9], that is to say that one must not be in capacity to re-identify data subjects from the data collected and processed. Subsequently, if a data subject cannot be identified from a dataset it means that the data are not related to an individual anymore, and thus it is anonymised data. The GDPR requires that data subjects should no longer be identifiable from the data to consider said data as anonymised. This suggests that once data are anonymised, it is impossible to trace back data subject.

On the other hand, pseudonymised data are data which have been replaced by a pseudonym, i.e "*[a]n assigned identity that is used to protect an individual's true identity*" as defined by the NIST [37]. When data are pseudonymised, the identity of the data subject is protected, which does not mean that the subject is not retrievable. Indeed, as the Norwegian DPA [15] puts it "*[p]seudonymisation is the replacement of directly identifiable parameters with pseudonyms, which will still constitute unique identifying indicators. A likelihood therefore exists that the specific individual may be indirectly identified*". As individuals are attributed specific indicators, they can be retrieved in an indirect manner, by linking these indicators to themselves.

The fundamental difference between anonymised and pseudonymised data is that while the former is de-identified in a permanent manner, the latter can be re-identified through the use of additional information.

Recital 26 indeed provides that "*[p]ersonal data which have undergone pseudonymisation, which could be attributed to a natural person by the use of additional information should be considered to be information on an identifiable natural person*" no matters who holds this information. This is why the debate on IP addresses is relevant to analyse the differences between pseudonymised and anonymised data. When IP addresses are involved, there is additional information held by the internet service provider which permit the re-identification of data subjects. The same goes with encryption for instance. Encryption is a typical example of pseudonymisation, as emphasised by the EDPS and the AEPD in [7]. When data are encrypted they are de-identified, but this process is reversible since the decryption key permits the re-identification of data subjects be it from some encrypted identifiers or whole encrypted dataset. Basically the opposition between anonymised and pseudonymised data can be summarized as follows:

- Pseudonymised data = Existence of re-identifying additional information
- Anonymised data = No additional information left, the process is irreversible

It remains to be seen what the consequences derived from this premise actually are. From this background it is worth emphasising that the GDPR expressly mentions that pseudonymised data are personal data, which means that data protection law applies to pseudonymised data (see Recital 26 of the GDPR [29]).

On the other hand, Recital 26 of the GDPR [29] provides that "*[t]he principles of data protection should therefore not apply [...] to personal data rendered anonymous in such a manner that the data subject is not or no longer identifiable*". Even though pseudonymisation and anonymisation are two de-identification methods, hey are diametrically opposed to each other from a legal point of view, since the nature of the data will condition the application of data protection law. This view is endorsed by the Spanish DPA and the EDPS which claim in a joint document [7] that "*the use of 'additional information' can lead to the identification of the individuals, which is why pseudonymous personal data are still personal data. Anonymous data, on the other hand, cannot be associated to specific individuals*". The identifiability of a data subject is thus the key concept to understand whether data are personal data. This identifiability criterion is determined by the (im-)possibility for the data recipient to identify data subjects - even indirectly - and is basically what distinguishes anonymised data from mere pseudonymised data. The problem is that there is always a risk of re-identification so that such a distinction is more theoretical than practical.

3.2 The distinction between anonymisation and pseudonymisation in practice

Data protection authorities and domestic courts have been challenged by data subjects or data controllers to determine whether specific data had to be considered as pseudonymised or anonymised. We will expose some of their decisions, in order to provide a better understanding of the way anonymised data and pseudonymised have been interpreted in context. Due to a restriction of time and paper length, the authors do not claim that this presentation is a comprehensive one, however it gives some clue of the way Courts and DPAs can construe these two concepts.

For instance, in its decision regarding the use of Google Analytics by websites, the CNIL[7] claimed that unique identifiers, such as IP addresses, although if they are said to be anonymised "*may make it possible to identify visitors to the website [...] on which Google Analytics is used. It is not necessary to know the visitor's name or postal address since [. . .] such individualisation of persons may be sufficient to make them identifiable*". The CNIL acknowledges that even when data are encoded, or not related to a direct identifier, like a name, or address or so, they can still be used to single out individuals. The fact of singling out an individual is enough to consider such processing of data as a processing of personal data, it is irrelevant to know whether the data controller is able to effectively identify the data subject by name. As long as there is a singling out risk data cannot be considered as anonymised. This is in line with the guidelines of the Working Party 29 on anonymisation [9].

Furthermore, the Irish Data Protection Commission, in a Guidance Note on anonymisation claimed in [12] that "[i]f the data controller retains the raw data, or any key or other information which can be used to reverse the 'anonymisation' process and to identify a data subject, identification by the data controller must still be considered possible in most cases. Therefore, the data may not be considered 'anonymised',

but merely 'pseudonymised' and thus remains personal data, and should only be processed in accordance with Data Protection law". In such a case when data controllers hold the additional information needed to re-identify data, data must be considered as pseudonymised and thus, as personal. Once again it appears to be consistent with what the GDPR provides, since, as long as there is additional information likely to re-identify data, data cannot be considered otherwise than pseudonymised.

Such a view has been expressly endorsed by the French *Conseil d'Etat* in the JC Decaux case. JC Decaux is a company which holds advertising spaces in the streets. This company wanted to launch an experiment to assess the pedestrian flow in an urban area at Paris. The system deployed involved the processing of pedestrians' smartphones MAC addresses (see [13]). These MAC addresses were subject to a hashing and salting method by the company. The *Conseil d'Etat* had thus to settle the issue as to whether data involved were personal data and thus whether data subjects could exercise their rights (including the right to information) by virtue of the GDPR [13]. The *Conseil d'Etat* claimed, to this end, that "*the "hashing" and "salting" processes, while intended to prevent third parties from accessing the data, enable the data controller to identify the data subjects and do not prevent records relating to the same individual from being correlated or information about that individual from being inferred*". Once again, although the MAC address is not by nature, personal data, it becomes personal data when the data controller keeps the ability to link data to a specific individual to infer information from the dataset or to single out the data subject.

The public rapporteur of the above-mentioned JC decaux case further explained in [13] that "*[t]he fact that anonymisation systems are more or less robust, and that the CNIL has to examine on a case-by-case basis whether a given anonymisation process really achieves its aim, in no way detracts from the fact that pseudonymisation may, for its part, be structurally fragile, in that it is resistant neither to individualization nor to correlation*". What is worth underlining is the structurally fragile nature of pseudonymisation. Indeed, it is not even argued that pseudonymisation can prevent re-identification since, by their very nature, pseudonymised data are reidentifiable.

These findings are in line with some DPA's decisions which had to settle the issue as to whether a de-identified dataset must be considered as pseudonymised, and thus, made up of personal data, or as anonymised i.e made up of non-personal data. Classifying a dataset as pseudonymised or anonymised can prove to be difficult in practice, as emphasised by use-cases tackled by data protection authorities in the EU.

To mention just one example, the Italian DPA ('Garante privacy') had to settle a dispute[9] where a company processed health data, once anonymised, for further statistical processing. The company claimed that it was not bound by the GDPR since it processed only non-personal data. On the opposite the Italian DPA emphasised that "*the mere substitution of the patient's ID with an irreversible hash code obtained from the patient does not constitute, under any circumstances, a suitable measure with respect to the requirement of removing singularities (single out) necessary to qualify the processing as anonymisation*".

From that premise, it concluded that "*the processing in question qualifies as a form of pseudonymisation within the meaning of Article 4(5) of the Regulation*'' so that the company should be considered as processing personal data[9]. This decision thus suggests that the reliability of an anonymisation technique is assessed as regards the inherent re-identification risks related to this specific technique.

These cases tend to prove that sometimes data controllers think that once they have implemented privacy enhancing techniques (PETs) they do not process personal data, but things are much more complex as the mere ability to single out individuals means that data are pseudonymised and thus personal. It shows that the qualification of data as anonymised is very difficult and depends on a wide array of parameters.

3.3 Introduction to the relative approach to anonymisation (and personal data)

The distinction between anonymised and pseudonymised data could seem pretty clear so far, however, as often, the devil lies in the details. Indeed, the GDPR discriminates personal and non-personal data, with regard to the identifiability criterion. From this background it must be recalled that pseudonymised data relate to identifiable persons whereas anonymised data do not.

[9] Garante per la Protezione dei Dati Personali, Provvedimento del 1° giugno 2023 [9913795], available online, https://www.garanteprivacy.it/web/guest/home/docweb/-/docweb-display/docweb/9913795

The interpretation of this "identifiability" criterion is nonetheless debated. Recital 26 of the GDPR [29] provides that "*[t]o 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*".

It does suggest that data can be considered as anonymised even when they are not fully de-identified, provided that re-identification is unlikely, taking into account the means reasonable that one may implement to re-identify data. As a matter of fact the Swedish DPA, stated in [51], with regard to a dispute involving a local telecom provider and Google that "*[i]t is also not required that Google or Tele2 intend to identify the complainant, but the possibility of doing so is in itself sufficient to determine whether it is possible to identify a visitor. Objective means that can reasonably be used either by the controller or by someone else are any means that can reasonably be used for the purpose of identifying the complainant*".

These elements are the starting point of an academic as well as a judicial debate on the relative or absolute nature of anonymisation. In a nutschell, Recital 26 of the GDPR [29] requires that data controllers take into account the possibility to re-identify data through the "*reasonable means*" criterion, but it is not clear what this exactly means. Some advocate that the possibility of re-identification must be regarded in an abstract manner, considering the inherent strength and weaknesses of an anonymisation process. On the other hand, some others claim that the capacity to re-identify data must be assessed in a case-by-case basis as regards the information and means available to the data holder. It is also a risk-based approach, since data controllers will not have to prove that there is no risk of re-identification at all to consider data as anonymised and thus, non personal.

The Breyer case, mentioned above in the section on IP addresses is said to acknowledge a relative approach as the CJEU followed in the Advocate General footsteps by acknowledging in [24] that the means to be used would be unreasonable 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*".

Therefore, one has to look to all these features to consider whether data are personal from the data recipient's eyes. This relative (or subjective) approach "*would require consideration of all factors within one's knowledge—specifically who has access to relevant data that enables identification*" as explained in [35]. As the Irish DPA [6] puts it "*[o]rganisations don't have to be able to prove that it is impossible for any data subject to be identified in order for an anonymisation technique to be considered successful. Rather, if it can be shown that it is unlikely that a data subject will be identified given the circumstances of the individual case and the state of technology, the data can be considered anonymous*".

In other words, data are not inherently anonymous, they can be considered anonymised as regards the incapacity of the data holder to re-identify individuals behind data.

On the other hand, some other institutions advocate for an objective approach. In particular, the Working Party 29 claimed in [9] that "*anonymisation results from processing personal data in order to irreversibly prevent identification*". According to such a view data are anonymised (and thus non-personal data) only when it is not possible to re-identify them at all. This view supports to a large extent the technical differences between pseudonymisation and anonymisation. Anonymisation is supposed to be a one-way ticket which turns irreversibly personal data into non-personal data. There should be no such thing as "half-anonymised data".

Paradoxically, according to Frederik Zuiderveen Borgesius in [60] the CJEU in its Breyer ruling acknowledges an absolute approach of personal data as well, based on the fact that Recital 26 of the GDPR emphasises that it is not only the ability of the data controller to re-identify data that must be assessed but also that of "another person". In other words, it means that it must be assessed whether the data controller or anybody else can re-identify data, so such a view seems to reject the subjective approach. From this standpoint, it is irrelevant to know who is able to re-identify data. As long as someone can re-identify data, even though such re-identification is only hypothetical, said data cannot be considered as anonymised. Indeed, if one has to look to the reasonable means – irrespective of the meaning of this criterion - that anybody on earth possesses to re-identify data, in an abstract manner, no data can be considered as fully anonymised. Things have become even more complex since the CJEU's ruling in the SRB vs EDPS case. This decision reflects, in our opinion, a controversial view of what constitutes non-personal data.

3.4 The controversial CJEU's view on anonymised data

Even though it is difficult to assess whether data are anonymised or pseudonymised, as mentioned previously, one thing remained certain (at least until now): **pseudonymised data are personal data**. As a matter of fact, data protection authorities [15], throughout Europe keep asserting that "[p]seudonymised data is not synonymous with anonymised data".

This issue has been directly tackled by the CJEU [21] in the SRB v. EDPS case [46]. In this case the Single Resolution Board (SRB), a EU institution released a decision aiming to adopt a resolution scheme in relation with a Spanish banking institution [10]. Basically, the SRB collected written comments from the bank shareholders and creditors to assess whether they are entitled to ask for compensation. It pseudonymised the comments by assigning an alphanumeric code to each contribution. The SRB then transfered the pseudonymised comments to a firm called Deloitte to help complete their assessement [10]. Some shareholders filed a complaint to the EDPS, in its role of supervisory authority, on the ground that they had not been informed of the transfer of their personal data. The case was brought before the CJEU which had to determine whether the SRB transfered personal data to Deloitte by sending the pseudonymised comments.

On this specific issue the EDPS considered in [27] that "*the comments provided by the complainants in the consultation phase of the right to be heard process were personal data, and they were transmitted by the Applicant as pseudonymised data to Deloitte*". Therefore, the EDPS concluded from this premise quite logically that the data transfer involved pseudonymised data which are personal data according to the GDPR [27]. This conclusion is in line with the clear distinction set up by recital 26 of the GDPR (Recital 16 of the EUDPR [32]). Here, the EDPS can be considered as adopting an absolute approach, claiming that alphanumeric codes equal pseudonymised data which are, by their very nature, personal, irrespective of whether Deloitte possesses any means to identify which individuals are hidden behind those codes.

The CJEU had thus to settle this issue and released a decision on 26 April 2023 [21]. Surprisingly, the stance taken by the CJEU is the opposite to the one of the EDPS. Although the court did not deny that the pseudonymised comments were personal data in the hand of the SRB, it claimed that it was insufficient to consider that Deloitte actually processed personal data. The Court basically held that the EDPS should have tried to figure out whether Deloitte had reasonable means to get access to the additional information needed to re-identify data subjects from the pseudonymised data. In this case the CJEU thus advocates for a relative approach [46,54]. Put differently, even when pseudonymised data are concerned, they are not necessarily personal data, according to the Court.

The CJEU has reiterated its conclusions on the occasion of another case involving a "German trade association representing wholesalers of motor vehicle parts, and Scania CV AB ('Scania'), a Swedish vehicle manufacturer" [20]. The CJEU claimed that a vehicle identification number, which is basically an alphanumeric code relating uniquely to a specific vehicle, can be considered as personal data when the data controller has "reasonable means" to identify data subjects from the VIN. It results from this that when the VIN appears in the registration certificate of the vehicle, associated with the name of the owner of the vehicle, it is personal data [20]. Here the issue is quite similar to the one in *Breyer* as the "pseudonymised data" concern an item, not a person. The nature of data is thus dependent on the context and the stakeholders perspective.

This judgment is clearly acknowledging the British view on what anonymisation means and sets the bar very high as regards what constitutes personal data. Besides, it underlines the very contingent nature of the personal data notion, since the qualification of data as personal will depend on the anonymisation technique used, the person who holds the data, the person who holds additional information or even whether there has been a data transfer involved. From this background the ICO claims for instance in Guidelines [2] released in November 2012 that "*[t]here is clear legal authority for the view that where an organisation converts personal data into an anonymised form and discloses it, this will not amount to a disclosure of personal data. This is the case even though the organisation disclosing the data still holds the other data that would allow re-identification to take place*". These guidelines result from some British case law [17] and in particular from a decision issued in 2011. In this dispute the department of health was asked to answer a FOIA request bearing on late term abortion statistics. With regard to the high sensitivity of said data, the department of Health [17] refused to release this data considering that "*any statistical information derived from reporting forms or patient records constitutes personal data*" in particular taking into account that "*the body which*

publishes this statistic has access to information which would enable it to identify each of them". On this issue, the Court [17] held that "*the requested statistics were fully anonymised. It follows that the Tribunal ought to have held that the disclosure of information to the public did not constitute the processing of personal data*". In similar cases, the Appellate Committee of the House of Lords [41] and the the British NHS[10] put the emphasis on the subjective definition of personal data.

This view thus gets rid of the clear distinction set up in the GDPR between pseudonymised and anonymised data. The only criterion which must be taken into account to qualify data as personal is - according to the CJEU - whether the entity which holds the data can reasonably re-identify them. This stance raises huge legal issues because it tries to conciliate a relative approach of re-identification, which lies on the means reasonably likely to be used, with a consequence from that premise which is necessarily absolute. Indeed data are or are not personal, there is no middle-way. This logical loophole thus raises many concerns with regard to EU data protection law and governance.

3.5 Consequences from these legal and technical inconsistencies

The main problem with the relative approach advocated by the CJEU, be it in its SRB vs EDPS Judgment or in prior decisions, is that it is a source of great legal uncertainty. If one considers that data have been anonymised, the GDPR no longer applies, which means that data are not subject - for instance - to data storage limitation period [23], or that data can be transferred to third parties even in third countries (including those which do not offer adequatehttps://fr.overleaf.com/project/655b1aec8425fb2e954ce182 data protection) without any further limitation as European data protection law does not apply anymore [18].

It also entails that the exact same data can be considered as pseudonymised for a certain party and anonymised for another. This reasoning tends to undermine the distinction between anonymised and pseudonymised data. As mentioned previously, the distinction between pseudonymised and anonymised data lies in the fact that, contrary to the former, the latter must result in an impossibility to re-identify data [9]. This impossibility must be interpreted in a dynamic manner, taking into account the development of re-identification techniques [6]. On the contrary, pseudonymised data bear in themselves the possibility of further re-identification. In the SRB vs EDPS decision, the CJEU concludes that even though initial data controllers can re-identify data, it must be assessed whether the data recipient can re-identify data to conclude that data are personal.

However, although one considers that Deloitte was unable to re-identify data, it should have been concluded that Deloitte still processed personal data since alphanumeric codes were attributed to specific individuals which could be singled out. It must be recalled that 'singling out' is one of the criteria used by the Article 29 working party [9] to assess the re-identification risks of said anonymised datasets. In any case, the SRB could still re-identify the data.

The problem with such a view is that it acknowledges, at least implicitly, that when de-identified data are involved, the nature of data as personal or non-personal depends on the context, on who holds data, and on the technical and financial means of the data recipient. This results in great legal uncertainty be it from the data subject's perspective or the data controller's one.

If we were to adopt a literal interpretation of the CJEU's SRB vs EDPS case it would mean that a same dataset could be published without restriction by certain people (those who cannot re-identify data) and not by others (those who can use reasonably likely means to re-identify data). This issue is very critical since some companies sell anonymisation techniques using the argument that data will not be subject to the GDPR anymore. In the same line of thought, this would run the risk of a data controller publishing pseudonymised data in an open access format, on the assumption that the data are anonymised for any third party whereas it is not the case. Indeed, it is hard, nay impossible to assess the 'means reasonably likely to be used' by any data recipient when data are published online without restriction. The means reasonably likely to be used by a big tech company like Google will not be the same as those of the average web user.

[10] Controllers and personal data in health and care research Last updated on 19 Apr 2018, https://www.hra.nhs.uk/planning-and-improving-research/policies-standards-legislation/data-protection-and-information-governance/gdpr-guidance/what-law-says/data-controllers-and-personal-data-health-and-care-research-context/

To overcome these shortcomings, the ICO for instance explained in its guidelines [2] that "*[t]he more aggregated and non-linkable the anonymised data is, the more possible it is to publish it*". Put it differently, data may be considered as more or less anonymised, anonymisation is not an achievable process with a beginning and an end. There is no such thing as anonymised data if we were to consider that some levels of anonymisation enable the free publication of data while others do not.

The problem, once again, is that data cannot be half-personal, they are or they are not personal. In the same vein, the Working Party 29 [9] emphasises that data controllers should update their anonymisation methods as regards the development of re-identification techniques. This does not look consistent, since, when a dataset is published, people will download it, so that once an anonymisation scheme is obsolete, any attempt to correct it will be useless as the original data will still be kept by those who downloaded them in the first place. The ruling of the CJEU, in the SRB vs EDPS case brings out the risk that data protection techniques and privacy enhancing technologies might be misused. We are not going to discuss all data protection techniques, there are too many of them, but we focus on classical encryption, as other forms of (homomorphic) encryption have been studied elsewhere (see [40]).

Encryption aims at protecting the confidentiality of data. An encryption algorithm processes personal data into some ciphertexts using an encryption key. The original data can be retrieved by using the proper decryption key and decryption algorithm. If the encryption key is the same as the decryption key, the algorithm is said to be symmetric otherwise it is called asymmetric. There are also homomorphic encryption algorithms [38] for which it is possible to perform operations on the plaintexts over the ciphertexts. Encryption has been intensively studied for decades (see [57]) and in this work, we focus only on classical symmetric or asymmetric schemes. It is possible to find discussion on homomorphic encryption in [45,40]. From a re-identification perspective, there are two situations to consider when discussing the case of encryption. Parties which possess the decryption keys can re-identify the personal data. There are also risks of re-identification from parties which do not possess the decryption keys. These additional risks are related to the imperfection of encryption and its implementation:

- **Advances in cryptanalysis** – New techniques are created to attack the encryption algorithms. It can be an algorithmic advance or it can be related to technical improvement like the quantum computer.
- **Side-channel attacks** – The physical implementation of a cryptographic algorithm leaks information about the encryption/decryption keys to an adversary [43].
- **Difficulty of managing properly cryptographic keys** – The key needs to be generated and store in a secure manner to prevent any weaknesses that would help the cryptanalysis of an adversary.
- **Key length deprecation** – The sizes of cryptographic keys evolve over the years to match the advances in cryptanalysis. Therefore, some parameters may be deprecated and therefore considered insecure (See https://www.keylength.com/en/).

The GDPR requires that data subjects should no longer be identifiable from the data to consider said data as anonymised. This suggests that it is possible to provide an impossiblity result. Unfortunately, it is difficult to achieve such results even in cryptography as explained by Kutylowski et al. in [45]. There are even negative results: privacy-preserving cryptographic protocols cannot achieve privacy only by itself [16]. Even with a technique as strong as encryption, there are still re-identification risks.

So far, encrypted data have be considered as pseudonymised data since there is a decryption key enabling the re-identification of individuals. In a joint paper, the Spanish DPA and the EDPS claimed in [7] that "*[e]ncryption is not an anonymisation technique, but it can be a powerful pseudonymisation tool*".

Following the CJEU's reasoning, encrypted data might potentially be published online since nobody, except the encryption/decryption key holder, can re-identify data. The problem with such a view is that encryption does entail vulnerabilities and, in any case, encryption keys may be broken at a certain time. Acknowledging such a relative or risk-based approach entails a risk that said non-personal data are turned into personal data in the future. Indeed, such risk must be assessed with regard to the evolution of the state of the art as mentioned in Recital 26 of the GDPR [29] and as recall the Irish DPA [12].

To mitigate these risks for personal data, some academics have proposed in [49] that re-identification should be prohibited and punishable by criminal law. Indeed, the CJEU mentions that for re-identification

the reasonable means have to be legal [24]. Such an approach relies on the fact that even in the case of pseudonymisation, only legitimate third parties should be able to re-identify data. In the scenario of encrypted data, only legitimate holders of the decryption key should be in capacity to re-identify data. Obviously, the same goes with anonymisation, where nobody should undertake the efforts necessary to re-identify data. The problem is that such a view would undermine data protection and data subjects' privacy in Europe. Sanctioning people afterwards is going to be a hard task, especially because when data are said to be anonymised, datasets can be released in open data. It multiplies the risks of data being re-identified. In this scenario, relying only on criminal law would be ineffective.

The CJEU's view raises huge legal liability issues. Indeed, data that are not personal for a data holder 'A' can be personal for a data holder 'B', which means that 'A' will not have to fulfill all the obligations under the GDPR while 'B' will have to. This system cannot stand since a same dataset would be subject to different rules, and nothing would prevent 'A' to publish data as he is not subject to the GDPR anymore.

Nonetheless, Article 33 of the GDPR [29] requires data controllers to notify a data breach to the supervisory authority when it can result in a risk to the rights and freedoms of a natural person. Can data controllers be held liable of a data breach when they anonymised a dataset which has been subsequently re-identified? It seems to us that it would be impossible since GDPR does not apply anymore in relation to the original data controller who published the data. For instance, the Norwegian DPA states in [15] to this end that "*[t]he advantage of anonymisation is that the further processing of the data can take place without incurring any form of processing liability*". So the only outcome would be to hold liable the data re-identifier who breaches the GDPR (and potentially criminal law) by re-identifying data. Legally speaking, this view can stand, but, on the other hand, most data re-identifiers will likely act in a covert manner so that DPAs and original data controllers would not even notice that a data breach has occurred. The only ones that would eventually bear the consequences of such a situation are the data subjects who will find it difficult to seek redress for an event they are not even aware of. In case of a re-identification, some [55] consider that "*the anonymized data become personal data again and we end up in a situation of joint controllership. Both the person responsible for the anonymization of the dataset and the recipient of the dataset responsible for combining the different sets and/or with the intention to identify or evaluate data*". Such an interpretation could seem surprising. If data protection authorities acknowledge that an anonymisation process has been carried out seriously, there is no reason why the original data controller would still be responsible for a further data breach.

4 Moving Beyond the Anonymisation Debate

While it is often argued that the GDPR is an obstacle to data sharing, this section tends to demonstrate that the GDPR provides a protective framework which does not necessarily prevent the transfer of data or their further use. Besides, EU lawmakers have initiated a move towards a better protection of data including non-personal data.

4.1 Enforcing data protection law rather than escaping from it.

Anonymisation is no silver-bullet since re-identification risks still exist. The logical consequence from this premise should be that anonymised data, also guaranteeing a **higher level of security**, must be considered as pseudonymised data. Accordingly, data protection law should still apply as long as re-identification risks is not zero. However, it would be a specific category of personal data because it is not argued here that anonymised data can be considered the same way as any plain text data. For instance, Recital 62 of the GDPR [29] mentions that "*the provision of information to the data subject [may prove] to be impossible or would involve a disproportionate effort*". It is possible to consider that when data subjects are not directly re-identifiable, and that data controllers have taken all reasonable steps to secure data, the provision of information to data subjects might possibly be seen as a disproportionate effort in this context.

Even though one may understand that sharing data in a global context of digital and economic competition is critical, it cannot be done at the expense of data protection in the EU. Furthermore, explaining that anonymised data must be considered the same way as pseudonymised data, does not equal an absolute

prohibition to share data. It would be possible to get inspired by the new proposal on the European Health Data space. Indeed, the latter permits access to pseudonymised data, but this access is strictly framed. In particular, applicants asking for access have to specify how the processing would comply with Article 6 of the GDPR and they have to assess the ethical aspects of the intended data processing [19].

It is argued here that protecting personal data does not equal an absolute prohibition of disclosure. However, it means that open data policies cannot be implemented when there is still a risk of re-identification. Hence, the need for a legal framework on data sharing. This is why going beyond this trade-off between the free flow of data and the respect for privacy is needed. In particular, if we analyse the SRB vs EDPS case through this prism, it must be recalled that this case deals with the right of data subjects to be informed. In particular, the complainants [21] "*relied on the fact that the SRB had failed to inform them that the data collected through the responses on the forms would be transmitted to third parties [...] in breach of the terms of the privacy statement*". In his pleadings [27], the EDPS underlined that "*[t]he complainants did not have additional insight into the details of the processing carried out by the Applicant other than the information they received during the Registration Phase*".

It should have been relevant to consider that SRB still transferred personal data and that Deloitte received, and thus, processed, personal data. Such processing would not be necessarily unlawful, but it would force data controllers to implement proper safeguards and legal basis for such processing. This view has been acknowledged by legal practitioners (as in [58]) that assert that in the context of transfer of pseudonymised data to third parties, one may consider "*such disclosures as instances of regular sharing of personal data. An example of the latter approach can be seen in recent policy documents published by NHS trusts which state that [...] 'A data sharing agreement should be in place when pseudonymised information is to be transferred to a third party*"'. The French *Conseil d'Etat* recently gave an example of the way de-identified data can be disclosed, even outside the EU borders, with proper safeguards. Indeed, the French court claimed that the company Doctolib, which enables people to take medical appointments online, acted in compliance with data protection law by using encryption methods to transfer data to a data storage solution belonging to an Amazon Branch located in the EU. Furthermore, to conclude this way, the French Judges also assess the contractual agreements between Doctolib and AWS (Amazon branch). It does mean that even pseudonymised data can be transferred when there is a legal basis and adequate security measures [42].

It can be concluded that the GDPR does not prevent data from being re-used in an absolute fashion. However, the GDPR is a framework which limits potential data processing abuses. For instance, when anonymised data are re-identified, it is not clear who will be held responsible of such a data breach. Hence, the need to keep anonymised data within the scope of the GDPR.

4.2 Towards a new notion of data under EU law

It must be emphasised that this move beyond the debate on anonymisation has been initiated by EU lawmakers as well. Indeed, sharing data has become a real economic challenge. At the same time, the multiplication of the sources, types and users of data makes the distinction between personal and non-personal data harder. While some advocate for a more flexible concept of personal data – which would facilitate their sharing – it is argued here that data protection should prevail over the logic of open data and data sharing without constraints.

As it has already been mentioned, when data are considered non-personal, their holder will not have to comply with all the requirements provided for by the GDPR, or any personal data protection legal framework. On the opposite, they are subject to the EU regulation on the free-flow of non-personal data [30]. In particular, Recital 9 of this regulation [30] provides that "*[s]pecific examples of non- personal data include aggregate and anonymised datasets used for big data analytics [...]. If technological developments make it possible to turn anonymised data into personal data, such data are to be treated as personal data, and Regulation (EU) 2016/679 is to apply accordingly*". The same goes for the Data Governance Act [31] which aims to facilitate the re-use of data. Although defending a logic of free-flow of data, Recital 15 of this regulation [31] expressly provides that "*non-personal data should be transmitted only where there is no reason to believe that the combination of non-personal data sets would lead to the identification of data subjects*".

In other words, these two texts expressly provide that the distinction between anonymised data and personal data is not that clear-cut due to the risk of re-identification and that, accordingly, if the anonymisation process is reversible, data protection law must be applied by default.

The EU lawmaker seems, to a certain extent, to embrace this ubiquitous re-identification risk and to consider that, even in the event of the processing of non-personal data, a certain degree of risk must be taken into account so that data still need to be protected. Put another way, non-personal data cannot be left in the wild, as regards the residual risk for privacy. The EU legislation thus tend to find a middle-way between the protection of personal data, and the free-flow of anonymised data. One of the clearest example of this trend lies in the adoption of the data governance act [31] which has already been mentioned. Recital 24 of the regulation [31] specifies that some non-personal data will be subject to stricter conditions with regard to their transfer when they are related to specific sectors, such as health for instance. What is worth emphasising here is that the sensitivity of such non-personal data is assessed "*including in terms of the risk of the re-identification of individuals*". The residual risk of re-identification is taken into account, as anonymised data are not considered as any other non-personal data, because of the risk of their re-identification.

This logic also transpires from the new Data Act [33], which, without dealing expressly with the protection of personal data provides in Article 28 that service providers must take proper measures to prevent international governmental access to non-personal data when it might run contrary to EU law. It tends to suggest that, to a certain extent, the logic of data protection also extends to non-personal data (and thus, said anonymised data). This view has been endorsed by academics. In particular Lazarotto and Malgieri [52] consider that "*[i]mpressive computational capabilities are making it possible to identify data subjects even in datasets that – until recently – we would have considered "*anonymous*". However, the need to guarantee digital users' protection goes beyond the mere issue of identification, considering that many risks to fundamental rights online can occur even without any personal data processing*". According to them, there is a pressing need to go beyond the duality between personal and non-personal data and to protect data subjects irrespective of whether data are considered personal, non-personal (and thus as pseudonymised or anonymised). They conclude from this premise that "*the measures proposed by the Data Act indicate that the EU institutions acknowledge the "inextricability" of personal and non-personal data, and are trying to create a non-explicit "third way" of protection by imposing new obligations to data processing services related to non-personal data*". As a conclusion for this section, it should be said that pseudonymised data are personal data, as recalled by data protection law. Furthermore, besides the debate on what proper anonymisation really is, data must be protected to ensure that individuals' privacy is not hindered within EU borders. **Put it another way, if it were to be a change in EU legislation it should be towards a better protection of non-personal (anonymised) data and not towards a less protective framework of personal (pseudonymised) data.**

5 Conclusion

This paper shows that although EU data protection law is not new, the exact perimeter of the definition of personal data remains uncertain, as illustrated by the debate on the legal status of IP addresses. Things are getting even more complex with the development of de-identification techniques and privacy enhancing technologies. Indeed, it is critical to determine whether those techniques can be used to pseudonymise or anonymise data. Theoretically, pseudonymised data are related to identifiable person while anonymised data are irreversibly de-identified.

In practice, courts and DPAs do not always agree on what constitutes anonymised data. The GDPR itself is ambiguous on this question since the qualification of data depends on the means reasonably likely to be used to re-identify data and on the person who has the knowledge of the additional information needed to re-identify. The CJEU has even considered that data which are pseudonymised in the hands of a certain party could be seen as anonymised when held by another. This trend towards a relative definition of personal data runs the risk of undermining the rights of EU citizens to data protection. This is why we advocate for a move beyond the anonymisation debate to ensure that data remain protected even when they are said to be anonymised.

The ECJ released on March 7th 2024, two very important decisions in case C-479/22 P and C-604/22 in which it upheld the relative approach of the identifiability criterion and on what can be considered personal data (see [47]).

References

1. Politi s.a.s. v Ministry for Finance of the Italian Republic. - Reference for a preliminary ruling: Tribunale civile e penale di Torino Italy. No. Case 43-71 (1971), https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:61971CJ0043
2. Anonymisation: managing data protection risk code of practice. Tech. rep. (2012)
3. Charter of fundamental rights of the european union. OJ pp. 391–407 (2012), https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:12012P/TXT
4. Consolidated versions of the treaty on european union and the treaty on the functioning of the european union. OJ (C 326/47), 43–390 (2012), http://data.europa.eu/eli/treaty/tfeu_2012/oj
5. Proposal for a REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on the protection of individuals with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation), chap. Explanatory Memorandum. No. COM(2012) 11 final 2012/0011 (COD) (2012), https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex:52012PC0011
6. Guidance on Anonymisation and Pseudonymisation. Tech. rep. (2019), https://www.dataprotection.ie/sites/default/files/uploads/2019-06/190614%20Anonymisation%20and%20Pseudonymisation.pdf
7. 10 Misunderstandings Related to Anonymisation. Tech. rep., AEPD and EDPS (2021), https://edps.europa.eu/system/files/2021-04/21-04-27_aepd-edps_anonymisation_en_5.pdf
8. Article 29 Data Protection Working Party: Opinion 4/2007 on the concept of personal data. Tech. Rep. 01248/07/EN (June)
9. Article 29 Data Protection Working Party: Opinion 05/2014 on Anonymisation Techniques. Tech. Rep. 0829/14/EN (April 2014)
10. Board, S.R.: Adoption of a resolution scheme in respect of banco popular español (srb/ees/2017/08). Decision of the single resolution board in its executive session (2017)
11. Cour d'Appel de Paris (13ème chambre, A.: Sté civile des producteurs phonographiques et a, c/ sebaux henri. Tech. Rep. 06/01954 (2007)
12. Commission, D.P.: In the matter of WhatsApp Ireland Limited Decision of the Data Protection Commission made pursuant to Section 111 of the Data Protection Act, 2018 and Articles 60 and 65 of the General Data Protection Regulation. Tech. Rep. IN-18-12-2 (august 2021), https://edpb.europa.eu/system/files/2021-09/dpc_final_decision_redacted_for_issue_to_edpb_01-09-21_en.pdf
13. Conseil d'État, 10ème - 9ème chambres réunies: Lecture du mercredi 08 février 2017. Tech. Rep. ECLI:FR:CECHR:2017:393714.20170208 (2016), https://www.legifrance.gouv.fr/ceta/id/CETATEXT000034017907/
14. Cour de cassation, criminelle, Chambre criminelle: Tech. Rep. 80.787 (2007)
15. Datatilsynet: A guide to the Anonymisation of Personal Data. Tech. rep. (November), https://www.datatilsynet.no/link/2e642d84d9214490866a297a71a44c78.aspx/download
16. van Dijk, M., Juels, A.: On the impossibility of cryptography alone for privacy-preserving cloud computing. In: Venema, W.Z. (ed.) 5th USENIX Workshop on Hot Topics in Security, HotSec'10, Washington, D.C., USA, August 10, 2010. USENIX Association (2010)
17. England and Wales High Court (Administrative Court): R (on the application of) v. Information Commissioner. Department of Health (April 2011)
18. Esayas, S.: The role of anonymisation and pseudonymisation under the EU data privacy rules: beyond the 'all or nothing'approach. European Journal of Law and Technology **6**(2) (2015)
19. EUROPEAN COMMISSION: Proposal for a REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on the European Health Data Space (2022/0140 (COD)) (2022), https://eur-lex.europa.eu/resource.html?uri=cellar:dbfd8974-cb79-11ec-b6f4-01aa75ed71a1.0001.02/DOC_1&format=PDF
20. European Court of Justice (Eighth Chamber, Extended Composition): Case C-319/22 Gesamtverband Autoteile-Handel e.V. v Scania CV AB. JUDGMENT OF THE COURT (2023), https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:62022CJ0319
21. European Court of Justice (Eighth Chamber, Extended Composition): Case T-557/20 Procedure for granting compensation to shareholders and creditors following the resolution of a bank – Decision of the EDPS in which it found that the SRB failed to fulfil its obligations concerning the processing of personal data . JUDGMENT OF THE COURT (2023), https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:62020TJ0557

22. European Court of Justice (Grand Chamber): C-275/06 Productores de Música de España (Promusicae) v Telefónica de España SAU. JUDGMENT OF THE COURT 62010CC0070 (2011), https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:62006CJ0275
23. European Court of Justice (Grand Chamber): Joined Cases C-404/15 and C-659/15 PPU Reference for a preliminary ruling - Police and judicial cooperation in criminal matters - Framework Decision 2002/584/JHA. JUDGMENT OF THE COURT (2016), https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:62015CJ0404
24. European Court of Justice (Second Chamber): REQUEST for a preliminary ruling under Article 267 TFEU from the Bundesgerichtshof (Federal Court of Justice, Germany), made by decision of 28 October 2014, received at the Court on 17 December 2014, in the proceedings Patrick Breyer v Bundesrepublik Deutschland. Judgment of the court (2016), https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:62014CJ0582
25. European Court of Justice (Third Chamber): Case C-70/10 Scarlet Extended SA v Société belge des auteurs, compositeurs et éditeurs SCRL (SABAM). JUDGMENT OF THE COURT 62010CC0070 (2011), https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:62010CJ0070
26. European Data Protection Supervisor: Guidelines on the protection of personal data processed through web services provided by EU institutions. Tech. rep. (November 2016), https://edps.europa.eu/sites/default/files/publication/16-11-07_guidelines_web_services_en.pdf
27. European Data Protection Supervisor: Oral Hearing in Case T-557/20 SRB v European Data Protection Supervisor. Tech. rep. (december 2022), https://edps.europa.eu/system/files/2022-12/22-12-01_t-555-22-pleading_edps_en.pdf
28. European Parliament, Council of the European Union: Regulation (EU) 2016/679 of the European Parliament and of the Council, chap. Article 4. https://data.europa.eu/eli/reg/2016/679/oj
29. European Parliament, Council of the European Union: Regulation (EU) 2016/679 of the European Parliament and of the Council (2016), https://data.europa.eu/eli/reg/2016/679/oj
30. European Parliament, Council of the European Union: Regulation (EU) of the European Parliament and of the Council (2018), https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32018R1807
31. European Parliament, Council of the European Union: Regulation (EU) 2018/1724 of the European Parliament and of the Council (2022), https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32022R0868
32. European Parliament, Council of the European Union: Regulation (EU) 2018/1725 of the European Parliament and of the Council (2022), https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32018R1725
33. European Parliament, Council of the European Union: Regulation (EU) 2023/2854 of the European Parliament and of the Council (2022), https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32022R0868
34. European Parliament, Council of the European Union: Regulation (EU) 2023/2854 of the European Parliament and of the Council (2023), https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=OJ:L_202302854&qid=1706695095736
35. Finck, M., Pallas, F.: They who must not be identified—distinguishing personal from non-personal data under the GDPR. International Data Privacy Law **10**(1), 11–36 (03 2020)
36. Fouad, I., Santos, C., Legout, A., Bielova, N.: My cookie is a phoenix: detection, measurement, and lawfulness of cookie respawning with browser fingerprinting. Proc. Priv. Enhancing Technol. **2022**(3), 79–98 (2022), https://doi.org/10.56553/popets-2022-0063
37. Garfinkel, S.: De-identification of personal information (2015-10-22 2015). https://doi.org/https://doi.org/10.6028/NIST.IR.8053
38. Gentry, C.: Fully homomorphic encryption using ideal lattices. In: Proceedings of the 41st Annual ACM Symposium on Theory of Computing, STOC 2009, Bethesda, MD, USA, May 31 - June 2, 2009. pp. 169–178. ACM (2009)
39. Guo, C., Liu, Y., Shen, W., Wang, H.J., Yu, Q., Zhang, Y.: Mining the Web and the Internet for Accurate IP Address Geolocations. In: INFOCOM 2009. 28th IEEE International Conference on Computer Communications, Joint Conference of the IEEE Computer and Communications Societies, 19-25 April 2009, Rio de Janeiro, Brazil. pp. 2841–2845. IEEE (2009), https://doi.org/10.1109/INFCOM.2009.5062243
40. Helminger, L., Rechberger, C.: Multi-Party Computation in the GDPR. Cryptology ePrint Archive, Report 2022/491 (2022)
41. House of Lords (Appellate Committee): Common Services Agency (Appellants) v Scottish Information Commissioner (Respondent) (Scotland). OPINIONS OF THE LORDS OF APPEAL FOR JUDGMENT IN THE CAUSE (2007)
42. Juge des référés: Lecture du vendredi 12 mars 2021 – Inédit au recueil Lebon. Tech. Rep. 450163, Conseil d'État (2021), ECLI:FR:CEORD:2021:450163.20210312

43. Kocher, P.C.: Timing Attacks on Implementations of Diffie-Hellman, RSA, DSS, and Other Systems. In: Advances in Cryptology - CRYPTO '96. Lecture Notes in Computer Science, vol. 1109, pp. 104–113. Springer (1996), https://doi.org/10.1007/3-540-68697-5_9
44. Kokott, J.: Opinion of Mrs Kokott – CASE C-275/06. Opinion of Advocate General (2007), https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:62006CC0275
45. Kutylowski, M., Lauks-Dutka, A., Yung, M.: GDPR - challenges for reconciling legal rules with technical reality. In: Computer Security - ESORICS 2020. Lecture Notes in Computer Science, vol. 12308, pp. 736–755. Springer (2020), https://doi.org/10.1007/978-3-030-58951-6_36
46. Lodie, A.: Are personal data always personal? case t-557/20 srb v. edps or when the qualification of data depends on who holds them. European Law Blog (2023), https://europeanlawblog.eu/2023/11/07/are-personal-data-always-personal-case-t-557-20-srb-v-edps-or-when-the-qualification-of-data-depends-on-who-holds-them/, accessed: 2023-11-27
47. Lodie, A.: Case c-479/22 p, case c-604/22 and the limitation of the relative approach of the definition of 'personal data' by the ecj. European Law Blog (2024), https://europeanlawblog.eu/2024/03/25/case-c-479-22-p-case-c-604-22-and-the-limitation-of-the-relative-approach-of-the-definition-of-personal-data-by-the-ecj/, accessed: 2024-15-04
48. Mishra, V., Laperdrix, P., Vastel, A., Rudametkin, W., Rouvoy, R., Lopatka, M.: Don't Count Me Out: On the Relevance of IP Address in the Tracking Ecosystem. In: Huang, Y., King, I., Liu, T., van Steen, M. (eds.) WWW '20: The Web Conference 2020, Taipei, Taiwan, April 20-24, 2020. pp. 808–815. ACM / IW3C2 (2020), https://doi.org/10.1145/3366423.3380161
49. Phillips, M., Dove, E.S., Knoppers, B.M.: Criminal prohibition of wrongful re-identification: Legal solution or minefield for big data? Journal of bioethical inquiry **14**, 527–539 (2017)
50. Porter, C.C.: De-identified data and third party data mining: The risk of re-identification of personal information. Shidler JL Com. & Tech. **5**, 1 (2008)
51. for Privacy Protection, S.A.: Supervisory decision under the general data protection regulation - tele2 sverige ab's transfer of personal data to third countries. Tech. Rep. DI-2020-11373
52. da Rosa Lazarotto, B., Malgieri, G.: The Data Act: a (slippery) third way beyond personal/non-personal data dualism? European Law Blog (2023), https://europeanlawblog.eu/2023/11/07/are-personal-data-always-personal-case-t-557-20-srb-v-edps-or-when-the-qualification-of-data-depends-on-who-holds-them/, accessed: 2023-11-27
53. Sanchez-Bordona, C.: Case C-582/14 Patrick Breyer v Bundesrepublik Deutschland(Request for a preliminary ruling from the Bundesgerichtshof (Federal Court of Justice, Germany)). Opinion of Advocate General, https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:62014CC0582
54. Spajic, D.: Anonymous vs. pseudonymous data: the CJEU reaffirms the relative approach to the concept of personal data. KU Leuven Centre for IT & IP Law (2023), https://www.law.kuleuven.be/citip/blog/anonymous-vs-pseudonymous-data-the-cjeu-reaffirms-the-relative-approach-to-the-concept-of-personal-data/, accessed: 2023-11-27
55. Stalla-Bourdillon, S., Knight, A.: Anonymous data v. personal data-false debate: an EU perspective on anonymization, pseudonymization and personal data. Wis. Int'l LJ **34**, 284 (2016)
56. Taylor, L.: From Zero to Hero: How Zero-Rating Became a Debate about Human Rights. IEEE Internet Comput. **20**(4), 79–83 (2016), https://doi.org/10.1109/MIC.2016.88
57. van Tilborg, H.C.A., Jajodia, S. (eds.): Encyclopedia of Cryptography and Security, 2nd Ed. Springer (2011)
58. Tran, D., Lee, S.: Pseudonymised data is personal data – but in whose hands? ico calls for views on third chapter of draft anonymisation guidance (February 2022), https://hsfnotes.com/data/2022/02/18/pseudonymised-data-is-personal-data-but-in-whose-hands-ico-calls-for-views-on-third-chapter-of-draft-anonymisation-guidance/, last retrieved January 17th
59. Villalón, C.: Opinion of Mr Cruz Villalón – CASE C-70/10. Opinion of Advocate General (20011), https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:62010CC0070
60. Zuiderveen Borgesius, F.: Breyer Case of the Court of Justice of the European Union: IP Addresses and the Personal Data Definition (Case Note). European Data Protection Law Review **3**(1) (June 2017)
61. Zuiderveen Borgesius, F.J.: Singling out people without knowing their names – Behavioural targeting, pseudonymous data, and the new Data Protection Regulation. Computer Law & Security Review **32**(2), 256–271 (2016), https://www.sciencedirect.com/science/article/pii/S0267364915001788