Scoping Personal Data

Towards a Nuanced Interpretation of the Material Scope of EU Data Protecton Law

Dalla Corte, Lorenzo

Publication date
2019

Document Version
Final published version

Published in
European Journal of Law and Technology

Citation (APA)
Dalla Corte, L. (2019). Scoping Personal Data: Towards a Nuanced Interpretation of the Material Scope of EU Data Protecton Law. *European Journal of Law and Technology*, *10*(1). http://ejlt.org/article/view/672/908

European Journal of Law and Technology, Vol 10, Issue 1, 2019

EJLT European Journal of **Law and Technology**

Scoping personal data: Towards a nuanced interpretation of the material scope of EU data protection law

Lorenzo Dalla Corte[1]

Abstract

The concept of personal data – any information relating to an identified or identifiable natural person – is a cornerstone of the European data protection framework since its very inception. The processing of personal data is a *conditio sine qua non* for the applicability of EU data protection law. Despite the crucial importance of the notion, the boundaries of the concept are often blurry. Ascertaining whether data is personal frequently depends on each individual processing's concrete context and characteristics. As a result of the contextual and relative character of the notion of personal data, in cases dealing with indirect identifiability, much is left to the discretion of the interpreter.

European doctrine and jurisprudence favour an expansive interpretation of the notion of personal data. In particular, the identifiability threshold is seen as very low; at the same time, the ways in which the information can be said to be relating to a natural person are manifold. The combination between the low identifiability threshold, and of the wide range of ways to satisfy the requirement for a relational link between data and natural person, leads to an extremely wide material scope for EU data protection legislation. Data protection is thus becoming, it has been argued, 'the law of everything'.

This paper responds to the growing concerns surrounding the perceived over-inclusiveness of the notion of personal data, suggesting a balanced approach to its interpretation. It starts by defining the concept of personal data in EU data protection, taking into account law, doctrine, and jurisprudence. It then delves into the two most crucial elements of the concept of personal data: identifiability, and the connection that must link information and natural person to make the data personal. The paper concludes by providing a balanced reading of the concept of personal data, pleading for a nuanced approach to its interpretation.

1. Introduction

The concept of personal data, defined as 'any information relating to an identified or identifiable natural person', is core to the European data protection framework since its very inception.[2] The qualification of data as personal is a *conditio sine qua non* for processing to be considered within the material scope of the General Data Protection Regulation (GDPR), and thus for the applicability

of EU data protection law.[3] Despite the crucial importance of the notion, its boundaries are however oftentimes blurry. Ascertaining whether data is personal frequently depends on each individual processing's context and characteristics. The qualification of data as personal is contextual: the same piece of information can be anonymous in one moment in time and personal in another.[4] As a result of the relative character of the notion of personal data, much is left to the discretion of the interpreter.

Doctrine and jurisprudence appear to favour an expansive interpretation of the notion of personal data, setting a very low identifiability threshold. On one hand, according to the GDPR,[5] identifiability must be ascertained taking into account every means reasonably likely to be used by the data controller or by any other person. On the other hand, the suitability of anonymisation as a protection mechanism, and of anonymity as a basis for policy, have been subject to intense doctrinal criticism,[6] especially following a number of successful re-identification attacks[7] performed on (purportedly) anonymised datasets.[8]

At the same time, the ways in which the information can be said to be relating to a natural person are manifold: information can relate to a natural person not only by virtue of its content, but also due to the purpose of the processing, or of its result.[9] Depending on each individual processing instance, all data can thus potentially become personal. The combination of the low identifiability threshold and the wide range of ways in which data can relate to a person can lead to an extremely wide material scope of EU data protection legislation. Data protection can become, it has been argued, 'the law of everything',[10] used to tackle problems which it was originally not meant to solve.

This paper responds to the growing concerns surrounding the perceived over-inclusiveness of the notion of personal data,[11] highlighting the elements that can be used to develop a balanced approach to its interpretation. The concept of personal data was drafted to be broad and technology-neutral enough to avoid leaving any personal data processing instance unprotected. Its boundaries are elastic, and their extension depends on the reading given by the interpreter, and eventually by the courts. Ultimately, doctrine and jurisprudence bind the reading of the law, and concur in setting the concrete extension of the concept of personal data. I argue that, while the notion of personal data lends itself to the possibility of an overly expansive interpretation, there is also room to construe it narrowly enough to withstand the tests to come.

While investigations on the tenability of the notion of personal data as a regulatory instrument are certainly necessary and welcome, the General Data Protection Regulation (GDPR) has just started being applicable and is expected to remain in force for quite some time. It seems thus necessary, in consideration of the concern about the breadth of the notion of personal data, to also highlight the elements that can prevent its overextension in the short run. Several authors pointed out the inconsistencies of the concept of personal data, sometimes advocating for a regulatory overhaul;[12] less attention has been given to how to ensure the tenability of the notion within the current framework. The paper starts by defining the concept of personal data under EU data protection law. It then delves into the critiques raised towards its two most crucial elements: identifiability, and the link between information and natural person. The paper concludes by providing the elements for a balanced reading of the notion of personal data, pleading for a nuanced approach in its interpretation.

2. Personal data in EU data protection law

Personal data is defined, within the EU data protection framework, as 'any information relating to

an identified or identifiable natural person'.[13] An identifiable natural person is one that can be identified, directly or indirectly, by reference to a direct identifier, such as a name or an identification number, or by a combination of indirect identifiers, e.g. location data, IP addresses, or other factors specific to her identity.[14] As the Article 29 Working Party (A29WP) highlights, there are thus four core components to the notion of personal data:[15] information, a natural person, identifiability, and a link ('relating to') connecting the information and the data subject. Those components are cumulative: each is necessary to qualify data as personal. Lacking one or more of those components (e.g. the identifiability of the natural person to which the information relates, or the 'relating to' link between information and natural person) the data is not personal. The boundaries of the concept of personal data depend on the interpretation of, and the relation between, those four concepts.

The notions of information and of natural person as defined in EU data protection law will be examined more briefly than the concept of identifiability, or the link connecting data and data subject, since their legal construction within the current legal framework makes them less prone to be modulated by interpretative means. Any kind of information,[16] regardless of its nature, content, format or the medium in which it is contained, can qualify as personal. It does not need to be truthful or objective, nor secret or private, nor kept in a particular format or medium. Any sort of data can be personal, if it relates to an identifiable natural person: EU data protection law constructs the concept of information (or data, which it treats as synonyms)[17] as broadly as possible.[18] What constitutes a natural person, on the other hand, does not seem to raise particular problematics in respect to the extension of the material scope of data protection law, [19] and is mostly left for MS law to determine.

The plasticity of the notion of personal data within the current legal framework derives mostly from the possibility to interpret identifiability and the meaning of 'relating to', rather than the (all-encompassing) view of information adopted by EU data protection law, or the qualification of what constitutes a natural person.

2.1 'Relating to'

The data and the natural person must be connected by a link: the information must be *relating to* the data subject. That does not necessarily mean that the *content* of the information must identify the data subject. The wording 'relating to' delineates a broader range of ways in which information can connect to a person. There are indeed three ways in which data can relate to the person: through its content, but also through its purpose, or the result of its processing.[20] Those ways are alternative: one suffices to link the data and the natural person.[21]

Information relates to a data subject when its content is about that specific natural person, regardless of its actual purpose or impact, 'the most obvious and common understanding in a society of the word 'relate''.[22] An identity card, for instance, is personal data in that its content is about a natural person, to which it links irrespective of the purpose or result of the processing; likewise, a medical analysis relates to the patient by virtue of its content, notwithstanding the purpose or result of the processing.

Data can also relate to the data subject when, despite not being about him or her by virtue of the content element, it is used or likely to be used with the purpose of evaluating, influencing, or generating consequences for that natural person. The readings of the accelerometer in a smart phone, despite being *about* the phone, can very well become personal data, for instance when processed by a fitness app, 'with the purpose to evaluate, treat in a certain way or influence the

status or behaviour of an individual'.[23]

The result element links the information and the data subject where, despite not being about that natural person, nor meant to be used with a particular purpose concerning her, the data will have, or is likely to have, an impact or effect on the data subject. It is not necessary for the impact to be major: 'It is sufficient if the individual may be treated differently from other persons as a result of the processing of such data'.[24] The result of the processing, whichever the content or the original purpose of the information, can thus render any kind of data personal. Asset monitoring through sensors involves data whose content is about the asset monitored, processed with the purpose of monitoring that asset. However, if such assets are entrusted to a natural person by her employer, then the data could arguably be used with the additional purpose of evaluating the employee's performance, and would thus be considered as relating to the employee.

The GDPR's text does not explicitly construe the 'relating to' link as being integrated by content, purpose, and/or result. The currently[25] dominant interpretation of what 'relating to' means derives from the A29WP Opinion on the concept of personal data.[26] While the A29WP's Opinions are not binding, they do carry a large degree of significance in EU data protection doctrine and practice. The Court of Justice of the European Union (CJEU), the chief judicial authority of the EU, tasked with ensuring uniform interpretation of EU law, itself implicitly adhered to the construction of the relational tie set forth by the A29WP's Opinion on the concept of personal data in its YS and Nowak judgements,[27] and explicitly referred to another A29WP opinion in the recent Jehovan todistajat case.[28] National courts and supervisory authorities consider them in their proceedings, too. The A29WP's interpretation of the wording 'relating to', and of the notion of personal data *tout court*, is thus prominent in the European data protection milieu, cemented by the CJEU's interpretation in the YS and Nowak cases.

It is hence not only the content that qualifies the information as relating to a natural person, but also the purpose of its processing, or its (likely) result. To integrate the definition of personal data, and thus to trigger the applicability of the GDPR, the natural person to which the information relates must however be also identified or identifiable.

2.2 Identified or identifiable

A natural person is considered identified when isolated from a group of reference, and identifiable when it has not, despite the concrete possibility to do so. EU data protection law applies in both cases: the mere possibility to identify the natural person to whom the information refers is sufficient to trigger the applicability of the GDPR, it is not necessary for the data subject to be identified already.[29]

Data subjects can be identifiable both directly or indirectly. Identification is performed through pieces of information commonly called 'identifiers',[30] which can be either direct or indirect. Direct identifiers are data that identifies a single individual, either without additional information (e.g. my employee ID number) or by cross-correlating it with other information.[31] Indirect identifiers are data that does not allow the identification of the data subject on its own, but can reduce the sample to which the data subject belongs until, by correlating enough indirect identifiers, he or she becomes unique within the record and can thus be singled out.[32]

According to the GDPR, to establish identifiability, account should be taken of all the means reasonably likely to be used to identify the data subject, directly or indirectly. To ascertain whether such means are reasonably likely to be used, the interpreter should consider all objective factors inherent to the processing, considering both the available technology at the time of the

processing and the foreseeable technological developments.[33] Such means, furthermore, do not need to be used directly by the controller to be relevant for the identifiability test, but can be employed by either the data controller or by another person as well.[34] The wording of Recital 26 'suggests that, for information to be treated as 'personal data' […] it is not required that all the information enabling the identification of the data subject must be in the hands of one person'.[35]

3. The potential over-inflation of the concept of personal data

The concept of personal data, as summarised above, is very wide:[36] its definition is broad and contextual, highly dependent on the reader's interpretation. It is also largely technology-neutral, in that it is applicable to any kind of data type and processing technique, and covers situations where the identification of the data subject is merely potential (e.g. when the natural person to whom the information refers is indirectly identifiable through the means available by a person other than the controller).

Several authors have thus pointed out, directly or indirectly, the deficiencies of the notion of personal data, highlighting how its current normative, doctrinal, and judicial construction may backfire in the near future, overly inflating the material scope of EU data protection law. This section briefly accounts for those critiques, which mainly involved two elements of the notion of personal data: identifiability and its threshold, and the meaning of the syntagm 'relating to'.

3.1 Does everything relate to everybody?

The A29WP's opinion on the concept of personal data, and the assenting jurisprudence (i.e. the YS and Nowak cases), clarified how the 'relating to' link between information and natural person can be constructed by the data's content, purpose, or (likely) result. Both the A29WP[37] and the CJEU[38] interpreted 'relating to' in a broad manner, which gave rise to academic concern regarding the effects of such interpretation on the notion of personal data, and thus on the material scope of EU data protection law.

Purtova, for instance, claims that 'in the age of the Internet of Things, datafication, advanced data analytics and data-driven decision-making, any information relates to a person in the sense of European data protection law'.[39] Granted, 'some information is perceived as relevant more easily',[40] but 'when increasing amounts of data are gathered in real time from increasingly connected environments, intended to be used in automated decision-making about us, and we do not know how the autonomous self-learning and self-managing computers draw meaning from data, we should always reasonably assume that any information is likely to relate to a person, since we cannot eliminate this possibility with certainty'.[41] Purtova argues that, in a ''smart' city where all aspects of the environment and people living in it are datified, and the inhabitants are subjected to a certain treatment in real time based on processing of the data, from the speed at which escalators are running to promote physical activity to the warmth and intensity of street lighting to prevent undesirable behaviour to targeted policing',[42] many categories of information which are not generally considered as 'relating to' natural persons will be increasingly linked with individual data subjects. 'In such a world, any information within the 'smart' environment can be used and all information is likely to be used with the purpose of adapting the environment and impacting people'.[43]

Playing the 'devil's advocate', Purtova then highlights how a literal interpretation of the notion of personal data, and of the currently dominant doctrine and jurisprudence framing it, may render

personal data also information traditionally not seen as personal, e.g. 'rainfall per hour, temperature, wind direction and speed, together referred to as 'weather'',[44] in the context of 'smart cities' and other large-scale 'smart' environments. She maintains that weather data is indeed information, and that – despite the fact that its content is not about a natural person – it may very well be used with a purpose involving (or likely to involve) a natural person.[45] Even when the purpose of the weather data processing does not involve a natural person, she argues that its result is likely to make it relate to people.[46] In such a 'smart city' context, Purtova assumes, weather data will thus relate to natural persons by virtue of its purpose or of its result.

It can be argued that such a reading can be overly broad, particularly if coupled with the low identifiability threshold discussed below. That line of reasoning could potentially lead to incongruous consequences even if one does not consider that, as Purtova does, 'the 'narratives of a frictionless world that surreptitiously adjusts the environment to the needs and desires of its users' are steadily on the way out of the realm of science fiction'.[47] Such expansive interpretation of the 'relating to' link, and therefore of the notion of personal data, does not need to be applied to the data processed by a futuristic self-adapting 'smart' environment to make the material scope of the GDPR go haywire. Let us say that a person has a car insurance contract with an insurance company, and that the insurance premium varies according to the residence or domicile of the insured party. Such premium would be calculated according to a number of factors. Let us also say that one of them is the yearly average precipitations' amount in the area – more rain might lead to more car accidents, and thus to a higher risk for the insurance company, that would factor that into its algorithm. Can we also claim that the yearly precipitations' average rate of the area, as processed by the insurance company's algorithm, is the customer's personal data? It is information, the customer is a natural person who is readily identifiable, and the data can be used with the purpose of calculating the user's premium, having the result of making him pay less or more. In light of the definition of personal data set by the GDPR, and of its interpretation as given by the A29WP,[48] precipitation rates could thus be argued as being customers' personal data. That line of reasoning is of course quite extreme, if not plainly absurd, but it *can* be sustained by relying on an extensive interpretation of the law and its doctrine and jurisprudence.

3.2 Is everyone identifiable?

The natural person to which the information relates must also be either already identified or just identifiable for the definition of personal data to be integrated, and thus for the GDPR to be applicable. Many authors, from different disciplines,[49] have however pointed out how the anonymity/identifiability dichotomy (on which the notion of personal data is based) is oftentimes blurry, warning that the line separating an anonymous person from an identifiable one is thinner than generally assumed. At the same time, the construction of the identifiability requirement in EU data protection law, doctrine, and jurisprudence, is quite expansive, so that the threshold after which to consider someone as identified or identifiable is quite low.

Any person – not just the controller or processor – can have the capability to identify a data subject. According to a strict interpretation of EU data protection law, controllers dealing with indirect identifiers must therefore consider, along with their own identification capabilities, the ones available to *any* other subject.[50] If one considers, for instance, the possibility for the controller to contact the competent authority to identify people when necessary to initiate criminal proceedings,[51] or the existence and operation of private investigation firms, it appears evident how the identifiability test adopted by EU data protection law and jurisprudence can

potentially have a very low threshold.

The means that the controller or any other person can theoretically use to identify the natural person to which the information relates are all the ones available at the time of processing, considering the state of the art, if likely reasonably to be used.[52] In determining reasonable likelihood, account should be taken of all objective factors involving the processing at hand. As the CJEU stated in the Breyer case, the only instances where means are considered not likely reasonably to be used are the practical impossibility or the illegality of the means considered.[53] EU data protection therefore sets, *prima facie*, an extremely low threshold for considering such natural person identifiable. In combination with the many ways in which information can be qualified as 'relating to' a natural person, this results in an expansive reading of the notion of personal data. At the same time, the faith held in the anonymity/identifiability dichotomy – and therefore on the notion of personal data – has been, to some extent, waning away.[54]

The constant increase in the availability of auxiliary data, technological development, and know-how, contributed in rendering the border between identifiability and anonymity somehow fuzzy. Tracking individuals, offline and online, has proven to be a lucrative endeavour:[55] an entire industry is focused on researching and developing ways to render people identifiable for a plethora of purposes, from behavioural advertising to cybersecurity to law enforcement. At the same time, computing became ubiquitous,[56] thanks to smartphones and wearables, and everyday objects are increasingly being networked and digitalised. Physical spaces are being increasingly instrumented with sensors, in a merger between code and space[57] – from the Internet of Things to the so-called 'smart city'.

Anonymisation – the process of irreversibly turning personal data into anonymous data[58] – has also been under heavy scrutiny. Ohm's stance[59] provides an egregious summary of the concerns surrounding anonymisation. Moving from the example given by three widely-known re-identification attacks,[60] Ohm worries that some powerful actors will eventually create a giant 'database in the sky', constantly feeding it with additional data until singling out specific individuals within 'anonymised' datasets becomes easy[61] due to the amount of supplementary information available. It is essentially impossible to know the auxiliary information an adversary has at its disposal,[62] and thus threat modelling and the quantification of the risk of re-identification are intrinsically arbitrary. It has been also held[63] that the expertise level needed to perform a re-identification attack, particularly when plenty of auxiliary information is available to the attacker, is not as sophisticated as to constitute a sizeable barrier anymore.

A dataset's utility and the privacy it guarantees are moreover inherently at odds:[64] significant gains in confidentiality (e.g. through obfuscation, generalisation, or aggregation) are bound to diminish considerably the dataset's granularity (and thus utility), and vice versa. In statistical databases, furthermore, it has been shown[65] how it is impossible to formally guarantee that access to the database cannot enable an attacker to learn anything about an individual that could not be learned without access to the database, due to the attacker possibility to exploit auxiliary information.[66] In any scenario where the attacker's capabilities are not constrained by reference to an artificial threat model there will always be auxiliary information that, in conjunction with statistical data, allows to infer the identity of an unidentified subject, or some previously unidentified attributes of a known one.

Recital 26 of the GDPR, similarly to what Recital 26 of the Data Protection Directive (DPD), models the potential attacker as the data controller plus 'another person'. It is not necessary that the means necessary to identify a data subject (e.g. auxiliary data contained in a private record) are directly available to the controller: they can be available to another person.[67] Those means must

also be 'likely reasonably to be used', which has been interpreted very broadly by the CJEU: only practical impossibility and illegality can rule out the means considered as available to the controller or to another person.[68] That appears particularly far reaching if one considers that, in many cases, the 'other persons' whose means would have to be considered will be competent authorities, which often have the means to coerce information disclosure from third parties.

EU data protection law and jurisprudence adopt a very broad implicit adversarial model when framing indirect identifiability's threshold, and thus consequently anonymity's one. It has been underlined how, for the GDPR, almost anyone[69] could be the adversary – the subject attempting the re-identification of a record, or the one possessing the auxiliary information necessary to that purpose.[70] Data controllers must model their own identification capabilities assuming that they will be deemed as able to get to any auxiliary information available to any third party, unless practically impossible or illegal.[71] A literal interpretation of EU data protection law and jurisprudence thus leans towards an absolute (or objective) stance on anonymity, where a person is considered identifiable by the controller if anyone can identify him or her, as opposed to a relative (or subjective) construction,[72] where the person is considered identifiable only if the controller itself has the means to do so.

4. Narrowing down personal data

The normative definition of personal data can thus be interpreted extremely broadly. It is, indeed, largely a matter of interpretation: the law in itself is dead letter, it is the reading given by the actor that applies it – the interpreter – that makes it 'living', fit to solve conflicts and to regulate behaviour. What will concretely set its scope is its interpretation by the qualified interpreter – the courts, lawyers, and all the actors involved in data processing endeavours. '(W)ords do not bind the interpreters; rather the interpreters give meaning to the words. […] The critical people are the users, not the writers, of words'.[73]

The interpreter is thus fundamental in defining the material scope of data protection law. Doctrinal and jurisprudential production[74] can provide enough direction to guide the interpreter through an educated determination of what is personal and what is not. In this regard, despite adopting an apparently objective[75] and formalist[76] approach to identification, and a wide reading of the relational link, the A29WP also clarifies a few interpretative criteria to avoid the over-inflation of the concept of personal data. Likewise, the CJEU jurisprudence that can be read as opening up the door for an unchecked expansion of the material scope of EU data protection law does not have to be necessarily interpreted as extensively as recent literature suggests.[77] While some elements of the doctrinal and judicial construction of the notion of personal data can indeed be framed as all-encompassing, the same literature and case law also contain guidance on how to keep them in check.

4.1 Flexibility by Design in the GDPR

The notion of personal data, to begin with, must be interpreted within its broader framework of reference – EU data protection law.[78] This means that the finality of data protection[79] must be taken into account when delimiting the scope of the notion of personal data.[80] The GDPR lays down rules relating to the protection of natural persons with regard to the processing of personal data, and rules about the free movement of personal data. It protects fundamental rights and freedoms, and sets out that the free movement of personal data within the EU must not be restricted nor prohibited for reasons connected with data protection.[81] As the A29WP pointed

out, the law's subject matter and objectives play a substantive role in determining how to apply its provisions to a number of situations where the rights of individuals are not at risk.[82] That appears to be even more true when considering the introduction of the so-called risk-based approach,[83] which provides for additional flexibility in the law's application.

Even when data is personal, the applicability and application of the GDPR are conditioned by several exemptions and derogations. Aside from what is *tout court* excluded from the GDPR's material and territorial scopes,[84] EU data protection embeds a degree of flexibility that mitigates the width of the definition of personal data it adopts, or the consequences thereof.[85] The breadth of the notion of personal data should not automatically lead to overstretching the scope of the GDPR, nor to the application of its rules to situations which were not meant to be regulated through data protection law.[86] At the same time, an overly restrictive interpretation of the concept[87] should be avoided, in that it would deprive individuals from the protection that the legislator meant them to receive.[88]

Apart from the holistic approach to the GDPR, and the teleological reading of the concept of personal data in light of the function of the right to data protection, there are additional factors that can guide the interpreter in determining the extent of the notion of personal data in a reasonable manner. Data protection law does not apply to personal data in a vacuum, but to its *processing*. That is to say that each concrete processing instance has additional context that is bound to be considered when determining whether the data is personal. Each distinct case has additional contextual elements that, read through the interpretative canons provided by law, doctrine, and jurisprudence, (co)determine the information's personal (or anonymous) nature.[89]

European data protection law is a highly convoluted framework, cutting through different normative levels, and operating across many jurisdictions. While legal certainty is undoubtedly an objective to pursue, the complexity of the normative framework of reference derives from the intricacy of the matter at hand: balancing the right to data protection with the free movement of data. The contextual and relative character of the definition of personal data, while possibly detrimental to legal certainty, allows European data protection to be flexible and technology-neutral enough not to leave any data processing situation that can have an impact on individuals outside of its scope. There is, in other words, a natural and unavoidable trade-off between the flexibility of data protection law and the legal certainty that a narrower notion of personal data would grant.[90]

4.2 The interaction between identifiability and the 'relating to' link

Any information – regardless of its content – can be used for a purpose involving an individual, or can be processed in a way that results in an impact for a natural person. Any kind of data can possibly relate to people. That does not, however, suffice to make the information personal data on its own: the natural person to which it refers must also be identified or identifiable. I argue that, in situations where the 'relating to' link is integrated by the purpose or result elements, its interaction with the identifiability requirement functions as a logical limit to the potential over-inflation of the concept of personal data, and thus of the material scope of EU data protection law.[91]

When the relational link between information and natural person is justified through purpose or result, rather than through content, additional information will necessarily be required to render such person identified or identifiable, and thus the data personal. When the information relates to the data subject due to its content, the very substance of the information may lead to the identifiability of the data subject – its content is about that specific person in that it has

biographical significance. An employee ID identifies through its content, no matter how it will be used, or the result its processing will have. When the relational link is given by the result or the purpose one, however, additional information will *always* be necessitated to satisfy the identifiability requirement, and hence the definition of personal data. When the link is constituted by the purpose or the result of the information, the data becomes personal only when coupled with auxiliary data whose content leads to the identifiability of the related natural person. If the data could lead to the identifiability of the data subject on its own, the relational link would be integrated by the content element, without the need to recur to the purpose or result ones.

In a way, qualifying the data as relating to a natural person through the purpose or result element naturally leads to a higher identifiability threshold than the one required when the information relates to the data subject through its content. To be clear, the relationship test and the identifiability test are undoubtedly different assessments, aiming at evaluating two distinct components of the notion of personal data. What I argue is that the 'relating to' and the identifiability elements of the concept of personal data are logically related, tied together, so that the potential expansive effect of the 'relating to' element of notion of personal data as framed by the Article 29 Working Party[92] and by the CJEU in the Nowak case is reined in through the identifiability test.

If the *content* of a given set of data does not relate to the data subject, and the link must be found in its *purpose* or *result*, the identification of the data subject will necessarily have to happen through auxiliary information: the data is not *about* the natural person in itself[93]. Since direct identifiers relate to the data subject due to their content, there is never the need to qualify them as relating to the data subject due to their purpose or result – the elements that justify the relational link are alternative, not cumulative. Data deemed as relating to a natural person due to its purpose or result, but not to its content, needs however to be tied to additional auxiliary information that render the person identified or identifiable to become personal.

In other words, the information must be actually relating to an identifiable data subject, not merely *relatable*.[94] The wording of the definition of personal data is clear: while the person can be just identifiable, the information must be necessarily related, not just relatable. The legislator explicitly avoided referring to the mere possibility of a relation, requiring an actual link between the information and the identified or identifiable natural person. If the link is satisfied through the content element, then of course the information and the natural person will be related, not just relatable – the information is about the person in the most literal sense. Even when the relational link is integrated by the purpose or result elements, the data will still have to be actually related to a specific person, rather than just relatable. That must necessarily happen through additional auxiliary information that makes the data subject individually distinct from the group of individuals to which the information is just relatable.

Data protection is an individual right,[95] its collective dimension currently marginal. It is meant to protect individuals, rather than the groups to which they belong. It does not protect all subjects to which an attribute (e.g. affluence, or being a single mother) pertains, as a collectivity, but does shield them from the moment where they become personally identifiable. Therefore, even if any information can possibly relate to natural persons due to its purpose or its result, despite the fact that its content does not relate to a natural person by itself, the data *always* needs to be actually tied to auxiliary information connecting it with the data subject interested by that concrete processing instance to become personal. Data about things may very well become personal data, but only when it can be tied through auxiliary data to an identified or identifiable natural person in a specific processing instance.

The interaction between the 'relating to' link and the identifiability requirement makes it so that the configuration of the 'relating to' nexus as satisfied by virtue of the purpose or result element, rather than by the content one, leads to the necessity of a higher amount of auxiliary information to integrate the notion of personal data. In a way, it functions as a system to differentiate data about *people* from data about *a person*, and hence data which is encompassed by the material scope of EU data protection law from information that is not. Granted, however, data about people can become data about a person, if so processed: the qualification of data as personal (or, conversely, anonymous) must be performed considering the information within its lifecycle, rather than statically.

4.2.1 Data lifecycle

If any information can somehow relate to natural persons, and identifying someone is gradually becoming easier and easier, then all data can potentially be personal, and data protection's scope becomes gargantuan. Such concern is legitimate, and a valid argument to make when questioning the regulatory structure and mechanics on which EU data protection is based; less so when applied to the concrete determination of what personal data within an actual processing instance is. Even if it is true that all data can become personal data, that does not mean that such data is personal all the time. Information relating to a natural person by virtue of the purpose or result elements will not necessarily be personal throughout its entire lifecycle.

Data has indeed a lifecycle:[96] it gets created, collected, processed, re-shaped, aggregated, stored, and eventually deleted. When it links to a natural person through its content, it will likely do so throughout its lifecycle, unless the content changes. When the information is linked to the natural person only through the purpose or result elements, it will relate to them just for a specific segment of its lifecycle. A passport number, data whose content relates to an individual, will keep relating to that individual throughout its entire lifecycle. Data about the performance of a vehicle will relate to an individual only if and when it has a purpose or result actually involving that particular person, for instance when it is used to evaluate a driver's performance. Likewise, let us assume that a public administration shares information with the public, for instance public transport information, which then gets used by another party (e.g. an employer) to generate particular consequences involving an identifiable individual (perhaps to see whether an employee's unpunctuality can be blamed on public transport). Such data could be personal, in connection with an individual's identifiers, only from the moment when it is used to generate an effect involving such individual (i.e. when it relates to him through the result element), since nor the information's content nor its purpose relate to natural persons.

In the processing instances where the relational link is justified through the purpose or the result element, rather than through the content one, the information considered will not necessarily[97] qualify as personal from its inception to its erasure, but only if and to the extent to which it is meant (or likely) to be used to generate effects involving specific natural persons. Moreover, as highlighted in the previous section, it will still need to be processed in connection with additional information that can lead to their identification to be qualifiable as personal. I believe it would be preferable to shift the general focus of the debate around the inflation of the material scope of data protection law from the static notion of personal data to the *processing* of personal data. Data protection law does not indeed apply to personal data in itself, but to its processing. As a naturally relational concept, personal data is better understood when considered, dynamically, within the concrete processing instance evaluated. The purported expansive effects deriving from the combination between the low identifiability threshold and the wide range of ways information

can relate to a natural person, as set by EU data protection law and jurisprudence, seem to be much less substantial when considering personal data that relate to the data subject by virtue of their purpose or result within their lifecycle, rather than statically.

4.2.2 Attribute protection

The core purpose of data protection is safeguarding individuals from the harms that might derive from unfair information processing.[98] The notion of personal data, on which EU data protection is based, is meant to be potentially applicable to any kind of data, since any information can be hypothetically used to generate harmful consequences affecting human beings. In other words, data protection law does not aim at protecting individuals only from the misuse of their identities, but from the misuse of their attributes[99] – their characteristics and defining traits – too. From this perspective, the fact that any data can become personal data is a regulatory 'feature', rather than a 'bug', necessary to provide the flexible, contextual protection in context EU data protection is meant to afford.

The concept of personal data must be framed diachronically: the exact same piece of information can be anonymous or personal depending on the context, actors, and time of processing.[100] Attributes, when the natural person they are linked to is not identifiable, are not directly covered by EU data protection legislation: they are data about people, rather than data about a person. That remains true up to a certain point in the information's lifecycle – the application of the inference to a specific natural person – after which data protection law kicks in.

Statistical information – e.g. an anonymised, aggregated dataset – for instance does not allow the singling out of specific individuals from the sample of reference. Its proper sanitization prevents identity inference, impeding that an attacker identifies a specific individual within the record. Depending on the concrete circumstances of the case, the attacker can however often infer meaningful attributes about a data subject, without being able to identify the record linked with that specific individual. EU data protection law, as an individual right, does not safeguard from group inferences *per se*,[101] but does grant attribute protection as soon as the processing purpose or result relate to a *specific* natural person, without limiting itself to identity protection. It does so by considering personal data also information whose content does not directly relate to natural persons, but that is still tied, through its purpose or result, to specific individuals through auxiliary information that render them identifiable. For instance, aggregated data showing the average income in my neighbourhood cannot be personal on its own: I cannot be singled out from the aggregated record, and statistical information is merely *relatable* to me, not actually related. As soon as the very same information gets linked with additional identifiers to be used with a particular purpose or result involving me as a specific identifiable individual – for instance, if it gets used as a proxy for credit rating – such data does however become personal.[102]

Any data can theoretically generate informational harm: not just the information whose content refers to an individual, but whichever data can be used with a particular purpose or to generate relevant consequences for the data subject. As the Article 29 Working Party states,[103] the capacity to infer information about a person (i.e. one of her attributes) must be considered when assessing the level of protection anonymization should afford, along with the possibility to single her out or to link different records concerning that person.[104] The mere possibility to infer an attribute applicable to a group (e.g. the average income in a certain area code) does not however render the data personal by itself: it merely makes it relatable to a person. The inference becomes personal when it begins to be actually related to the data subject, which requires auxiliary information leading to that specific person's direct or indirect identifiability. EU data protection is

thus meant to safeguard whichever attribute might characterise a person, but only from the moment where the group inference starts to be related to a *specific* identifiable person.[105]

4.3 Modulating identifiability

The identifiability test adopted by European data protection law and jurisprudence is, potentially, very broad. Natural persons are deemed identifiable if the controller or another person can single them out, directly or indirectly, through all means likely reasonably to be used – i.e. unless those means are practically impossible or illegal[106] – in consideration of the state of the art. Individuals can theoretically be identified through several means and by many actors. At the same time, singling people out from a group has arguably never been easier for a motivated attacker. Competent authorities, for instance, are well within the meaning of 'another person', and have often the power to compel the production of the auxiliary information necessary to tie all sort of data to specific natural persons. An order from a competent authority will thus often be a 'means likely reasonably' to be used to identify the natural person to which the information relates. This logically leads to a low threshold after which a person is to be considered as legally identifiable.

This should not however lead an 'identifiability assumption', where – *in* case of doubt – individuals are to be considered as identifiable by default under the GDPR. While the necessity to avoid elusion mandates the adoption of a low and flexible identifiability thresholds, EU data protection law, jurisprudence, and doctrine, also provide a number of criteria and considerations through which to modulate the identifiability test.

4.3.1 Means to identify and reasonable likelihood

Determining identifiability must be done on a case-by-case basis, for each concrete processing instance, through the interpretative criteria made available by the law and its jurisprudence. The mere theoretical possibility of identification is not sufficient to render someone identifiable: such possibility must be concrete, modelled according all objective factors of the processing.[107] The 'means likely reasonably to be used by the data controller or by another person' should be interpreted in light of the functioning of EU data protection law, which singles out each processing instance by reference to the data processed, the actors involved, and the purposes of the processing.

To ascertain whether means are reasonably likely to be used to identify the natural person, one should consider all objective factors pertaining to the processing:[108] account must be taken of 'all' the means 'likely reasonably' to be used for identification by the controller and third parties, paying special attention to the current state of technology, and the constant increase in computing power, know-how, and tools available.[109] However, mere theoretical chance is not sufficient to deem a person identifiable; if, taking into account 'all the means likely reasonably to be used by the controller or any other person', that possibility does not exist or is negligible, the person should not be considered as identifiable.[110] That is the case if the identification of the data subject is prohibited by law, or practically impossible, which happens when the risk of identification is deemed, in reality, to be insignificant.[111] 'Objective' factors, I believe, should be interpreted as referring to the contextual and environmental elements of the processing, rather than to the mere existence of the means that can be used to identify a person.

To be clear, I am not arguing for a switch to a paradigm where identifiability depends on the subjective, relative capacity of the entity performing the identification attempt – the controller or 'another person'. I believe, however, that artificially removing contextual considerations from the

identifiability assessment can lead to perverse results, as much as basing the assessment on the capabilities of the controller alone would render data protection law too easily avoidable. The objectivity of the factors to be considered when determining identifiability should be referred to the environment in which the controller operates: the CJEU does indeed refer to *practical* impossibility,[112] not to impossibility *tout court*, when discussing the means that are not likely reasonably to be used to identify the data subject.

The GDPR exemplifies the factors to be considered when determining a person's identifiability by referring to the costs of and the amount of time required for identification, taking into consideration the state of the art.[113] Other important factors are, for instance, the intended purpose of the processing, the way it is structured, the advantage expected by the controller and the interests at stake for individuals, as well as the risk of organisational and technical failures.[114] Indeed, the subjective motivation of the attacker might impact the identifiability threshold: some datasets will be more desirable than others. For example, information might be more likely to lead to the identification of the data subject if it has a significant commercial value, or if it can be used for law enforcement or intelligence purposes. Attackers might be also drawn if the information reveals newsworthy information about public figures, or if it can be used for political purposes, or even if it could just raise curiosity.[115]

The availability of auxiliary data that might be used to identify the data subject must be contemplated too, both in terms of its public availability as well as considering how certain attackers might have access to privileged information that allow to identify individuals. Identification risks increase where someone is likely to know a large amount of auxiliary data about a person, such as family members, colleagues, doctors, or other professionals. The auxiliary data needed to perform re-identification could of course also be information available to businesses or organisations, including law enforcement, judicial authorities, or other public-sector bodies; it could also be published on the Internet, available to everyone.[116]

The identifiability test is meant to be dynamic, and should consider both the state of the art at the time of the processing and the possibilities for future development during the foreseen processing period. Storage time is a paramount factor in determining identifiability's threshold. What may not be likely reasonably possible today, or in a month, might become feasible in a decade. Identification must be put in relation to the information's lifecycle, and thus the controller should consider the possibility of future identification, which may make the data personal (from that moment on, not retroactively). Data controllers should stay aware of relevant developments, and enact the necessary technical and organisational measures.

Technical and organisational considerations are particularly important in determining the factors that may render the data subject identifiable, and thus the controller might consider testing the likelihood that the identification, if attempted, would be successful.[117] In some circumstances it can be difficult to establish such risk, particularly where complex statistical methods might be used by a third party to match various pieces of anonymised data. It can be good practice to perform a re-identification test to ascertain the potential for re-identification, attempting to re-identify individuals from the datasets that are being assessed.[118] Such 'motivated intruder'[119] test essentially involves considering whether an attacker would be capable to achieve re-identification, if adequately motivated. The approach assumes that the attacker is motivated, competent, and has access to resources commensurate with the motivation it may have for the re-identification.[120]

The criteria that have been provided to help in the interpretation of the notion of personal data, particularly with respect to the identifiability test, seem to mirror what has been defined, in the

field of statistical confidentiality, as the 'data environment':[121] the set of formal and informal structures, processes, mechanisms and agents that act on, define, control, or interact with data, or provide interpretable context for such data. Each data environment has been framed as consisting of four components: data, agency, governance, and infrastructure.[122] The data element considers the information available within the environment; the agency one models how agents might act on and in the environment; the governance element identifies the discipline of the users' interaction with the data; and the infrastructure component considers the physical and logical structures and processes that regulate the data environment. While the concept of data environment pertains to the statistical confidentiality milieu, it might well provide for a way to formalise identifiability – and perhaps the status of information as personal *tout court* – within the context of the right to personal data protection.[123]

4.3.2 From 'any other person' to 'another person'

The interpretation of what constitutes a means reasonably likely to identify a person is not the only factor that can lead to the inflation of the identifiability component of the notion of personal data. The fact that identifiability must be tested against the means available to both the controller and 'another person' can also lead to an excessively low identification threshold, and hence, in conjunction with the width of the 'relating to' link, to the disproportionate width of the definition of personal data. In this regard, I hold that the 'another person' Recital 26 of the GDPR refers to should be interpreted teleologically:[124] it should be 'likely reasonably' for the controller or the processor to have access to the third party possessing the auxiliary information that permit the identification of the data subject.[125] The mere existence of a third party with powerful means to which the controller cannot actually have access should not be considered as 'likely reasonably' to be used to identify a person.

To determine whether a natural person is identifiable, the GDPR mandates the consideration of all the means reasonably likely to be used, either by the controller or by another person, to identify the data subject, directly or indirectly.[126] For data to be personal, it is not necessary that it alone identifies the data subject; it is not required for all the information enabling the identification of the data subject to be in the hands of a single entity, either[127] – as long as access to that entity is reasonably likely for the data controller. Indeed, a strict literal interpretation of the concept of personal data could expand its scope to all information all the time, regardless of the information's inability to reveal the data subject on its own: '(i)t would never be possible to rule out, with absolute certainty, the possibility that there is no third party in possession of additional data which may be combined with that information and are, therefore, capable of revealing a person's identity'.[128]

Just as the means that may be used by the controller must be constrained to the likely reasonably ones only, so the third parties who may be approached by a controller to identify the data subject should be understood as the ones that can likely reasonably be accosted. Reasonable likelihood would not occur when contact with those third parties is exceedingly costly, considering both human and economic capital, practically impossible, or prohibited by law.[129] Such an interpretation has been put forth by Advocate General Campos Sànchez-Bordona in the Breyer v Germany case; while the CJEU did not reject it, it did not explicitly confirm its validity either. I believe clarifying that the same 'reasonable likelihood' test applicable to the means for identification could (and should) be applied to the 'any other person' that can identify the data subject in lieu of the data controller would have provided valuable guidance from the CJEU.

In this sense, it seems significative to point out a notable difference between Recital 26 of the

GDPR, and its predecessor, Recital 26 of the DPD. While Recital 26 of the DPD specified that, to determine whether a person is identifiable, one should consider 'all the means likely reasonably to be used either by the controller or by *any other* person', Recital 26 of the GDPR indicates that 'account should be taken of all the means reasonably likely to be used […] either by the controller or by *another* person'. That would seem a meaningful semantic difference. The syntagms 'any other person' and 'another person' are not equivalent: the former conveys that the means to be considered when determining the identifiability of a data subject can be available to *any* person; the latter clarifies that such means do not necessarily have to be available to the controller, but can be available to *another* person too. It is still too early to see how courts will interpret this lexical change, but moving from the postulation that the concept of personal data in the GDPR mirrors the one in the DPD, it is reasonable to assume that the legislature meant to clarify that there are situations where the fact that a person is identifiable by a particularly resourceful controller does not make that person identifiable by *any* controller by default.

The breadth of the identifiability element can thus be tempered by a number of considerations and parameters which, while guiding in the interpretation of the concept of personal data, also contribute in narrowing it down to a reasonable dimension. The flexibility and contextual nature of the notion of personal data requires a case-by-case approach to determining whether information is personal or not. The discretion of the interpreter is however bound by a number of factors that, if followed, would prevent outcomes deviating from the *ratio legis* of EU data protection law.

5. Conclusion

The combination between the low identifiability threshold and the multiple ways in which information can relate to natural persons render the notion of personal data potentially very broad. Individuals are ubiquitously tracked through the devices they carry, and advances in technology and research make identification a progressively easier endeavour. At the same time, information can be linked to (identifiable) natural persons in many ways – i.e. through its content, its purpose, or its result. The notion of personal data, and the identifiability/anonymity dichotomy on which it is based, can thus be seen as overly fuzzy and expansive concepts. Such concern is more than legitimate from a regulatory perspective, thinking *de iure condendo*. However, focusing on it while determining, *de iure condito*, what constitutes processing of personal data, can lead to perverse results. Law, jurisprudence, and doctrine, nonetheless, provide enough exegetic tools to ensure that the concept of personal data remains flexible enough not to deprive individuals from their right to data protection, while still sufficiently narrow as not to cover all data all the time.

I have argued that the potential over-inflation of the concept of personal data within the current regulatory framework is due, in particular, to the possibility to interpret too extensively the identifiability requirement and the meaning of 'relating to'. I have also argued, however, that the law and the related doctrine and jurisprudence also contains elements that can permit to avoid an overly extensive interpretation. The GDPR embeds a certain degree of flexibility in its application, which reflects also on the notion of personal data and on the elements that compose it. The potential width of the 'relating to' link is tempered by its interaction with the identifiability requirement, particularly if one considers (personal) data within its lifecycle, and the fact that EU data protection law is meant to regulate the processing of individual attributes too, not just identifiers. On the other hand, the expansive effect of the identifiability requirement can be reined in by applying the plethora of criteria devised to perform the 'reasonable likely' test not only to the means that can be used to identify the data subject, but also on the persons to whom those means can be available, and on their relationship with the controller.

Despite its indeterminate and relative character, I believe that the notion of personal data can still be a viable basis for policy and legislation. A nuanced epistemological approach to determining what constitutes personal data, through the appropriate interpretative criteria, either mitigates the dreaded effects of an overextension of the material scope of EU data protection law, or justifies the rationale for which information should be considered personal.

Acknowledgements

This research was performed with the financial support of the Dutch STW-Maps4Society program (project number 13718). The author would like to thank Nadya Purtova, Eleni Kosta, Bastiaan van Loenen, and the anonymous reviewers for their helpful comments.

Bibliography

Aldhouse F, 'Anonymisation of Personal Data – A Missed Opportunity for the European Commission' (2014) 30 Computer Law & Security Review 403

Ambrose ML, 'It's About Time: Privacy, Information Life Cycles, and the Right to Be Forgotten' (2012) 16 Stan. Tech. L. Rev. 369

Arkhipov V and Naumov V, 'The Legal Definition of Personal Data in the Regulatory Environment of the Russian Federation: Between Formal Certainty and Technological Development' (2016) 32 Computer Law & Security Review 868 <https://www.sciencedirect.com/science/article/pii/S0267364916301236>

Article 29 Data Protection Working Party, 'Opinion 4/2007 on the Concept of Personal Data WP136' (2007)

——, 'Opinion 06/2013 on Open Data and Public Sector Information ('PSI') Reuse WP207' (2013)

——, 'Opinion 05/2014 on Anonymisation Techniques WP216' (2014)

——, 'Statement 14/EN WP 218 on the Role of a Risk-Based Approach in Data Protection Legal Frameworks' (2014)

Bratton BH, *The Stack: On Software and Sovereignty* (MIT Press 2016)

Brickell J and Shmatikov V, 'The Cost of Privacy: Destruction of Data-Mining Utility in Anonymized Data Publishing', *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining* (ACM 2008)

Bygrave LA, 'Information Concepts in Law: Generic Dreams and Definitional Daylight' (2015) 35 Oxford Journal of Legal Studies 91

Christl W, Kopp K and Riechert PU, 'Corporate Surveillance in Everyday Life' (2017) <http://crackedlabs.org/en/corporate-surveillance>

Christl W and Spiekermann S, *Networks of Control* (Facultas 2016) <http://www.privacylab.at/wp-content/uploads/2016/09/Christl-Networks__K_o.pdf>

Culnane C, Rubinstein B and Teague V, 'Health Data in An Open World: A Report on Re-Identifying Patients in The Mbs/Pbs Dataset and the Implications for Future Releases of Australian Government Data' (2017) <https://regmedia.co.uk/2017/12/17/report_on_deidentification.pdf> accessed 11 January 2018

de Montjoye Y-A and others, 'Unique in the Crowd: The Privacy Bounds of Human Mobility' (2013) 3 Scientific Reports 1376 <http://dx.doi.org/10.1038/srep01376>

de Montjoye Y-A, Radaelli L and Singh VK, 'Unique in the Shopping Mall: On the Reidentifiability of Credit Card Metadata' (2015) 347 Science 536

Dwork C, 'Differential Privacy', *33rd International Colloquium on Automata, Languages and Programming, part II (ICALP 2006)* (Springer Verlag 2006) <https://www.microsoft.com/en-us/research/publication/differential-privacy/>

Dwork C and Naor M, 'On the Difficulties of Disclosure Prevention in Statistical Databases or the Case for Differential Privacy' (2008) 2 Journal of Privacy and Confidentiality 8

Easterbrook FH, 'Legal Interpretation and the Power of the Judiciary' (1984) 7 Harv. JL & Pub. Pol'y 87

El Emam K and Alvarez C, 'A Critical Appraisal of the Article 29 Working Party Opinion 05 / 2014 on Data Anonymization Techniques' (2014) 5 International Data Privacy Law 1

Elliot M and others, 'Functional Anonymisation: Personal Data and the Data Environment' [2018] Computer Law & Security Review

Elliot M and Mackey E, 'The Social Data Environment' in Keiron O'Hara, MH Carolyn Nguyen and Peter Haynes (eds), *Digital Enlightenment Yearbook* (2014)

Gellert R, 'Data Protection: A Risk Regulation? Between the Risk Management of Everything and the Precautionary Alternative' (2015) 5 International Data Privacy Law

——, 'Understanding Data Protection As Risk Regulation' (2015) 18 Journal of Internet Law 3

——, 'We Have Always Managed Risks in Data Protection Law: Understanding the Similarities and Differences Between the Rights-Based and the Risk-Based Approaches to Data Protection' (2016) 2 European Data Protection Law Review 481 <https://doi.org/10.21552/EDPL/2016/4/7>

Greenfield A, *Everyware: The Dawning Age of Ubiquitous Computing* (New Riders 2010)

Harbinja E, 'Does the EU Data Protection Regime Protect Post-Mortem Privacy and What Could Be the Potential Alternatives' (2013) 10 SCRIPTed 19

Hildebrandt M, *Smart Technologies and the End (s) of Law: Novel Entanglements of Law and Technology* (Edward Elgar Publishing 2015)

——, 'Law as Information in the Era of Data-Driven Agency' (2016) 79 The Modern Law Review 1

Kitchin R and Dodge M, *Code/Space: Software and Everyday Life* (Mit Press 2011)

Koops B-J, 'The Trouble with European Data Protection Law' (2014) 4 International Data Privacy Law 250 <http://dx.doi.org/10.1093/idpl/ipu023>

Leenes R, 'Do They Know Me? Deconstructing Identifiability' (2008) 4 University of Ottawa Law & Technology Journal 135

Mackey E and Elliot M, 'Understanding the Data Environment' (2013) 20 XRDS: Crossroads 36

Mantelero A, 'Personal Data for Decisional Purposes in the Age of Analytics: From an Individual to a Collective Dimension of Data Protection' (2016) 32 Computer Law & Security Review 238 <https://www.sciencedirect.com/science/article/pii/S0267364916300280> accessed 2 February 2018

——, 'From Group Privacy to Collective Privacy: Towards a New Dimension of Privacy and Data Protection in the Big Data Era' in Linnet Taylor, Luciano Floridi and Bart van der Sloot (eds), *Group Privacy* (Springer 2017)

Millard C and Hon WK, 'Defining "personal Data" in e-Social Science' (2012) 15 Information, Communication & Society 66

Mourby M and others, 'Are "Pseudonymised"Data Always Personal Data? Implications of the GDPR for Administrative Data Research in the UK' [2018] Computer Law & Security Review

Narayanan A and Felten EW, 'No Silver Bullet: De-Identification Still Doesn't Work' (2014)

Narayanan A and Shmatikov V, 'Robust De-Anonymization of Large Datasets (How to Break Anonymity of the Netflix Prize Dataset)' (2008)

——, 'Myths and Fallacies of Personally Identifiable Information' (2010) 53 Communications of the ACM 24

Ohm P, 'Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization' (2009) 57 UCLA Law Review 1701

Ohm P and Peppet S, 'What If Everything Reveals Everything?' in Cassidy R Sugimoto, Hamid R Ekbia and Michael Mattioli (eds), *Big Data Is Not a Monolith* (MIT Press 2016)

Purtova N, 'The Law of Everything. Broad Concept of Personal Data and Future of EU Data Protection Law' [2018] Law, Innovation and Technology 1 <https://doi.org/10.1080/17579961.2018.1452176>

Rubinstein IS and Hartzog W, 'Anonymization and Risk' (2016) 91 Washington Law Review 1

Schwartz PM and Solove DJ, 'The PII Problem: Privacy and a New Concept of Personally Identifiable Information'

(2011) 86 New York University Law Review 1814

Schwartz PM and Solove DJ, 'Reconciling Personal Information in the United States and European Union' (2014) 102 California Law Review 877

Solove DJ and Schwartz PM, 'PII 2.0: Privacy and a New Approach to Personal Information' [2012] Privacy and Security Law Report 2011

Spicer K and others, 'Intruder Testing : Demonstrating Practical Evidence of Disclosure Protection in 2011 UK Census Intruder Testing : Demonstrating Practical Evidence of Disclosure Protection in 2011 UK Census' 28

Sweeney L, 'Simple Demographics Often Identify People Uniquely' (2000) 3 Carnegie Mellon University, Data Privacy Working Paper

Sweeney L, Abu A and Winn J, 'Identifying Participants in the Personal Genome Project by Name' (2013)

Sweeney L and Yoo JS, 'De-Anonymizing South Korean Resident Registration Numbers Shared in Prescription Data' [2015] Technology Science

Tudor C, Cornish G and Spicer K, 'Intruder Testing on the 2011 UK Census: Providing Practical Evidence for Disclosure Protection' (2014) 5 Journal of Privacy and Confidentiality 3

Urgessa WG, 'The Protective Capacity of the Criterion of "Identifiability" Under EU Data Protection Law' (2016) 2 European Data Protection Law Review

van Loenen B, Kulk S and Ploeger H, 'Data Protection Legislation: A Very Hungry Caterpillar: The Case of Mapping Data in the European Union' (2016) 33 Government Information Quarterly <http://www.sciencedirect.com/science/article/pii/S0740624X16300326>

Wu FT, 'Defining Privacy and Utility in Data Sets' (2012) 84 University of Colorado Law Review 1117

Yakowitz J, 'Tragedy of the Data Commons' (2011) 25 Harvard Journal of Law & Technology 1

Zuboff S, 'Big Other: Surveillance Capitalism and the Prospects of an Information Civilization' (2015) 30 Journal of Information Technology

Zuiderveen Borgesius FJ, 'Singling out People without Knowing Their Names–Behavioural Targeting, Pseudonymous Data, and the New Data Protection Regulation' (2016) 32 Computer Law & Security Review

---

[1] Tilburg Law School, TILT; TU Delft, A+BE

[2] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (GDPR). 4.5.2016, OJ L 119/1, Art. 4.1. See also Directive 95/46/EC 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 (DPD). 24.10.1995, OJ L 281/31, Art. 2(a); Council of Europe, Convention for the protection of individuals with regard to automatic processing of personal data, ETS no. 108 of 28 January 1981 (Convention 108), Art. 2(a).

[3] GDPR, Art. 2(1).

[4] Christopher Millard and W Kuan Hon, 'Defining "Personal Data" in e-Social Science' (2012) 15 Information, Communication & Society 66, 69.

[5] GDPR, Recital 26.

[6] E.g. Paul Ohm, 'Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization' (2009) 57 UCLA Law Review 1701; Arvind Narayanan and Edward W Felten, 'No Silver Bullet: De-Identification Still Doesn't Work' (2014); Arvind Narayanan and Vitaly Shmatikov, 'Myths and Fallacies of Personally Identifiable Information' (2010) 53 Communications of the ACM 24; Ira S Rubinstein and Woodrow Hartzog, 'Anonymization and Risk' (2016) 91 Washington Law Review 1; Francis Aldhouse, 'Anonymisation of Personal Data – A Missed Opportunity for the European Commission' (2014) 30 Computer Law & Security Review 403. *Contra*: Jane Yakowitz, 'Tragedy of the Data Commons' (2011) 25 Harvard Journal of Law & Technology 1. See also Felix T Wu, 'Defining Privacy and Utility in Data

Sets' (2012) 84 University of Colorado Law Review 1117.

[7] This paper defines attacker as the subject attempting to identify or re-identify an individual within a record, and the attack the activities performed by the attacker for such purposes.

[8] E.g. Latanya Sweeney, 'Simple Demographics Often Identify People Uniquely' (2000) 3 Carnegie Mellon University, Data Privacy Working Paper; Latanya Sweeney, Akua Abu and Julia Winn, 'Identifying Participants in the Personal Genome Project by Name' (2013); Latanya Sweeney and Ji Su Yoo, 'De-Anonymizing South Korean Resident Registration Numbers Shared in Prescription Data' [2015] Technology Science; Arvind Narayanan and Vitaly Shmatikov, 'Robust De-Anonymization of Large Datasets (How to Break Anonymity of the Netflix Prize Dataset)' (2008); Chris Culnane, Benjamin Rubinstein and Vanessa Teague, 'Health Data in An Open World: A Report on Re-Identifying Patients in The Mbs/Pbs Dataset and the Implications for Future Releases of Australian Government Data' (2017) <https://regmedia.co.uk/2017/12/17/report_on_deidentification.pdf> accessed 11 January 2018; Yves-Alexandre de Montjoye and others, 'Unique in the Crowd: The Privacy Bounds of Human Mobility' (2013) 3 Scientific Reports 1376 <http://dx.doi.org/10.1038/srep01376>; Yves-Alexandre de Montjoye, Laura Radaelli and Vivek Kumar Singh, 'Unique in the Shopping Mall: On the Reidentifiability of Credit Card Metadata' (2015) 347 Science 536.

[9] Article 29 Data Protection Working Party, 'Opinion 4/2007 on the Concept of Personal Data WP136' (2007) 6.

[10] Nadezhda Purtova, 'The Law of Everything. Broad Concept of Personal Data and Future of EU Data Protection Law' [2018] Law, Innovation and Technology 1 <https://doi.org/10.1080/17579961.2018.1452176>.

[11] Bert-Jaap Koops, 'The Trouble with European Data Protection Law' (2014) 4 International Data Privacy Law 250 <http://dx.doi.org/10.1093/idpl/ipu023>; Bastiaan van Loenen, Stefan Kulk and Hendrik Ploeger, 'Data Protection Legislation: A Very Hungry Caterpillar: The Case of Mapping Data in the European Union' (2016) 33 Government Information Quarterly <http://www.sciencedirect.com/science/article/pii/S0740624X16300326>; Purtova (n 9).

[12] See e.g. Ohm (n 5); Koops (n 10); Purtova (n 9); Daniel J Solove and Paul M Schwartz, 'PII 2.0: Privacy and a New Approach to Personal Information' [2012] Privacy and Security Law Report 2011; Paul M Schwartz and Daniel J Solove, 'The PII Problem: Privacy and a New Concept of Personally Identifiable Information' (2011) 86 New York University Law Review 1814.

[13] GDPR, Art. 4(1). See also DPD, Art. 2(a); Convention 108, Art. 2(a).

[14] See Court of Justice of the European Union, Lindqvist, C-101/01, 6 November 2003, ECLI:EU:C:2003:596, s 27.

[15] Article 29 Data Protection Working Party, 'Opinion 4/2007 on the Concept of Personal Data WP136' (n 8) 6.

[16] The GDPR, as the DPD before it, uses data and information interchangeably. On the 'underdeveloped, if not poor' understanding of information in legal scholarship, see Lee A Bygrave, 'Information Concepts in Law: Generic Dreams and Definitional Daylight' (2015) 35 Oxford Journal of Legal Studies 91, 91.

[17] Bygrave (n 15) 113.

[18] '(A)s the law stands, and depending on the theoretical perspective towards the meaning of information, everything can still be plausibly argued either to be or to contain information that can be personal data provided the other requirements of the definition are met': Purtova (n 9) 27.

[19] The GDPR does not apply to the personal data of deceased persons; however, Member States are allowed to regulate with national law the processing of personal data of deceased persons (GDPR, Recital 27). On the topic, see Edina Harbinja, 'Does the EU Data Protection Regime Protect Post-Mortem Privacy and What Could Be the Potential Alternatives' (2013) 10 SCRIPTed 19.

[20] Article 29 Data Protection Working Party, 'Opinion 4/2007 on the Concept of Personal Data WP136' (n 8) 9 ss. See also Court of Justice of the European Union, Peter Nowak v Data Protection Commissioner, C-434/16, 20 December 2017, ECLI:EU:C:2017:994, which expands Court of Justice of the European Union, YS (C-141/12) v Minister voor Immigratie, Integratie en Asiel, and Minister voor Immigratie, Integratie en Asiel v M, S, Joined Cases C-141/12 and C-372/12, 17 July 2014, ECLI:EU:C:2014:2081.

[21] Article 29 Data Protection Working Party, 'Opinion 4/2007 on the Concept of Personal Data WP136' (n 8) 10.

[22] Article 29 Data Protection Working Party, 'Opinion 4/2007 on the Concept of Personal Data WP136' (n 8) 10. See also Peter Nowak v Data Protection Commissioner (n 19) s 37; 43.

[23] Article 29 Data Protection Working Party, 'Opinion 4/2007 on the Concept of Personal Data WP136' (n 8) 10. See also Peter Nowak v Data Protection Commissioner (n 19) s 38; 43.

[24] Article 29 Data Protection Working Party, 'Opinion 4/2007 on the Concept of Personal Data WP136' (n 8) 11. See also Peter Nowak v Data Protection Commissioner (n 19) s 39; 43.

[25] While the A29WP's Opinion refers to the concept of personal data in the Data Protection Directive, the notion did not change with the enactment of the GDPR (see Opinion of AG Kokott in case Peter Nowak v Data Protection Commissioner, C-434/16, 20 July 2017, ECLI:EU:C:2017:582, s 3), and what the Opinion holds is expected to remain valid: Purtova (n 9) 6.

[26] Article 29 Data Protection Working Party, 'Opinion 4/2007 on the Concept of Personal Data WP136' (n 8) 9 ss.

[27] CJEU, YS and others (n 19); Peter Nowak v Data Protection Commissioner (n 19).

[28] Court of Justice of the European Union, Jehovan todistajat, C-25/17, 10 July 2018, ECLI:EU:C:2018:551, s 21: 'it is clear from Opinion 1/2010 of 16 February 2010 on the concepts of 'controller' and 'processor' produced by the Working Group set up pursuant to Article 29 of Directive 95/46, that, in particular, the 'effective control' and the conception that the data subject has of the controller must be taken into account'. While case C-25/17 is to my knowledge the first time where the Court referred to an A29WP opinion explicitly, implicit references can be found in other cases, e.g. YS and others, Nowak, and Google Spain. Furthermore, several Opinions of the Attorney General, which have been relied upon by the Court, explicitly refer to A29WP opinions.

[29] GDPR, Art. 4.1; Recital 26. See also Article 29 Data Protection Working Party, 'Opinion 4/2007 on the Concept of Personal Data WP136' (n 8) 12.

[30] For an original taxonomy on identifiers in data protection, see Ronald Leenes, 'Do They Know Me? Deconstructing Identifiability' (2008) 4 University of Ottawa Law & Technology Journal 135. The definition of identifiers this paper relies on is taken from ISO/TS 25237:2008 (Health informatics — Pseudonymization).

[31] See CJEU, Lindqvist (n 13) s 27.

[32] 'The use by the EU legislature of the word 'indirectly' suggests that, in order to treat information as personal data, it is not necessary that that information alone allows the data subject to be identified': Court of Justice of the European Union, Patrick Breyer v Bundesrepublik Deutschland, Case C-582/14, 19 October 2016, ECLI:EU:C:2016:779, s 41. See also Article 29 Data Protection Working Party, 'Opinion 4/2007 on the Concept of Personal Data WP136' (n 8) 13.

[33] GDPR, Recital 26; Article 29 Data Protection Working Party, 'Opinion 4/2007 on the Concept of Personal Data WP136' (n 8) 15.

[34] GDPR, Recital 26; CJEU Breyer (n 31) s 43; Article 29 Data Protection Working Party, 'Opinion 4/2007 on the Concept of Personal Data WP136' (n 8).

[35] GDPR, Recital 26; Article 29 Data Protection Working Party, 'Opinion 4/2007 on the Concept of Personal Data WP136' (n 8). See also CJEU Breyer (n 31) s 43.

[36] Court of Justice of the European Union, Rijkeboer, C-553/07, 7 May 2009, EU:C:2009:293, s 59.

[37] Article 29 Data Protection Working Party, 'Opinion 4/2007 on the Concept of Personal Data WP136' (n 8).

[38] CJEU, YS and others (n 19); Peter Nowak v Data Protection Commissioner (n 19).

[39] Purtova (n 9) 3. Purtova's article is, at the time of writing, the main scholarly critique to the current doctrinal and judicial construction of what 'relating to' means, and of its consequences for the tenability of the notion of personal data.

[40] Purtova (n 9) 16.

[41] Purtova (n 9) 16.

[42] Purtova (n 9) 17.

[43] Purtova (n 9) 17.

[44] Purtova (n 9) 19.

[45] The author mentions influencing people's behaviour as a purpose of processing, by e.g. adapting the environment's lighting on the basis of the data processed, to modify deviant behaviour - Purtova (n 9) 19.

[46] The author exemplifies by reference to knowledge discovery in databases (KDD): 'In the context of a large knowledge-discovery database built for advanced data analytics, by several different public and private parties with

varying interests […] the weather information will still be relating to people in impact' - Purtova (n 9) 19.

[47] Purtova (n 10) 17, citing Mireille Hildebrandt, 'Law as Information in the Era of Data-Driven Agency' (2016) 79 The Modern Law Review 1, 4. The reference is at what Hildebrandt calls 'onlife', 'pre-emptive computing systems that calculate our inferred future behaviours and engage in pervasive and continuous adaptations of our life word': Mireille Hildebrandt, *Smart Technologies and the End (s) of Law: Novel Entanglements of Law and Technology* (Edward Elgar Publishing 2015) 15; 41 ss.

[48] Article 29 Data Protection Working Party, 'Opinion 4/2007 on the Concept of Personal Data WP136' (n 8).

[49] See *supra* (n 7, 11).

[50] Worku Gedefa Urgessa, 'The Protective Capacity of the Criterion of "Identifiability" Under EU Data Protection Law' (2016) 2 European Data Protection Law Review 522.

[51] As the Court does in CJEU Breyer (n 31) s 47.

[52] GDPR, Recital 26.

[53] CJEU Breyer (n 31) s 46.

[54] 'For years, it was widely believed that as long as data sets were 'anonymized,' they posed no risk to anyone's privacy. […] Unfortunately, the notion of perfect anonymisation has been exposed as a myth': Rubinstein and Hartzog (n 5) 704.

[55] See Wolfie Christl and Sarah Spiekermann, *Networks of Control* (Facultas 2016) <http://www.privacylab.at/wp-content/uploads/2016/09/Christl-Networks__K_o.pdf>; Wolfie Christl, Katharina Kopp and Patrick Urs Riechert, 'Corporate Surveillance in Everyday Life' (2017) <http://crackedlabs.org/en/corporate-surveillance>. See also Shoshana Zuboff, 'Big Other: Surveillance Capitalism and the Prospects of an Information Civilization' (2015) 30 Journal of Information Technology.

[56] See Adam Greenfield, *Everyware: The Dawning Age of Ubiquitous Computing* (New Riders 2010).

[57] See Rob Kitchin and Martin Dodge, *Code/Space: Software and Everyday Life* (Mit Press 2011); Benjamin H Bratton, *The Stack: On Software and Sovereignty* (MIT Press 2016).

[58] See Article 29 Data Protection Working Party, 'Opinion 05/2014 on Anonymisation Techniques WP216' (2014).

[59] Ohm (n 5). After dealing with the purported failure of anonymization, and thus of identity protection, in a following contribution the author also muses about the future of attribute inference: see Paul Ohm and Scott Peppet, 'What If Everything Reveals Everything?' in Cassidy R Sugimoto, Hamid R Ekbia and Michael Mattioli (eds), *Big Data Is Not a Monolith* (MIT Press 2016).

[60] Latanya Sweeney's work, the Netflix re-identification attack (see Narayanan and Shmatikov (n 9)) and the AOL de-identification debacle (see Ohm (n 7)).

[61] This has been defined as 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': Ohm (n 5) 1746.

[62] Narayanan and Felten (n 5).

[63] Narayanan and Felten (n 5). *Contra*: 'Linkage attacks, however, are much more complicated than they sound' - Rubinstein and Hartzog (n 5) 711.

[64] See Wu (n 5); Justin Brickell and Vitaly Shmatikov, 'The Cost of Privacy: Destruction of Data-Mining Utility in Anonymized Data Publishing', *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining* (ACM 2008); Mark Elliot and others, 'Functional Anonymisation: Personal Data and the Data Environment' [2018] Computer Law & Security Review.

[65] See Cynthia Dwork, 'Differential Privacy', *33rd International Colloquium on Automata, Languages and Programming, part II (ICALP 2006)* (Springer Verlag 2006) <https://www.microsoft.com/en-us/research/publication/differential-privacy/>; Cynthia Dwork and Moni Naor, 'On the Difficulties of Disclosure Prevention in Statistical Databases or the Case for Differential Privacy' (2008) 2 Journal of Privacy and Confidentiality 8.

[66] Simply put, attackers can draw from auxiliary data, extraneous from the anonymized dataset of reference; such auxiliary data, which cannot be modelled *ex ante*, can be cross-correlated with a purportedly anonymized dataset,

potentially de-anonymizing the individuals to which the dataset refers: Dwork (n 64) 2. Dwork proved that it is impossible, for a statistical database, to ensure that an attacker cannot learn, through the database, anything about an individual that could not be learned without access to such database: 'in any 'reasonable' setting there is a piece of information that is in itself innocent, yet in conjunction with even a modified (noisy) version of the data yields a privacy breach' - Dwork and Naor (n 64) 93.

[67] GDPR, Recital 26; see also DPD, Recital 26. With respect to the DPD, Urgessa claims that 'any other person' means that a person is to be considered identifiable 'if only another person than the controller is able to link a person to a data' - Urgessa (n 49) 522. The wording of Recital 26 of the GDPR, however, refers to 'another person', not to 'any other person' anymore: see section 4.3. below.

[68] CJEU Breyer (n 31) s 46.

[69] With regard to identification, in the Breyer case, the CJEU adopted an objective (or absolute) criterion, where the natural person is deemed as identifiable if any subject can do so, and rejected the subjective (or relative) criterion, for which a person is deemed identifiable if the data controller can identify her by relying only on its own capacity. The Russian Federation, in contrast, reportedly adopted the subjective criterion: despite the fact that the personal data definition adopted by the Russian Federation corresponds to the one set by Convention 108, Russian judicial practice indicates that, when determining identifiability, courts consider only whether the data subject is identifiable by the controller alone: see Vladislav Arkhipov and Victor Naumov, 'The Legal Definition of Personal Data in the Regulatory Environment of the Russian Federation: Between Formal Certainty and Technological Development' (2016) 32 Computer Law & Security Review 868 <https://www.sciencedirect.com/science/article/pii/S0267364916301236>.

[70] See Urgessa (n 49) 529.

[71] The CJEU ruled that the means to identify a data subject would not be considered as likely reasonably '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': CJEU Breyer (n 31) s 46.

[72] See *supra* n 68.

[73] Frank H Easterbrook, 'Legal Interpretation and the Power of the Judiciary' (1984) 7 Harv. JL & Pub. Pol'y 87, 87.

[74] While this paper's jurisprudential analysis has been limited to the judgements of the CJEU, national courts and supervisory authorities provide additional, detailed guidance on what is personal data.

[75] See *supra* n 67.

[76] In the policy debate regarding the suitability of anonymisation (and hence of the anonymity/identifiability dichotomy) as a basis for policy, Rubinstein and Hartzog identify two major groups: 'formalists (for whom mathematical proof is the touchstone of any meaningful policy) and pragmatists (for whom workable solutions should prevail over theoretical concerns)': Rubinstein and Hartzog (n 5) 706.

[77] E.g. *inter alios* Aldhouse (n 5); Purtova (n 9).

[78] Article 29 Data Protection Working Party, 'Opinion 4/2007 on the Concept of Personal Data WP136' (n 8) 4.

[79] I.e. the protection of fundamental right and freedoms with regard to the processing of personal data, and the free movement of such data.

[80] See YS and others v. Minister of Immigration, Integration and Asylum (n 19) s 41-42.

[81] GDPR, Art. 1.

[82] On the other hand, the Working Party also cautions against an overly restrictive interpretation, which would leave individuals deprived of protection. 'The scope of the data protection rules should not be overstretched': Article 29 Data Protection Working Party, 'Opinion 4/2007 on the Concept of Personal Data WP136' (n 8) 5.

[83] See Raphaël Gellert, 'We Have Always Managed Risks in Data Protection Law: Understanding the Similarities and Differences Between the Rights-Based and the Risk-Based Approaches to Data Protection' (2016) 2 European Data Protection Law Review 481 <https://doi.org/10.21552/EDPL/2016/4/7>; Raphaël Gellert, 'Understanding Data Protection As Risk Regulation' (2015) 18 Journal of Internet Law 3; Raphaël Gellert, 'Data Protection: A Risk Regulation? Between the Risk Management of Everything and the Precautionary Alternative' (2015) 5 International Data Privacy Law. See also Article 29 Data Protection Working Party, 'Statement 14/EN WP 218 on the Role of a Risk-Based Approach in Data Protection Legal Frameworks' (2014).

[84] GDPR, Art. 2, 3.

[85] E.g. GDPR, Art. 23; Art. 85 ss.

[86] Article 29 Data Protection Working Party, 'Opinion 4/2007 on the Concept of Personal Data WP136' (n 8) 5.

[87] E.g. the subjective, restrictive interpretation adopted by the courts in the Russian Federation: see Arkhipov and Naumov (n 68).

[88] See Commission of the European Communities, 'Commission Communication on the protection of Individuals In relation to the processing of personal data In the Community and Information security' COM(90) 314 final. Brussels, 13.09.1990, p. 19: 'As in Convention 108, a broad definition is adopted in order to cover all Information which may be linked to an individual. Depending on the use to which it is put, any Item of data relating to an individual, harmless though It may seem, may be sensitive […] to avoid a situation in which means of indirect identification make it possible to circumvent this definition'.

[89] In other words, wondering 'is this personal data?' is a disingenuous exercise, in that it does not lead to the same reasoning as 'is this the *processing* of personal data?' by removing the contextual and environmental considerations that, when data does not relate to a person by virtue of its content, make or break its qualification as personal.

[90] See the discussion on the differences between the EU notion of personal data and the US concept of personally identifiable information in Paul M Schwartz and Daniel J Solove, 'Reconciling Personal Information in the United States and European Union' (2014) 102 California Law Review 877.

[91] Methodologically, the analysis has been carried out by enucleating all the possible configurations of the interaction between the 'relating to' link and the identifiability requirement (i.e. when the data subject is identified or identifiable, directly or indirectly, through the means available to the controller or to another person, and the data relate to him or her by virtue of their content, or their purpose, or their result), systematising them in a table, and populating the table with examples of processing instances strictly suiting each possible configuration (e.g. information relating, by virtue of its purpose, to a natural person that is indirectly identifiable through the means available to a person other than the controller, or information relating, by virtue of its content, to a person that is directly identifiable by the controller).

[92] Article 29 Data Protection Working Party, 'Opinion 4/2007 on the Concept of Personal Data WP136' (n 8).

[93] Admittedly, what 'information relating to a natural person by virtue of its content' means is still largely up for debate. I take a somewhat narrow view, and consider it as data whose content relates to the data subject no matter how it is used or the result of its processing, e.g. an employee number. An IP address is information whose content is about a machine, but whose purpose or result may relate to the natural person using that machine. Conversely, my employee ID number is information whose content relates to me no matter its purpose or impact. In this sense, I believe the CJEU, in its Nowak judgement, framed the concept of 'relating to through content' in a confusing manner, conflating the content element and the purpose and result ones: see e.g. 'the content of those answers reflects the extent of the candidate's knowledge and competence in a given field and, in some cases, his intellect, thought processes, and judgment' (Nowak (n 19) s 37) and '(t)he content of those comments reflects the opinion or the assessment of the examiner of the individual performance of the candidate in the examination, particularly of his or her knowledge and competences in the field concerned' (Nowak (n 19) s 43).

[94] An anonymous reviewer suggested a helpful parallelism with Durant v Financial Services Authority [2003] EWCA Civ 1746. In Durant, the Court of Appeal of England and Wales tied the qualification of information as personal data to its existence in 'a continuum of relevance or proximity to the data subject', to be determined considering 'whether the information is biographical in a significant sense, […] going beyond the recording of the putative data subject's involvement in a matter or an event that has no personal connotations' and that the information 'should have the putative data subject as its focus' (§28). The Durant v FSA criterion of biographical significance is arguably too narrow and in contradiction with successive ECJ jurisprudence. However, the judgement still provides helpful guidance on the interaction between identifiability and the 'relating to' link when it recognizes as a 'reason for hesitation […] that in some cases it is Mr. Durant's identity that leads to the information, rather that the information leading to Mr. Durant' (§78).

[95] See Alessandro Mantelero, 'Personal Data for Decisional Purposes in the Age of Analytics: From an Individual to a Collective Dimension of Data Protection' (2016) 32 Computer Law & Security Review 238 <https://www.sciencedirect.com/science/article/pii/S0267364916300280> accessed 2 February 2018; Alessandro Mantelero, 'From Group Privacy to Collective Privacy: Towards a New Dimension of Privacy and Data Protection in the Big Data Era' in Linnet Taylor, Luciano Floridi and Bart van der Sloot (eds), *Group Privacy* (Springer 2017).

[96] See Meg Leta Ambrose, 'It's About Time: Privacy, Information Life Cycles, and the Right to Be Forgotten' (2012) 16 Stan. Tech. L. Rev. 369.

[97] Information that does not relate to a data subject due to its content but by virtue of its purpose or result can still happen to relate to the data subject for the entirety of its lifecycle, depending on the kind of data and how it is processed, but that is a mere possibility. Information whose content relates to a data subject will keep relating for a data subject for its entire lifecycle no matter how it is processed.

[98] Along with ensuring the free movement of personal data: GDPR, Art. 1.

[99] 'What matters here, however, is not simply whether the individual with prior knowledge can identify the data subject concerned but whether he/she will learn something new from the information obtained through re-identification': Article 29 Data Protection Working Party, 'Opinion 06/2013 on Open Data and Public Sector Information ('PSI') Reuse WP207' (2013) 15.

[100] See e.g. Schwartz and Solove (n 89) 877; 892.

[101] Granted, the same information can relate to more than a single natural person; it may also constitute personal data for each of them, provided that each person is also identified or identifiable (see CJEU, Nowak (n 19) s 45). See also Article 29 Data Protection Working Party, 'Opinion 4/2007 on the Concept of Personal Data WP 136' (n 7) 11: 'the same piece of information may relate to different individuals at the same time, depending on what element is present with regard to each one'.

[102] Borgesius follows the same line of reasoning with regard to online behavioural advertising: 'Some data processing activities for behavioural targeting do not concern personal data. A company can use personal data to construct a model […] Such models do not consist of personal data, as they do not relate to a specific person. As soon as a company applies the model to an individual, however, the information relates to this person because of its purpose or its result' - Frederik J Zuiderveen Borgesius, 'Singling out People without Knowing Their Names–Behavioural Targeting, Pseudonymous Data, and the New Data Protection Regulation' (2016) 32 Computer Law & Security Review 260.

[103] Article 29 Data Protection Working Party, 'Opinion 05/2014 on Anonymisation Techniques WP216' (n 57) 3. On the opinion, see Khaled El Emam and Cecilia Alvarez, 'A Critical Appraisal of the Article 29 Working Party Opinion 05 / 2014 on Data Anonymization Techniques' (2014) 5 International Data Privacy Law 1.

[104] ''(I)dentification' not only means the possibility of retrieving a person's name and/or address, but also includes potential identifiability by singling out, linkability and inference': Article 29 Data Protection Working Party, 'Opinion 05/2014 on Anonymisation Techniques WP216' (n 57) 10.

[105] See Zuiderveen Borgesius (n 101) 260.

[106] CJEU Breyer (n 31) s 46.

[107] GDPR, Recital 26; Article 29 Data Protection Working Party, 'Opinion 4/2007 on the Concept of Personal Data WP136' (n 8) 15 ss.

[108] GDPR, Recital 26.

[109] Article 29 Data Protection Working Party, 'Opinion 4/2007 on the Concept of Personal Data WP136' (n 8) 6.

[110] Article 29 Data Protection Working Party, 'Opinion 4/2007 on the Concept of Personal Data WP136' (n 8) 15.

[111] CJEU, Breyer v Germany (n 31) s 45-46. It is reasonable to assume that the Court did not mean to equate impossibility with zero probability, but with a negligible level of the latter – 'so that the risk of identification appears in reality to be insignificant'.

[112] CJEU, Breyer v Germany (n 31) s 46.

[113] GDPR, Recital 26.

[114] Article 29 Data Protection Working Party, 'Opinion 4/2007 on the Concept of Personal Data WP136' (n 8) 15.

[115] Article 29 Data Protection Working Party, 'Opinion 06/2013 on Open Data and Public Sector Information ('PSI') Reuse WP207' (n 98) 16–17.

[116] Article 29 Data Protection Working Party, 'Opinion 06/2013 on Open Data and Public Sector Information ('PSI') Reuse WP207' (n 98) 14–15.

[117] Article 29 Data Protection Working Party, 'Opinion 06/2013 on Open Data and Public Sector Information ('PSI') Reuse WP207' (n 98) 14.

[118] Article 29 Data Protection Working Party, 'Opinion 06/2013 on Open Data and Public Sector Information ('PSI') Reuse WP207' (n 98) 17.

[119] See e.g. Keith Spicer and others, 'Intruder Testing: Demonstrating Practical Evidence of Disclosure Protection in 2011 UK Census Intruder Testing: Demonstrating Practical Evidence of Disclosure Protection in 2011 UK Census' 28; Caroline Tudor, George Cornish and Keith Spicer, 'Intruder Testing on the 2011 UK Census: Providing Practical Evidence for Disclosure Protection' (2014) 5 Journal of Privacy and Confidentiality 3.

[120] Article 29 Data Protection Working Party, 'Opinion 06/2013 on Open Data and Public Sector Information ('PSI') Reuse WP207' (n 98) 16.

[121] See Elaine Mackey and Mark Elliot, 'Understanding the Data Environment' (2013) 20 XRDS: Crossroads 36; Elliot and others (n 63).

[122] Mark Elliot and Elaine Mackey, 'The Social Data Environment' in Keiron O'Hara, MH Carolyn Nguyen and Peter Haynes (eds), *Digital Enlightenment Yearbook* (2014) 256.

[123] See Miranda Mourby and others, 'Are "Pseudonymised"Data Always Personal Data? Implications of the GDPR for Administrative Data Research in the UK' [2018] Computer Law & Security Review.

[124] In this sense, see Opinion of AG Campos Sànchez-Bordona in case Breyer v Germany, C-582/14, 12 May 2016, ECLI:EU:C:2016:339, s 67-68.

[125] See the A29WP on how the disclosure of public-sector information impacts the assessment of the 'other person' who may re-identify the data subject: 'once data are publicly released for reuse, there will be no control over who can access to the data. The likelihood that 'any other person' will have the means and will use those means to re-identify the data subjects will increase very significantly. Therefore, and irrespective of the interpretation of recital 26 in other contexts, when it comes to making data available for reuse […] utmost care should be taken to ensure that the datasets to be disclosed should not include data that can be re-identified by means likely reasonably to be used by any person, including potential re-users, but also other parties that may have an interest in obtaining the data, including law enforcement' - Article 29 Data Protection Working Party, 'Opinion 06/2013 on Open Data and Public Sector Information ('PSI') Reuse WP207' (n 98) 13.

[126] GDPR, Recital 26.

[127] CJEU, Breyer v Germany (n 31) s 41-44.

[128] Opinion of AG Campos Sànchez-Bordona in case Breyer v Germany (n 122) s 65.

[129] Opinion of AG Campos Sànchez-Bordona in case Breyer v Germany (n 122) s 68.