Dynamic IP Addresses Can be Personal Data, Sometimes. A Story of Binary Relations and Schrödinger's Cat

Alessandro EL KHOURY*

Case C-582/14 Patrick Breyer v Bundesrepublik Deutschland ECLI:EU:C:2016:779

Dynamic IP addresses can be personal data for the online media provider who stores them, when it has the legal means to obtain additional information to identify the data subject.

I. Introduction

According to Article 2(a) of Directive 95/46/CE (Data Protection Directive, henceforth DPD),[1] personal data is any information relating to an identified or identifiable natural person. Sometimes it is obvious which information constitutes personal data; some other times the exercise becomes complex and may lead to unexpected results. The paramount principle upon which EU data protection law is based is the possibility to qualify certain information as "personal data". Whenever is possible to separate the personal element from the information itself, the rules and safeguards stemming from EU data protection law become inapplicable. In fact, from the legal perspective, personal data has been defined in a binary fashion: it's either personal, or not.[2]

Since the *Lindqvist*[3] case the Court of Justice has interpreted the notion of personal data very broadly and the Article 29 Data Protection Working Party followed with *Opinion 4/2007 on the concept of personal data.*[4] The possibility to identify *directly* or *indirectly* a person through a number of pieces of information alone or combined (and possibly held by different data controllers) highlights the complexities that data protection experts are experiencing nowadays when dealing with BigData, data mining techniques and the collection of information through the Internet of Things. Machines around us are constantly collecting information to provide services, but not everything is

* Research assistant at HEC Paris and ICT external expert at European Commission; email: alelkhoury@gmail.com and Twitter @alelkhoury. The author wishes to thank Mr Luca Vogna, system architect at the European Commission.

[1] Directive 95/46 of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data, OJ L 281, 23.11.1995, p. 31.

[2] H Kuan, C Millard, and I Walden, "The problem of 'personal data' in cloud computing: what information is regulated?—the cloud of unknowing" (2011) 1(4) *International Data Privacy* 37.

[3] Case C-101/01 *Lindqvist* [2003] ECLI:EU:C:2003:596, para. 24. In this case a telephone number alongside a name was deemed to be personal data.

[4] Art. 29 Data Protection Working Party, *Opinion 4/2007 on the concept of personal data*, 01248/07/EN, WP 136.

doi:10.1017/err.2016.26

https://doi.org/10.1017/err.2016.26 Published online by Cambridge University Press

personal data: an Internet Protocol (IP) address is just a numerical label used to identify a device connected to a network. Who is the person operating that device in a given moment is not information that an IP address alone can provide.

In the judgment delivered on 19 October 2016 in the *Breyer*[5] case, after having carried out the exercise of deciding what is personal data and what is not, the Court of Justice ruled that dynamic IP addresses constitute personal data in relation to a certain provider, where it has the legal means which would enable it to identify the data subject through additional data held by another provider. The Court substantiated one of the major breakthroughs in modern data protection law and *de facto* recognized the existence of a *grey zone* where data can be personal and non-personal at the same time.[6]

II. Facts

The judgment originates from a request for preliminary ruling from the German Federal Court of Justice, in relation to an action brought by Mr Patrick Breyer against the Federal Republic of Germany, concerning the registration and storage by the latter of the IP address allocated to Mr Breyer alongside the date when he accessed several internet sites run by German Federal institutions, the terms entered in the search fields and the quantity of data transferred.

The action has not been brought before court by chance. As member of the PiratenPartei and lawyer expert in data protection Mr Breyer holds the view that the systematic retention of personal data is a harmful operation.[7] His commitment within the PiratenPartei revolves around the idea that the State's observation of citizens who are not suspected of a crime is an unacceptable violation of civil rights.

The question whether IP addresses should be considered personal data is an old one, and the Court of Justice addressed it in *Scarlet Extended* where it concluded that static IP addresses are protected personal data because they allow the precise identification of users.[8] The difference between *static* and *dynamic* IP addresses can be easily explained through an example: if we imagine IP addresses as coats of different colours, used to identify doctors in an hospital, using a static IP address means that each doctor will always wear the same coat, whereas using a dynamic IP address means that every time a doctor enters the hospital he will be assigned one random coat from the ones available.[9] The coat alone does not provide enough information to enable anyone to identify the doctor wearing it. However, identification is possible if the coloured coat is combined with other additional data.

Recital 26 of the DPD is a fundamental provision used to ascertain the nature of IP addresses as it states that "account should be taken of all the means *likely reasonably to*

[5] Case C-582/14, *Patrick Breyer v Bundesrepublik Deutschland* ECLI:EU:C:2016:779.

[6] A similar concept is enshrined in Austrian data protection law (*Datenschutzgesetz, 2000*). Referred to as "indirect personal data", it entails that information is "indirectly personal data" if the controller, processor or recipient of the data cannot identify the individuals using legally permissible means.

[7] See P Breyer, "Illegale Vorratsdatenspeicherung in der Telekommunikationsbranche?" (2012) 1-2 *Neue Juristische Wochenschrift* 14.

[8] Case C-70/10 *Scarlet Extended* ECLI:EU:C:2011:771, para. 51.

[9] In Information Technology this is technically known as DHCP (Dynamic Host Configuration Protocol).

https://doi.org/10.1017/err.2016.26 Published online by Cambridge University Press

be used" to identify a person. The recital generated a large debate in academia, which polarized around two criteria: according to the *objective* criterion, IP addresses are personal data when a user can be concretely identified, regardless of the abilities and the means of a service provider to do so; on the other hand, the supporters of the *subjective* criterion uphold the view that IP addresses become personal data only when there is the concrete capacity of a subject who has access that information, to use his own resources (i.e. gather additional data) to identify a data subject.[10]

The Federal Court of Justice (FCJ) referred two questions to the European Court of Justice (ECJ). With the first one, the FCJ asks essentially whether Article 2(a) of the DPD must be interpreted as meaning that a dynamic IP address stored by an online media provider (i.e. an operator of a website) has to be considered already personal data for that provider, where only a third party – such as the Internet Service Provider (ISP) – has the additional information necessary to identify the data subject which used that dynamic IP address. The second question concerns the balancing operated by Article 7(f) of the DPD, which determines the criteria for making data processing legitimate. According to it, personal data may be processed only if the processing is necessary for the purposes of the legitimate interests of the controller, except when it is overridden by the data subject's fundamental right. What the FCJ seeks to understand is whether the scope of Article 7 can be reduced, and this balancing can be operated *a priori* by a national law.

III. Analysis of the judgment

The Court first describes the relevant provisions (paras. 1–12). Paragraphs 12 and 15 of the German Law on Telemedia (TMG)[11] state that a service provider may collect and use personal data to make telemedia available only in so far as the TMG or another legislative provision permits, or if the user has given his consent. For the TMG, the collection, use and combination of personal data is allowed only to the extent that this is necessary for the purposes of charging the user for the service.

After analysing the proceedings *a quo*, and after recalling the academic debate on the *objective* and *subjective* criteria to determine the nature of dynamic IP addresses (paras. 13–25), the ECJ deals with the two focal points of the case: whether dynamic IP addresses should be considered as personal data and, if so, whether the national legislation is compliant with the DPD.

The first point concerns the factual circumstances in which the dynamic IP of Mr Breyer's computer could be considered personal data. The German Court of Appeal refers to the dynamic IP alongside the other data stored in the logfiles as "specific data on Mr Breyer's factual circumstances", because it reveals his behaviour on a certain website.[12] However, no matter how specific, it does not allow the identification of

[10] See also *Opinion of Advocate General Campos Sanchez-Bordona* in *Breyer* ECLI:EU:C:2016:339, paras. 52–53 and footnote 11 for a summary of German academic positions.

[11] Telemendiengesetz (Law on telemedia) of 26 February 2007, BGBI. 2007 I, p. 179.

[12] Regretfully, it has to be noted that none of the judicial bodies noticed that an IP address, besides its static or dynamic nature, is often sufficient to reveal the geographical location of the devices to which it is assigned, thus revealing the physical location of a data subject at a specific time. On the risks of geolocation see Art. 29 Data Protection Working Party, *Opinion 13/2011 on geolocation services on smart mobile devices*, 811/11/EN, WP 185, p. 7.

https://doi.org/10.1017/err.2016.26 Published online by Cambridge University Press

Mr Breyer: this becomes possible only when the ISP reveals information sufficient to identify the person operating behind that dynamic IP address at a specific time (paras. 23–24). That is the fundamental element that provides a bridge between the personal data held by the ISP, and the additional specific data held by German institutions.

The ECJ recalls the *Scarlet Extended* case, and specifies that those findings were related to the identification of users carried out by the ISP itself, whereas in *Breyer* the IP address is stored by an online media provider which does not have the additional data necessary to identify the users (paras. 31–36). Therefore, having stated the impossibility of German institutions identifying Mr Breyer only with the available data (paras. 37–39), the ECJ explores the possibility of *indirect identification* as set out in Article 2(a) of the DPD. The reasoning necessarily passes through Recital 26, which refers to "all the means likely reasonably to be used either by the controller or by any other person to identify the said person" (para. 42).

Paragraph 47 is the keystone of the judgment: the ECJ states that "although German law does not allow the ISP to transmit directly to the online media provider the additional data necessary for the identification" of Mr Breyer, it is possible that in the event of cyber-attacks "legal channels exist so that the online media services provider is able to contact the competent authority, so that the latter can take the necessary steps necessary to obtain that information from the ISP" and to file charges against the eventual perpetrators. Thus, the ECJ concludes that dynamic IPs have to be considered as personal data for the online media provider who stores them, when it has the legal means to obtain additional data to identify the data subject from the ISP (para. 49).

With the second question the referring court asks whether the storage of those IP addresses at the end of that consultation period is authorised by Article 7(f) of the DPD, considering that para. 15 of the TMG allows online media service providers to collect and use personal data of a user only to the extent of facilitating and charging for the use of that telemedia, highlighting the fact that, according to Germany, the storage is "necessary to guarantee the security and continued proper functioning of the online media service", by enabling cyber-attacks to be identified and combated (paras. 26–27). The referring court upholds the same view, as the whole operation should be interpreted as means "to facilitate the use of telemedia".

First, the ECJ ascertained that the German institutions operated *sine imperio* within the framework of Article 3(a) DPD, as their activities as online media providers do not qualify as activities of the State in areas of criminal law (paras. 50–54). As regards the compatibility of the TMG with Article 7(f) of the DPD, the TMG is interpreted restrictively by the referring court, meaning that it authorises the collection and use of personal data relating to a user, without his consent, only to the extent necessary to facilitate and charge for the specific use of the online media service, despite the fact that the objective of ensuring the general capacity related to the functioning of the website may be a valid justification too (para. 55). Recalling the reasoning of *ASNEF and FECEMD*,[13] the ECJ concludes that Article 7(f) of the DPD establishes an open balancing between the legitimate interests pursued by the controller and the fundamental rights and freedoms of the data subject, which has to be assessed on a case-by-case basis.

[13] Cases C-468/10 and C469/10 *ASNEF and FECEMD* [2011] ECLI:EU:C:2011:77, paras. 30–33.

The available criteria are listed exhaustively in Article 7 and "Member States cannot add new principles relating to the lawfulness of the processing of personal data or impose additional requirements that have the effect of amending the scope of one of the six principles provided for in that Article" (para. 57). The ECJ adds that Member States are allowed to further specify elements in accordance with the mechanism provided by Article 7, but the balancing between the free movement of personal data and the protection of private life should not be prevented *a priori*. The ECJ concludes that Article 7(f) of the DPD prevents a national law from reducing the scope of that article and to operate a balancing *a priori* (para. 64).

IV. Comment

The answers given by the ECJ are not ground-breaking: knowing that a specific type of data (dynamic IP addresses) can be considered personal data, sometimes, and that the scope of Article 7 of the DPD cannot be reduced by national law is not particularly interesting from the risk regulation perspective. What is interesting is the legal reasoning leading to those results, as it could be applied to the same category of data and subjects.

There are three key elements in the judgment. First, dynamic IP addresses and the data mentioned in the logfiles (paras. 23–25) belong to the general category of *metadata* that is, bluntly put, "data about data".[14] Technically speaking, metadata is not personal data, but it contains data about other data, which could be personal data. If we recall the classic binary theory of personal data, it becomes evident that metadata, by not being personal data, is outside the scope of the DPD. Thus, every processing related to it is possible[15] including collection, organization, adaptation... The list is long and it entails also the possibility of selling data and combining it with additional data, which happens to be one of the core businesses of BigData. Again, all these operations are possible outside the protections of EU law and beyond the consent of the data subject, due to the fact that *metadata* is not personal data. The (wrong) assumption is that the data subject's rights are not affected by the processing.

The second element rests on the fact that the ECJ based the second answer on the precondition that, while recording the logfiles of Mr Breyer, German Federal institutions acted *sine imperio*, thus making that legal reasoning applicable to subjects acting in the same position, such as any market operator.

The third element consists of the legal reasoning used by the ECJ in para. 47. In *Breyer,* "legal channels" consisted in the possibility for the German institutions to contact the competent authorities *sine imperio*, in order to oblige the ISP to disclose the missing information to identify Mr Breyer. The ECJ conceives these "legal channels" as any possible channel not prohibited by law. If we abstract from the specific case, other "legal channels" could be, for instance, contractual clauses or even specific contracts, to be enforced judicially or extra judicially. Nothing would prevent a company from

[14] J Gantz. and D Reinsel, "Extracting value from chaos" (2011) IDC iView Article sponsored by EMC Corporation, available online at <https://www.emc.com/collateral/analyst-reports/idc-extracting-value-from-chaos-ar.pdf> (last accessed 19 December 2016), 2.

[15] For the implications of metadata with data ownership and data mining see A El Khoury, "Data Protection & Risk Regulation. Cloud Computing: a case study" (2016) Thesis on file at LUISS School of Government, pp. 44–47.

https://doi.org/10.1017/err.2016.26 Published online by Cambridge University Press

concluding contracts containing these clauses and they would undoubtedly be "legal channels". All the more so if the requested data in the contract is *merely* metadata. All the more so if a company has enough contractual power to insert these clauses into general service contracts, and apply a take it or leave it policy.

The three key elements are glued together by "all the means likely reasonably to be used" of Recital 26 of the DPD. The ECJ concluded that dynamic IP addresses should be considered personal data for the German Federal institutions because the possibility to acquire the additional data held by the ISP, that allows the identification of Mr Breyer, is not prohibited by law and is not particularly complex, as it could happen at the request of competent authorities: this constitutes the *means likely reasonably to be used* by the German Federal institutions. "That would not be the case 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" (para. 46), so that the risk of identification appears to be insignificant.

If we broaden the reasoning, it should be noted that the concept of *proportionality* necessarily requires at least two terms of comparison. Something is proportionate or disproportionate compared to something else. An action may be *proportionate* or *disproportionate* according to the means available to a certain subject. Moreover, the notion of "means likely reasonably to be used" has an even a more relative nuance, as it depends on how valuable a certain result is for a subject. In other terms, there is a *double relativity*: one concerning the means vis-á-vis the subject using them, and the other one concerning the means vis-á-vis the objective that that subject wants to achieve.

The conclusion is that dynamic IPs are not personal data *per se*, but they can become so for a certain controller, if he has the lawful means to obtain any further data which would allow the identification of the data subject through them. The same logic could be applied *mutatis mutandis* to all metadata, which begs the question: considering the outcome of *Breyer*, is any information potentially personal data?

V. Conclusions – a story of binary relations and Schrödinger's cat

Schrödinger's cat is a thought experiment imagined by physicist E. Schrödinger in 1935[16] and is often used to describe the state of quantum superposition in quantum mechanics. A cat is placed in a sealed box with a minuscule radioactive source, a Geiger counter and a bottle of poison. The radioactive source has a 50% chance of decaying and being detected by the Geiger counter. When sensing decay, the Geiger counter activates a small hammer that breaks the bottle, releases the poison and kills the cat. Only after opening the box it is possible to see whether the cat is dead or alive, but before that moment the cat is potentially dead and alive at the same time. The cat being dead or alive is the equivalent of the possibility to qualify certain data as personal or not at a certain moment. Only at the end of the exercise was the ECJ able to determine that a dynamic IP is personal data in that specific situation.

[16] E Schrödinger, "The present situation in quantum mechanics" (1935) 23(48) *Naturwissenschaften* 807.

https://doi.org/10.1017/err.2016.26 Published online by Cambridge University Press

What *Breyer* shows is that, from a risk a regulation perspective, the binary notion of personal data is not particularly useful, considering that data collection and information flows are tremendously big and complex. A binary system is way too simple to be used as a basis for the applicability of EU data protection law. If we consider the so called "accretion problem" in data anonymization,[17] it becomes much more *likely reasonable* that certain non-personal data (e.g. a dynamic IP address) is combined with personal data, thus becoming personal too. The risk of identification increases with the number of databases and possible correlations that can be made, which makes the whole exercise of risk assessment somehow dependent on the *means likely reasonably* to be used for it. Account should be also given of the fact that today, cloud computing permits access to a wide variety of complex computing services at very low prices, which makes those *means* even more *likely reasonably to be used.*

These means depend upon the controller's willingness and ability to achieve a certain result. A correct risk assessment should quantify the likelihood of certain risks, and if we take into account all of the above this does not seem feasible. In other terms, we are trying to tell if Schrödinger's cat is dead or alive before opening the box. Does it mean that every piece of information is potentially personal data? There is no univocal answer. Yes, because the likelihood of identification is high and grows with time. No, because not every piece of data can reveal factual circumstances or behaviours of data subjects.

Our solution would be to embrace the idea that personal data is necessarily a relative concept, and safeguards should be applicable beyond the notion of personal data. This would entail acknowledging a grey zone where certain provisions of the DPD are not applicable, namely, the ones linked with the exercise of the data subject's rights. All the other provisions concerning the processing, and especially the ones related to the quality of data (Article 6), confidentiality and security of the processing (Section VIII), notification (Section IX) and data transfers to third countries (Chapter IV), should remain applicable. In practice, the risk would be regulated through the processing operations rather than through personal data, as the ECJ did in *Breyer*. Dynamic IP addresses have been considered personal data for the German Federal institutions because they could *likely reasonably* put in place processing which would allow the identification of Mr Breyer. Like Schrodinger's cat, the fate of dynamic IP addresses has been revealed only at the very end.

[17] A Narayanan and V Shmatikov, "Robust de-anonymization of large sparse datasets" (IEEE Symposium on Security and Privay, 18–22 May 2008), 111–125. For the accretion problem, once an adversary has linked two anonymized databases together, he can add the newly-linked data to his collection of outside information and use it to help unlock other anonymized databases. See also P Ohm, "Broken promises of privacy: Responding to the surprising failure of anonymization" (2010) 57 *UCLA Law Review* 1701, 1746.

https://doi.org/10.1017/err.2016.26 Published online by Cambridge University Press