The law of everything. Broad concept of personal data and overstretched scope of EU data protection law

Nadezhda Purtova
Associate Professor, Tilburg Institute for Law, Technology, and Society (TILT), Tilburg University, The Netherlands, P.O. Box 90153. 5000 LE Tilburg

THIS PAPER IS A WORK IN PROGRESS; ITS ARGUMENT IS BEING REFINED. IT IS PLACED ON SSRN FOR COMMENTS AND FEEDBACK, AND THE FINAL VERSION MAY SUBSTANTIALLY DIFFER

Abstract
Has the concept of personal data become too broad? Article 29 Working Party’s position facilitates a plausible argument that everything can be personal data, leading to the application of data protection law to unconventional situations, like observing weather and analysing wastewater. The reason goes beyond the rapid progress of technology towards perfect (re)identification of datasets. Datafication and advances in data analytics make everything (contain) information, and in increasingly ‘smart’ environments any information is likely to relate to an individual. The case-law of the EU Court of Justice, in particular Breyer, YS, and Nowak, is either in line with the broad approach of WP29, or has limited potential to avert the imminent explosion of the situations falling within the scope of data protection. The GDPR with its intensive compliance regime is on the verge of becoming the law of everything, well meant but impossible to comply with.

Keywords
Concept of personal data, material scope of data protection, information relating to an individual, Breyer, YS, Nowak

I Introduction
In his “Morality of Law” Lon Fuller tells a tale of a young ruler who undertakes to reform the law of the land. After a few attempts at it, all met by public discontent, the ruler wants to teach his subjects a lesson and makes it a crime “to cough, sneeze, hiccup, faint or fall down in the presence of the king … [and] not to understand, believe in, and correctly profess the doctrine of evolutionary, democratic redemption.”1 Unsurprisingly, the citizens threaten to disregard its provisions, since “[t]o command what cannot be done is not to make law; it is to unmake law, for a command that cannot be obeyed serves no end but confusion, fear and chaos.”2 This could have as well been said about the EU data protection law as it currently develops.

1 LL Fuller, The Morality of Law (1969) Yale University Press, 36-37
2 Fuller, 37
3 Directive 95/46/EC on the protection of individuals with regard to the processing of personal data and
This paper will argue that the material scope of the current 1995 Data protection directive\(^3\) (‘DPD’) and the new General Data Protection Regulation\(^4\) (‘GDPR’) is growing so broad that the good intentions to provide the most complete protection possible are likely to backfire soon and result in the ridicule of the data protection as it was intended. The concept ‘personal data’ determining the material scope of data protection is bound to expand due to the in-built possibilities for broad interpretation, exploding generation and aggregation of data, as well as advances in data analytics. I will demonstrate that literally everything can nowadays be plausibly argued to be personal data: weather, waste water, exam scripts. As a result, the European data protection law is now on the verge of becoming “the law of everything”, meant to deliver the highest legal protection under all circumstances, but in practice impossible to comply with and hence ignored or discredited as conducive to abuse of rights and unreasonable.

This discussion on the expansion of scope of personal data has been brewing for quite some time in the data protection community. Many privacy and data protection scholars have been critical of the concept of personal data as growing too broad. The mainstream literature is centred around one element of the concept of personal data, i.e. identifiability of a person, and one strand of technological development, namely, re-identification and de-anonymisation algorithms, and one corresponding element of the notion ‘personal data’ – identifiability to a particular person. To name just a few key authors arguing to that effect, the works of Ohm\(^5\), Sweeney\(^6\), Schwartz and Solove\(^7\) suggest that, given the progress of data processing technologies and the amount of data available for analysis, absolute and irreversible anonymity is no longer possible. Tene and Polonetsky note how Big Data analytics makes a binary distinction between identifiable and non-identifiable information meaningless.\(^8\)

Schwartz and Solove propose to keep personal data (or personally identifiable information, a functional equivalent of personal data in the US) as a threshold of protection, but define it sharper, namely, based on the risk of identification from ‘0’ (zero risk of identification) to ‘identified’, and to treat information with varying degrees of identifiability differently.\(^9\) The Court of Justice has recently adopted a broad approach to identifiability in the Breyer case.\(^10\)

What these authors seem to overlook is that the problem with the concept of personal data goes beyond simple identifiability and the second element of the concept ‘personal data’, i.e. the relation of information to a person, is problematic as well. In the age of autonomous machines, when we cannot possibly know how the autonomous self-learning and self-managing computers draw meaning from data, any information can relate to a person.

At the time of writing, the Court of Justice was asked to form an opinion on the meaning of ‘information relating to a natural person’ in the Nowak case.\(^11\) The decision will either support

---

3 Directive 95/46/EC on the protection of individuals with regard to the processing of personal data and on the free movement of such data, Official Journal L 281, 23/11/1995 p. 0031 – 0050
10 Court of Justice of the European Union, Case C-582/14, Patrick Breyer v. Bundesrepublik Deutschland, judgment of 19 October 2016, ECLI:EU:C:2016:779 (hereinafter Breyer)
11 Court of Justice of the European Union, Case C-434/16, Reference for a preliminary ruling from Supreme Court (Ireland) made on 4 August 2016, Peter Nowak v Data Protection Commissioner OJ C
the trend towards interpreting personal data as an all-encompassing concept, or reverse it. Whichever the outcome of the Nowak case, it will be of significant impact for the future of data protection law. In other words, there is no better time for the analysis in this paper as present.

The argument will proceed as follows. I will start by breaking up the definitions of personal data under the DPD and the GDPR to demonstrate how those are inherently flexible and capable to stretch to accommodate for changing context (Part II). I will then review how the concept of personal data has been applied so far, mainly referring to the Article 29 Working Party (‘WP29’) opinion on the concept of personal data as, albeit non-binding, the primary EU guideline for data protection compliance in practice. I will show how the notion of personal data according to WP29, broadly defined and in light of the rapidly advancing data analysis techniques, potentially renders nearly everything personal data and hence subject to the data protection regime (Part III). I will subsequently review the case-law of the Court of Justice and examine if the Court follows the footsteps of WP29 (Part IV). I will then discuss implications of the analysis for the future of the data protection law in Europe (Part V), followed by a conclusion (Part VI).

II  Legal definition of personal data: flexibility, adaptability, and uncertainty

‘Personal data’ is one of the key notions of data protection law determining the material scope of both the current DPD and the new GDPR which will be effective from 25 May 2018. When personal data is ‘processed’, the data protection principles, rights and obligations apply (Article 3(1) DPD and Article 2(1) GDPR). When there is no personal data, they do not.

DPD defines personal data as

\[
\text{any information relating to an identified or identifiable natural person ('data subject'); an identifiable person is one who can be identified, directly or indirectly, in particular by reference to an identification number or to one or more factors specific to his physical, physiological, mental, economic, cultural or social identity.}
\]

Article 4(1) GDPR closely follows DPD.\(^{13}\)

\[
'\text{personal data’ means any information relating to an identified or identifiable natural person ('data subject'); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person.}
\]

Anonymous data is the opposite to personal data and refers to the information that does not relate to an identified or identifiable person, or to personal data “rendered anonymous in such a manner that the data subject is not or no longer identifiable”.\(^{14}\) Processing anonymous data does not trigger application of data protection law. Pseudonymous data, i.e. personal data after pseudonymisation, is still information pertaining to an identifiable person and subject to the data protection law.\(^{15}\)

---

\(^{364}\), 3.10.2016, p. 11–11 (hereinafter Nowak)

\(^{12}\) Article 29 Working Party opinion 4/2007 on the concept of personal data, 20 June 2007, p. 3 (WP 136)

\(^{13}\) This extended part of the definition in GDPR also clarifies some issues that raised questions under the DPD, e.g. whether or not genetic data is personal data. However, this discussion is outside of the scope of this paper and will not be pursued here any further.

\(^{14}\) Recital 26 DPD and GDPR

\(^{15}\) Recital 26 GDPR; ‘presudonymisation’ means “the processing of personal data in such a manner that
Schwartz and Solove characterize the EU definition of personal data as broad, flexible, and adaptable to technological context. Two elements of the definition in particular ensure such flexibility and allow for the definition of personal data to ‘stretch’. The first is identifiability. Indeed, in order to be personal data, information does not have to relate to a person already known, or ‘identified’, but could also be ‘identifiable’, i.e. a mere possibility of identification is of relevance. This reference to the possibility of identification naturally creates a window for a judgement on the extent of this possibility, namely, what chance of identification would be sufficient to meet the legal standard, and whether or not a chance of identifiability is sufficient to render data ‘personal’. To provide some legal certainty as to how this element should be applied, some guidance is needed, and given in a non-binding Recital 26 of both DPD and GDPR. Although the two recitals are in line with each other, from here on, my analysis will focus on the GDPR only.

Recital 29 GDPR adopts a test of reasonable likelihood of identification not only by the current controller, but “by the controller or by another person”, taking into account not the subjective ability to identify, but the state of art of technology at the time of processing:

To ascertain whether a natural person is identifiable, account should be taken of all the means reasonably likely to be used […] by the controller or by another person to identify the natural person directly or indirectly. To ascertain whether means are reasonably likely to be used, account should be taken of all objective factors, such as the costs of and the amount of time required for identification, taking into consideration the available technology at the time of the processing and technological developments.

As Schwartz and Solove point out, the explanation in Recital 26 GDPR makes the GDPR concept of ‘personal data’ suitable for “a tailored, context-specific analysis for deciding whether or not personal data is present.” According to the Recital, it is possible for the same piece of data not to be personal at the time of collection, but turn into personal later, just sitting there, simply by virtue of progress of technological tools.

In addition to identifiability, there is another element of the definition that invites context-dependent assessment. Like ‘identifiability’, ‘information relating to’ a natural person may be interpreted broadly and narrowly, and invites a judgement on what degree of relationship to a person is sufficient, and if this relationship is present in particular circumstances. However, the preamble to the GDPR (or of the DPD) gives no guideline as to the meaning of ‘relating to’.

Due to the flexibility of the definition in the DPD, and despite the DPD being an instrument of complete harmonization, there have been a significant divergence in how the national legislation so far implemented the definition of personal data; the complete harmonization only precludes Member States from adding additional elements to the harmonised provisions,
but still allows “deciding the details or choosing between options”.\(^{22}\) Also, the Recitals briefing and clarifying the definition of personal data are non-binding, and the Member States are under no obligation to include those in their national implementation. It remains to be seen if the choice of a regulation as a legislative instrument will result in a uniform application of the GDPR definition of personal data across the EU member states.

### III Weather as personal data? Article 29 Working Party opinion

1 **Introducing WP 136: streamlining national implementation and relevance after May 2018**

To streamline the national implementation of the definition of personal data, Article 29 Working Party, a EU advisory authority on the matters of data protection, adopted a non-binding opinion on the concept of personal data.\(^{23}\) Even WP29 opinion concerns the concept of personal data in the DPD, the opinion will most likely remain significant for data protection compliance even after the GDPR becomes effective, since as Advocate General Kokott observes, “the latter will not affect the concept of personal data.”\(^{24}\)

The concept ‘personal data’ – as interpreted in WP136 – lives up to its potential to become an all-encompassing notion and drive the data protection law over the cliff. WP29 begins with policy considerations, the most relevant of which for this analysis are that the notion of personal data is broad, and intends to cover all information which may be linked to an individual; the DPD is sufficiently flexible to strike a balance between the data subjects’ rights and legitimate interests of others; while the DPD aims to protect individuals, the scope of data protection rules should not be overstretched at the risk of “of ending up applying data protection rules to situations which were not intended to be covered by those rules and for which they were not designed by the legislator”; at the same time, unduly restrictive interpretation of personal data should be avoided “so that it can anticipate evolutions and catch all ‘shadow zones’ within its scope.”\(^{25}\)

WP29 breaks up the definition of personal data into four elements. Personal data is: (a) information, (b) relating to (c) identified or identifiable (d) natural person. I will only consider the first three as relevant for the argument. Pointing to the flaws of the ‘identifiability’ element is less controversial compared to the other two, so I will start with it.

2 **“Identified or identifiable [natural person]”**

WP29 adopts a broad understanding of what “identified or identifiable” means, making it very difficult to sustain a meaningful distinction between personal and not personal data. “Identified” refers to a person who is known, or distinguished in a group, and “identifiable” is a person who is not identified yet, but identification is possible.\(^{26}\) This reference to the possibility of identification opens the window for a broad interpretation and raises most discussions. Further, one is directly identified or identifiable most commonly by reference to a name, in combination with additional information if the name is not unique;\(^{27}\) one is

---


\(^{23}\) WP 136

\(^{24}\) Opinion of Advocate General Kokott, Case C-434/16 *Peter Nowak v Data Protection Commissioner*, delivered on 20 July 2017, para. 3.

\(^{25}\) WP 136, 4-5

\(^{26}\) WP 136, 12

\(^{27}\) WP 136, 13
“indirectly identifiable” by the so-called “unique combinations” of not unique identifiers that allow the individual to be singled out, not necessarily by name.

What receives most attention, however, is whether or not the means of identification are “reasonably likely to be used” under Recital 26 DPD. WP29 follows the language of the Recital closely, restating that the means of identification are “reasonably likely to be used by the controller or any other person”, often interpreted as anybody, which is a significantly broader interpretation allowing for more data to be considered personal, as opposed to the narrow “by the controller”. The former approach is often called “absolute”, or “objective,” and the latter – “relative”.

While one may argue that at least somebody is bound to have sufficient capacity to identify, WP29 clarifies that a “purely hypothetical possibility” of identification is insufficient to meet the standard of “reasonably likely”. Instead, “all the factors at stake” should be considered to assess this possibility. Examples of such factors are:

- the cost of identification;
- the intended explicit or implied purpose of processing (when “the processing … only makes sense if it allows identification of specific individuals and treatment of them in a certain way”, the availability of tools of identification should be presumed reasonably likely;
- the risk of organisational dysfunctions (e.g. breaches of confidentiality duties) and technical failures, likely including data breaches;
- the state of the art in technology at the time of the processing, including possible technological developments in future, within the lifetime of processing;
- Measures to prevent data identification (i.e. to maintain anonymity) are of importance as a means of avoiding processing personal data altogether, rather than in fulfilment of data security obligations under DPD.

The resulting standard of the reasonable likelihood of identification is quite broad and context-dependent, leading to one major consequence: the status of data as ‘personal’ is dynamic, i.e. the same dataset may not obviously be personally identifiable at the start of processing, or from the perspective of the controller, given the tools and data available to him, but become, or appear to have been all along, identifiable for another person or once the circumstances change. Rephrasing Koops, just as light sometimes acts as a particle and sometimes as a wave, data sometimes act as identifiable and at other times as non-identifiable.

It has become a widely accepted point in the privacy and data protection literature that as the data processing technologies advance, and the pool of data which can be combined grows, the state of the art in technological means of identification will change. At one time, it was thought that data which were initially not identifiable would remain so. However, with the development of new technologies, such as the ability to combine data from different sources, it has become possible to identify individuals who were previously unidentifiable.

---

30 Art 1(1) 1998 Data Protection Act, where the individual needs to be identifiable “(a) from those data, or (b) from those data and other information which is in the possession of, or is likely to come into the possession of, the data controller”.
31 E.g. FJ Zuiderveen Borgesius ‘Singling out people without knowing their names – Behavioural targeting, pseudonymous data, and the new Data Protection Regulation’ Computer Law and Security Review 32(2), April 2016, 256, 264
32 WP 136, 15
33 WP 136, 15-16
34 WP 136, 16
35 WP 136, 17
and as combining databases becomes daily practice of intelligence agencies, ‘smart city’ municipalities, and advertising, so does the reasonable likelihood of somebody being able to link any piece of information to an individual. To name just a few key authors arguing to that effect, I refer to the works of Ohm,27 Sweeney,28 Schwartz and Solove,29 and Tene and Polonetsky.30 Data processing practice is full of examples where data first considered anonymous was identified. In 2000 the combination of a ZIP code, date of birth, and gender was enough to identify 87% of the US population.41 Famously, film rating records of 500,000 Netflix subscribers were re-identified in 2008 using the openly accessible Internet Movie Database.42 In 2013 travel routes of celebrities such as Bradley Cooper and Olivia Munn, including street addresses, and whether or not they left a tip, where deduced from the “anonymised” public database of the New York taxi rides which contained no passenger information, and paparazzi pictures.43 In 2014 knowing location of credit card holders on 4 occasions allowed to re-identify 90% of 3 months of credit card transactions, chronicling the spending of 1.1 million people in 10,000 shops, having access only to amounts spent, shop type and a code representing each person. Knowing the amounts spent on these 4 occasions lead to re-identification of nearly all card-holders.44

The examples demonstrate that the capacity of (re-)identification is increasing each year, and although perfect identification may still be a myth, one wonders for how long. Many legal (e.g. the often cited Ohm) and technical scholars (e.g. Narayanan) agree that at this rate, a meaningful distinction between identifiable and non-identifiable information is not sustainable much longer.47 In Europe this conclusion is further reinforced by a complementary WP29 opinion on anonymisation: the data is not identifiable, i.e. anonymous, only when anonymisation is irreversible.48

32 Narayanan, A. & Shmatikov, V. (2008) ‘Robust de-anonymisation of large datasets. (How to break anonymity of Netflix Prize dataset)’
36 A Narayanan and EW Felten ‘No silver bullet: De-identification still doesn't work’ randomwalker.info, published 9 July 2014, arguing that, while data science is encouraged, maintaining the distinction is “incompatible with the reality of re-identification science”.
37 But see A Cavoukian & D Castro, ‘Big Data and Innovation, Setting the Record Straight: De-identification Does Work’ (2014), available at http://www2.itif.org/2014-big-data-deidentification.pdf , arguing that proper de-identification works. However, they also agree that de-identification does not work for the “high-dimensional data.”
38 Article 29 Working Party opinion 05/2014 on Anonymisation Techniques, 10 April 2014, p. 3 (WP 216), 3, 5-7
3 “Any information”

Unlike the “identifiability”, the remaining two elements of the definition receive little attention in the literature. Yet, as WP29 interprets them, they significantly contribute to the imminent explosion of the kinds of data that can be considered personal, and are therefore not less controversial. While explaining the element “any information”, WP29 does not examine what information means, likely considering it self-evident, and focuses immediately on what kinds of information would fall under “any information”. In the age of ubiquitous computing and advanced analytics algorithms, this omission may have a truly explosive effect on the range of situations that would fall within the scope of data protection law.

a Broad but undefined concept of information

WP29 starts with a broad declaration of the intent of the legislator “to design a broad concept of personal data” and further explains that any information can fall under the concept “personal data” regardless of its nature, content, or format. To be considered personal data, the nature of information is of no significance: it can be true or inaccurate, objective and subjective, including opinions and assessments.

Next, there are no particular requirements to the content of information. Information does not have to concern private or family life, and could pertain to life of the individual in his or her professional and other capacities, which is consistent with the aim of DPD to protect “the fundamental rights and freedoms of natural persons, and in particular [but not exclusively] their right to privacy”.

Finally, information can constitute personal data regardless of the format, medium, or form, which could be “alphabetical, numerical, graphical, photographical or acoustic”, “kept on paper [or] stored in a computer memory” as a binary code, structured or unstructured, provided the other criteria of the definition are met. Video and voice recording can be such information, as well as a child’s drawing that could contain personal data of both the child and the parents.

Aside from this, the explanation by WP29 refers to “information” as a concept the meaning of which is self-evident. In any case, no further clarification as to what is meant by ‘information’, and how it relates to or differs from other information-related concepts, such as ‘data’, ‘meaning’, ‘knowledge’, or information artefacts (‘information carriers’, e.g. a book, a CD, etc.) is given. An unconvincing attempt is made in a short paragraph regarding human tissue samples. According to WP29, these are “sources” of biometric data but “not biometric data themselves.”

Therefore the extraction of information from the samples is collection of personal data, to which the rules of the Directive apply. The collection, storage and use of tissue samples themselves may be subject to separate sets of rules.

---

49 WP 136, 6
50 WP 136, 6
51 WP 136, 6-7
52 WP 136, 7
53 WP 136, 7
54 WP 136, 8
55 WP 136, 8
57 WP 136, 9
58 WP 136, 9
WP29 seems to opine that the samples are information carriers, like e.g. CDs, rather than information. At the same time, the exact wording of the paragraph is not this explicit. Specifically, a number of questions arise from this brief paragraph: first, is there a meaningful difference between tissue samples and the child’s drawing that the WP29 used earlier as an example of information and not a source of information, when both do not provide information immediately but require “information extraction”; second, if the tissue samples “are not biometric data themselves”, are they biometric data in combination with additional information or in a certain technological context; and third, if the collection, storage and use of such samples “may be subject to separate sets of rules,” are those rules in addition to data protection, lex specialis in relation to the general rules of DPD, or does the data protection law not apply?

b  “Everything is information”

The striking consequence of the declared commitment of WP29 to the all-encompassing interpretation of “any information” and the vagueness of WP29 on the meaning of information is that a viable argument can be built that everything is information.

Information is a notoriously nebulous concept that has different meanings. These meanings vary over time, across and within disciplines, or depending on one’s philosophical inclinations. While law is generally characterised by poor conceptualisation of information, several analyses have adopted a General Definition of Information (‘GDI’) as an operational standard: “information is data + meaning.” Data is the first element of the definition and stands for the lack of uniformity in the world (what Floridi poetically calls “the fractures in the fabric of being”), or between at least two physical states (a higher or lower charge in a battery), or between two symbols (letters A and B, or numerals 0 and 1). Put differently, data is “a description of something that allows it to be recorded, analysed, and reorganized.” People measured and quantified the world long before the Digital Age, e.g. in the form of maths and language. However, the arrival and proliferation of computing turbocharged this process, enabling to measure and quantify, and hence harness data from literally everything. Meaning is the second element of the GDI. To perceive data as information we need to make sense of it. For instance, we understand combinations of letters


as words and sentences, or observe that the battery charge is low and conclude that the battery will soon need to be recharged. However, human cognition is limited in the extent to which it can make sense of data. Some data colloquially referred to as ‘raw’ has no or very limited meaning to a human brain, like zeros and ones of a binary code. The way all computers attach meaning to data is different from human cognition, e.g. computers have a different language and process data faster. Humans – albeit with proper training and effort - can still follow how traditional computers make sense of data: these computers follow deductive models, rules and cases. However, how meaning is ‘attached’ to data by modern machines is beyond the grasp of human mind. The game-changer is a new generation of data-processing algorithms based on machine learning. Machine learning is the ability of computer algorithms to learn from data and make predictions for new situations, and improve automatically through experience. The new algorithms are autonomous, i.e. self-learning, self-repairing, and self-managing and form the core of the modern approach to Artificial Intelligence (‘AI’), a strand of computer science aimed to build computers as intelligent agents. The way advanced AI self-learning algorithms make sense of data is not transparent even for their designers. Hence, the new AI algorithms work as a black box that is truly beyond human cognition. These AI self-learning autonomous machines together with unprecedented amount of data already stored in databases or live-streamed, form the essence of Big Data and have the ability to harness information in fundamentally novel ways. In effect, we can no longer say that some data has no meaning for we really have lost to computers the monopoly of deciding that. In fact, it is safer to assume that all data – the number and frequency of steps or key strokes one makes daily, the colour of one’s eyes or even how many leaves grow on a tree - potentially has meaning, even if not for humans. Hence, everything is data and all data has meaning; hence, everything is information.

4 “Relating to”

What “relating to” means, according to WP29, is “crucial as it is very important to precisely find out which are the relations/links [between the information and the individual] that matter and how to distinguish them.” Again, the WP29 construes “relating to” very broadly.

WP29 points out that in some situations this relationship is quite obvious, while in others the link is not self-evident. In particular, the latter is the case where information relates to an object, e.g. a value of a house, or a process or event, e.g. data on the functioning of a machine that requires human intervention. In these cases there would be an indirect relationship to

73 WP 136, 9
people owning or otherwise interacting with the object. Presumably, WP29 considers the indirect relationship sufficient. In all cases, all circumstances of the case need to be taken into account in assessment.\textsuperscript{74}

Information can “relate” to an individual in \textit{content, purpose, or result}, meaning that information “relating to” a natural person includes but is broader than the information “about” that person. The meaning of “relating to” grows even broader considering that these three conditions are meant as alternative and not as cumulative ones.\textsuperscript{75}

Information relates to a person more obviously in content, i.e. when it is \textit{about} that person. However, even the information that is not in any way \textit{about} someone may be found to “relate to” a person. Information relates to a person in purpose “when the data \textit{are used or are likely to be used … with the purpose} to evaluate, treat in a certain way or influence the status or behaviour of an individual”\textsuperscript{76} [emphasis added]. Finally, information regardless of its content or any purpose of processing may relate to a person in result when “their use is \textit{likely to have an impact} on a certain person's rights and interests” even in a minor way.\textsuperscript{77} “It is sufficient if the individual may be treated differently from other persons as a result of the processing of such data.”\textsuperscript{78}

Remarkably, similar to the criterion of identifiability, the relationship in terms of purpose and result will occur not only when data are \textit{already used}, but also where it is \textit{likely to be used} with the purpose or effect of impacting people “taking into account all the circumstances surrounding the precise case”.\textsuperscript{79} I have already discussed how the reference to likelihood results in the standard of the reasonable likelihood of identification that is both broad and context-dependent, making the status of data as ‘personal’ dynamic (see III.2). The same argument equally applies here, with a significant correction: while the likelihood of identification must be “reasonable”, which WP29 interprets as more that a purely hypothetical possibility, there is no such clarification here, and hence a lower threshold of what is likely to relate to a person applies.

In other words, some information is perceived as relevant more easily, for instance, information ‘generated’ \textit{by} (observing) people (e.g. administrative records of people’s off-line lives, and digital records of online behaviour like websites visited, texts and images uploaded; information generated through use of ‘smart’ objects and devices like phones or fitness bracelets), or objects people interact with (their cars, homes, computers). At the same time, some information is hard to intuitively place in any connection of relevance for anyone: e.g. the amount of weight a block of concrete can withstand, or the amount of sand crystals in a cubic meter of sand in the Sahara desert. However, when increasing amounts of data are gathered in real time from increasingly smart environments, intended to be used in automated decision-making about us, and we cannot possibly know how the autonomous self-learning and self-managing computers draw meaning from data, we should always reasonably assume that – in the computer’s ‘mind’ – any information is likely to relate to a person, since we cannot eliminate this possibility with certainty.

5 Playing devil’s advocate: weather is personal data

In this section I will apply the elements of definition of personal data as explained by WP29 to demonstrate on a provocative example how literally everything can nowadays be plausibly argued to be personal data.

\textsuperscript{74} WP 136, 9-10
\textsuperscript{75} WP 136, 11
\textsuperscript{76} WP 136, 10
\textsuperscript{77} WP 136, 11
\textsuperscript{78} WP 136, 11
\textsuperscript{79} WP 136, 11
Weather, while neutral and of no consequence as a subject of small talk, may be considered information relating to an identified or identifiable natural person in the context of a living lab, for instance the Stratumseind 2.0 smart city project in Eindhoven, the Netherlands. In short, Stratumseind is the longest going-out street in the Netherlands, with many bars and ‘coffeeshops’ selling cannabis. The problem the street faces is the declining numbers of visitors and, as a result, decrease in turnover of the local businesses. This is said to be the result of criminal behaviour and vandalism on the street. In response, the municipality of Eindhoven, together with the police and a number of private parties (the Stratumseind establishments association, local real-property owners, etc.), started a smart city project Stratumseind 2.0. The project includes a number of smaller data-driven projects, some of which aim at predicting, preventing and de-escalating deviant behaviour on the street, among others, by engaging the police or adapting the street lighting. Private companies such as Atos, Intel and Philips, who see Stratumseind as a test-bed for their smart-city products, supply the technology enabling the projects.

The data is gathered from the multiple sensors installed on the street, including video and acoustic cameras, sound sensors, WIFI tracking and a weather station. These sensors, among others, measure how many people pass by the street per day and how they move, and where they come from, sound on the street, what it is and where it comes from, rainfall per hour, temperature, wind direction and speed. All the data is stored in a database which enables knowledge discovery, although the actual currently performed analysis is limited.

Would rainfall per hour, temperature, wind direction and speed, together referred to as “weather” be personal data following the WP29 guidelines? I would argue that they are. The weather is observed and recorded by the weather station, i.e. “datified”, and hence it is information fitting the broad approach of WP29. Although not about people, in the context of Stratumseind 2.0, this information is collected in a database that will likely be used for a purpose to assess and influence their (deviant) behaviour, and hence it is information relating to people in purpose. In the context of a large knowledge discovery database built for advanced data analytics, by several different public and private parties with varying interests, it is possible to imagine that not all results of data analytics pertaining to human behaviour on the street would be intended, both in terms of the opaqueness and unpredictability of the advanced analytics algorithms and possible lack of absolute agreement between the involved parties on their intentions. In that part, the weather information will still be relating to people in result. Finally, each visitor of Stratumseind is highly likely to be identifiable if not by the weather information alone, certainly in combination with the data from the WIFI tracking sensors, voice recordings or video footage; if not by a weather station operator, certainly by some other project partners who, being technology companies, possess the tools and expertise to do so. Alternatively, since the purpose of the project is not only to predict but also address criminal and other deviant behaviour, this implies that the perpetrators’ identification is intended. Hence, the identification is “reasonably likely” as III.2 explains. Therefore, under these circumstances, weather is information that relates to a number of personally identified or identifiable natural persons, and is personal data. A similar argument may be constructed in relation to waste water should the project partners decide to include the data from monitoring the sewer (e.g. for the purpose of detecting drug labs) in the database.

An objection may be raised that the weather and waste water could only be brought in a relationship to an identifiable person and hence contain personal data in very particular circumstances of what Cavoukian and Castro call “high-dimensional data” which “consists of numerous data points about each individual, enough that every individual’s record is likely to be unique, and not even similar to other records.” However, as Narayanan and Felten

---

80 In the factual description of the Stratumseind 2.0 case I refer to the ongoing PhD research of Maša Galic on privacy and big data-based security in public spaces.

81 A Narayanan and EW Felten ‘No silver bullet: De-identification still doesn’t work’ randomwalker.info, published 9 July 2014,
explain, “high-dimensional data is now the norm, not the exception. … [T]hese days it is rare for useful, interesting datasets to be low-dimensional.”

This weather example may come across as provocative, but points to a real concern about applying the same data protection rules to any information-related phenomenon that we intuitively feel should be protected or trigger protection. This trend will eventually lead to results that were hardly intended at the origins of the data protection law. Think of the weather station as data controller to whom data protection obligations apply, including data security obligations to protect weather information from unauthorized access or destruction, obligation to inform data subjects about third parties to whom the weather information is transferred, and data subject’s rights to object to processing of weather information. More and more situations will fall within the concept “personal data” as interpreted by WP29. This will eventually turn the data protection law in an uneconomical exercise of regulating everything, and deprive its protection of meaning. As a result, the current personal data-centred paradigm of legal protection will not be sustainable much longer.

IV The Court of Justice case law

1 Significance of WP29 opinions for the Court of Justice jurisprudence

This Part will examine if the threat of data protection law turning into the law of everything is materialising in the case-law of the EU Court of Justice, and if the Court’s interpretation of “personal data” is in line with WP29.

As explained earlier, while opinions of WP29 are influential in data protection practice, and are key compliance guidelines, they are not a source of the data protection law and not binding. Under Article 29(1) DPD, the role of the Working Party is “advisory.” Therefore, when ruling on the meaning of personal data, the Court is free to set aside WP29 opinions, including WP 136. As a matter of fact, the Court never cites WP29 opinions in its data protection jurisprudence, and on a number of occasions has ruled contrary to the WP29’s earlier opinions. A recent example is the Google Spain case where the Court found search engine providers to be controllers with regard to indexing and making available via its search personal data published on third-party websites. Hereby, without expressly mentioning it, the Court overruled WP29 earlier position that the search engine providers cannot be “the principal controller” when they act “purely as an intermediary” with regard to the third-party content containing personal data, and that the search engine providers can only be controllers “with regard to the removal of personal data from their index and search results, [while] the extent to which an obligation to remove or block personal data exists, may depend on the general tort law and liability regulations of the particular Member State”.

At the same time, Advocates General (‘AG’) in their opinions for the Court do, although not always, refer to WP29, and hence WP 136 may still be of indirect influence on how the

82 A Narayanan and EW Felten ‘No silver bullet: De-identification still doesn’t work’ randomwalker.info, published 9 July 2014.
84 Case 131/12 Google Spain SL, Google Inc v Agencia Española de Protección de Datos and Mario Costeja González [2014] ECR I-000 (nyr), para. 31 et seq.
85 Article 29 Working Party Opinion 1/2008 on data protection issues related to search engines, adopted on 4 April 2008 (WP 148), 14
86 AG Sharpston in YS, M and S does not rely on WP29 in her argument on whether or not legal analysis is personal data, and even expressly states that WP29 opinions are not binding (footnote 40)
87 Most recently Opinion of Advocate General Sanchez-Bordona, Case C-582/14 Patrick Breyer v Bundesrepublik Deutschland, delivered on 12 May 2016; and Opinion of Advocate General Kokott, Case C-434/16 Peter Nowak v Data Protection Commissioner, delivered on 20 July 2017
concept “personal data” develops in the case-law, if not in terms of substantive outcomes, surely in terms of providing the AGs and the Court with a list of issues to consider.

2 Development of CJ case law on the concept of personal data

This section will very briefly sketch the development of the EU case-law on the meaning of personal data from Lindquist until roughly 2014, focusing on the big lines rather than exhaustive description. The year 2014 marks a different stage in the Court of Justice case-law, since this was the year when in *YS and others* the Court for the first time produced extensive analysis of the meaning of ‘information relating to an individual’.88

The Court of Justice ruled on the meaning of personal data in the Directive in its very first data protection case, Lindquist, and on a number of occasions since then. However, the Court’s judgements are not nearly as comprehensive as WP 136. Nearly all of them are given in the context of a reference for preliminary ruling, where the Court is constrained by precise questions from the national courts and the circumstances of each case.

The bulk of the cases simply name a particular type of data involved in that case, ruling that this information indeed constitutes personal data without much analysis of what the various parts of the concept ‘personal data’ mean. AG Sharpston89 names the examples of such types of data: “the name of a person in conjunction with his telephone coordinates or information about his working conditions or hobbies”,90 his address,91 his daily work periods, rest periods and corresponding breaks and intervals,92 monies paid by certain bodies and the recipients,93 amounts of earned or unearned incomes and assets of natural persons.94

There is often a reference to the scope of the Directive which is very wide and the personal data covered by the Directive are varied.95

The Court has also dealt with the argument that a broad interpretation of the DPD scope would lead to the outcomes which are unreasonable and disproportionate. For instance, in Lindquist, Mrs Lindqvist argued that it was “unreasonable to take the view that the mere mention by name of a person or of personal data in a document … on an internet page constitutes automatic processing of data.”96 In the more recent *Google Spain*, AG Jääskinen submitted that

> the broad definitions of personal data, processing of personal data and controller are likely to cover an unprecedentedly wide range of new factual situations due to technological development. ... This obliges the Court to apply a rule of reason, in other words, the principle of proportionality, in interpreting the scope of the Directive in order to avoid unreasonable and excessive legal consequences.97

However, in both cases, the Court did not accept that the proportionality and ‘reasonableness’ had to be taken into account on the stage of determining the Directive’s scope. Rather, according to Lindquist, the Directive itself has a degree of flexibility98 to enable proportionate application of its rules. In effect, the principle of proportionality is respected on the stage of national implementation of the Directive and is secondary to the issue of scope.99 Similarly, in

---

88 See IV.4.a for analysis of YS and others
89 Opinion of AG Sharpston in YS, M and S, para. 44
90 AG Sharpston cites Case C-101/01 Lindqvist [2003] ECR I-12971, paragraph 24;
92 Case C-342/12 Worten [2013] ECR, paragraphs 19 and 22.
93 C-139/01 Österreichischer Rundfunk and Others [2003] ECR I-4989, paragraph 64
94 Satakunnan Markkinapörssi and Satamedia [2008] ECR I-9831, paragraphs 35 and 37
95 like in Österreichischer Rundfunk and Others, paragraph 43; Lindqvist, paragraph 88; and Case C-553/07 Rijkeboer [2009] ECR I-3889, paragraph 59.
96 Lindquist, para. 20
97 Opinion of AG Jääskinen in Google Spain, para. 30
98 Lindquist, para. 83
99 Lindquist, paras. 87-88
Google Spain the Court did not accept the proportionality argument and explained that the aim of the Directive was “to ensure a high level of protection of the fundamental rights and freedoms … with respect to the processing of personal data”;\(^{100}\) and that in so far as the Directive’s provisions are “liable to infringe fundamental freedoms”, they must be interpreted in the light of fundamental rights.\(^{101}\) Hence, the position of the Court so far has been to mitigate the possibly undesirable impact of the wide scope of the data protection law not concerning the scope, but applications of the particular provisions.

Only relatively recently, as the data processing situations which the national courts are dealing with got more complex, did the courts begin to refer questions that required a more detailed analysis of what the particular elements of the concept “personal data” meant.

To date, the Court has not yet dealt with the issue of meaning of information in the definition of personal data, and hence, WP 136 remains the only compliance guideline on this matter. The following sections will focus on, respectively, Breyer and YS and others and Nowak. The analysis of these cases will show that the case law of the Court of Justice is either in line with the broad approach to the notion of personal data that WP29 adopts, or has limited potential to avert the imminent explosion of the situations falling within the scope of data protection law.

3 ‘Identified or identifiable’: Breyer

The Court dealt with the meaning of “identifiable” in the reference for preliminary ruling in the Breyer case, the reference for preliminary ruling brought in 2014 and the ruling given in 2016. The judgement was generally received as a confirmation of the absolute approach to identifiability in the EU data protection law, which was declared in Recital 26 DPD, extended by WP29, and now received the assent of the Court.\(^{102}\) I will argue, however, that while the Court indeed went for a broad interpretation of identifiability, its reading is more restrained and measured compared to WP136, and narrows down the scope of the concept “personal data” somewhat.

The main dispute was between Mr Breyer, a German activist, and the Federal Republic of Germany. The former visited some publicly accessible websites of the German Federal institutions, and the latter, for the purposes of preventing denial of service attacks, stored IP addresses of the websites’ visitors, including Mr Breyer’s dynamic IP addresses. Mr Breyer objected to the retention of his dynamic IP addresses as his personal data.

On an earlier occasion, in Scarlet Extended, the Court has already found static IP addresses to be personal data for the internet service provider “because they allow users to be precisely identified” by the internet service provider.\(^{103}\) However, Breyer presented a different issue. Namely, unlike a static IP address “which [is] invariable and allow[s] continuous identification”,\(^ {104}\) a dynamic IP address changes each time there is a new Internet connection, and does not allow to establish a link, “through files accessible to the public”, between a given computer and the Internet connection.\(^ {105}\) In fact, the Federal Republic of Germany, the website provider who kept the IP addresses, could not identify Mr Breyer to his dynamic IP address\(^ {106}\) and would need additional information in possession of the Internet service provider to be able to do so. It was “common ground” that a dynamic IP address does not constitute information relating to an identified natural person, since the identity of the owner

---

\(^{100}\) Google Spain, para. 66

\(^{101}\) Google Spain, para. 68

\(^{102}\) e.g. FJ Zuiderveen Borgesius ‘Breyer case of the Court of Justice of the European Union: IP addresses and the personal data definition’ European Data Protection Law Review 2017, Volume 3, Issue 1

\(^{103}\) Breyer, paras 33-34, citing Scarlet Extended (C-70/10, EU:C:2011:771), para. 51

\(^{104}\) Breyer, para. 36

\(^{105}\) Breyer, para. 16

\(^{106}\) Breyer, para. 35, 37
or another user of a computer bearing the IP address is not directly revealed.\textsuperscript{107} Hence, the
central issue before the Court was whether such an IP address constitutes information relating to an
\textit{identifiable} natural person in relation to the website provider where the additional data
necessary for identification of the website visitor were held by the visitor’s internet service
provider.\textsuperscript{108}

This time, the Court skipped the usual reference to the scope of the Directive which is very
wide and the personal data covered by the Directive are varied.\textsuperscript{109} The argument developed in
several steps: first, to be considered personal data, it is not necessary for information on its
own to allow identification of the data subject (follows from the definition of personal data as
relating to identifiable person, either directly or indirectly).\textsuperscript{110} Second, neither is it necessary
“that all the information enabling the identification … must be in the hands of one person”\textsuperscript{111}
which follows from Recital 26 DPD and its two-fold standard of (1) “all the means likely
reasonably likely to be used”, (2) “either by the controller or by any other person.”\textsuperscript{112} Here the
judgment follows the objective, or absolute, approach to defining identifiability as WP29
advances in WP 136.

The Court further moved to assess “whether the possibility to combine a dynamic IP address
with the additional data held by the Internet service provider constitutes a \textit{means likely
reasonably to be used} to identify the data subject.”\textsuperscript{113} The Court took over the AG’s argument
that the possibility to combine a dynamic IP address with additional data \textit{would not be reasonably likely}
if it was “prohibited by law or practically impossible” due to “a disproportionate effort in terms of
time, cost and man-power, so that the risk of identification appears in reality to be insignificant.”\textsuperscript{114} Although the German law does not allow the Internet
service providers to transfer data to the website providers, in case of cyber attacks the website
providers can contact a competent authority who can obtain information from the Internet
service providers to start criminal proceedings. Hence, the website provider is found to have
the means likely reasonably to be used to identify the website visitors on the basis of a
dynamic IP address with the help of third parties, namely, the Internet service provider and
the competent authority.\textsuperscript{115} Hence, the dynamic IP addresses were found to be personal data in
this case.

While agreeing with WP29 on the outcome (WP29 considers both dynamic and static IP
addresses personal data in WP 136\textsuperscript{116}), the Court seems to adopt a more measured approach to
what \textit{means of identification} are reasonably likely to be used. Namely, to the cost and effort
of identification that WP29 includes among the factors that need to be taken into account, the
Court adds the factor of legality. By itself, it would not have any restrictive effect on the
scope of personal data compared to WP 136. Indeed, a legal possibility for the controllers to
obtain additional information that enables them to identify data subjects is merely one of “all
the factors at stake” that should be taken into consideration to assess the possibility of
identification according to WP29.\textsuperscript{117} The restrictive potential of \textit{Breyer} lies in the assessment
that starts from what would be \textit{unreasonable}, where the Court follows the AG’s opinion that
identification would not be reasonably likely if prohibited by law. This seems to be in
contradiction to WP 136 which, among the relevant factors, names the possibility of

\begin{itemize}
\item \textsuperscript{107} Breyer, 38
\item \textsuperscript{108} Breyer, 39 (emphasis added)
\item \textsuperscript{109} like in Österreichischer Rundfunk and Others, paragraph 43; Lindqvist, paragraph 88; and Case
\item \textsuperscript{110} Breyer, paras. 40–41
\item \textsuperscript{111} Breyer, 43
\item \textsuperscript{112} Breyer, para. 42 (emphasis added)
\item \textsuperscript{113} Breyer, para. 45 (emphasis added)
\item \textsuperscript{114} Breyer, para. 46
\item \textsuperscript{115} Breyer paras. 46–49
\item \textsuperscript{116} WP 136, 16.
\item \textsuperscript{117} WP 136, 15-16
\end{itemize}
organisational dysfunction, which implies that identification can also result from data security breaches which result from illegal acts (e.g. hacking). It seems that a more nuanced argument would be that a legal prohibition to combine data for identification would make the means of identification “less reasonably likely to be used”, rather than “not reasonably likely”.

In addition, the Court does not consider the nature of processing in assessment of identifiability. Namely, under WP 136, the means of identification should be considered reasonably likely to be used when “the processing … only makes sense if it allows identification of specific individuals and treatment of them in a certain way”.118 It appears from the Court’s reasoning in para. 47 that, under the circumstances of the case, registering IP addresses for the purposes of dealing with cyber attacks implies identification, as there is an established mechanism of identification of the perpetrators by their IP addresses for the purposes of bringing criminal charges where the Internet service providers, competent authorities, and the website providers collaborate. In brief, while the case presented an opportunity for the Court to confirm or reject WP29 guideline “implied identification = reasonably likely identification,” the Court chose not to do so.

In sum, what is the impact of the Breyer case on the notion of personal data in part regarding identifiability? The case certainly reaffirmed the broad reading by WP29 of “all the means likely reasonably to be used either by the controller or by any other person.”119 Breyer was the first case where the Court explicitly stated that it is not necessary “that all the information enabling the identification … must be in the hands of one person”.120 At the same time, the ruling that a legal ban on identification would make the identification means not reasonably likely to be used is a notable deviation from WP 136 and has a potential to limit a number of situations where data is considered identifiable in Europe. Crucially, this potential should not be overestimated. This is mainly because the outcome of the assessment of what the Court considered “identification measures reasonably likely to be taken” was phrased by the Court in a manner specifically tailored to the narrow question posed, and the circumstances of the case. Citing the AG, it was not in dispute in this case “whether the classification of dynamic IP addresses as personal data is necessary as soon as there is a third party, irrespective of who it may be, capable of using those dynamic IP addresses to identify network users.”121 Answering the question posed this way negatively would indeed have significantly (and undesirably) limited the range of situations where information would be considered identifiable to a natural person, and hence personal data. However, the Court did not do so. Hence, WP 136 guidelines on identifiability have not been overruled and are still in play.

4 ‘Relating to’: YS and others and Nowak

The Court dealt with the meaning of “information relating to a natural person” in 2014 in the joint cases YS and others122, and – in a situation where a broad interpretation of personal data would lead to what the Court saw as unintended results – adopted a limited interpretation compared to WP 136.

a YS and others

On various stages of immigration process in the Netherlands, YS, M and S, third country nationals, requested access to the “minute” in their immigration file on the ground of the right of access to their personal data. The “minute” is a note drafted by an immigration case officer without authority to decide on an application for residence permit to motivate his or her

118 WP 136, 16
119 Breyer, para. 42 (emphasis added)
120 Breyer, 43
121 Opinion of Advocate General in Breyer, para. 50
122 Court of Justice of the European Union, Joint cases C-141/12 and C- 372/12 YS and M. and S. v. Minister of Immigration, Integration and Asylum, judgment of 17 July 2016, ECLI:EU:C:2014:2081 (hereinafter YS and others).
recommendation to a senior officer before he or she makes such a decision. The minute
generally contains three types of information: (1) information relating to the case officer, i.e.
name, telephone and office number; (2) “data relating to the applicant”, including name, date
of birth, nationality, gender, ethnicity, religion and language; procedural history; statements
the documents submitted; and the applicable law; and (3) the “legal analysis”, i.e. “an
assessment of the foregoing information in the light of the applicable legal provisions.”

Until 14 July 2009, it was the policy of the Dutch government to make a copy of the minute
available upon request. Since then the policy has changed, and the applicant now receives “a
summary of the personal data contained in the document,” excluding the “legal analysis”. YS, M and S were refused a copy of the full minute, and two sets of proceedings in Dutch
courts have led to the references for a preliminary ruling by the Court of Justice. One of the
central issues before the Court was if the legal analysis in the minute should be considered
personal data. The question is likely to have been answered positively following WP 136:
not in itself constitute such data”. The Court followed AG Sharpston and used two arguments to support its interpretation: first,
“a legal analysis is not information relating to the applicant … , but at most, in so far as it is not … a purely abstract interpretation of the law, is information about the assessment and application … of that law to the applicant’s situation;” second, granting access rights in
legal analysis is not compatible with “the objective and general scheme” of DPD. According to the Court, the data subject’s rights in DPD exist, among others, so “that person
may be certain that the personal data concerning him are correct and that they are processed
in a lawful manner”, and the right of access exists to enable the data subject “to carry out the
necessary checks … to obtain, depending on the circumstances, the rectification, erasure or
blocking of his data” Legal analysis “is not in itself liable to be the subject of a check of its
accuracy by that applicant and a rectification”. Crucially, granting such a right of access
“would not in fact serve the directive’s purpose of guaranteeing the protection of the … right
to privacy with regard to the processing of data relating to him, but would serve the purpose
of guaranteeing … a right of access to administrative documents, which is not however
covered by Directive 95/46.”

Consequently, the Court appears to interpret “information relating to” narrowly, as
information about an individual, and rejecting a broader approach of WP 136 where
information can also “relate to an individual” indirectly, in purpose or effect. The Court
does not refer to WP 136 in its judgment, or phrase the issue in similar words, i.e. as a choice
between “information about an individual” and “information relating to an individual in
purpose or effect”. However, it specifically rejects the argument brought on behalf of M and S
that the legal analysis should be considered as relating to an individual “if it is decisive for
the assessment of the application … and is applied to the specific case of the applicant.”

---

123 YS and others, para. 14
124 YS and others, para. 16-17
125 Since the status of “the data relating to the applicant” as personal data was not in dispute (YS and
others, paras. 33-34, 38.
126 WP 136, 10
127 YS and others, para. 39 (emphasis added)
128 YS and others, para. 40
129 YS and others, para. 41
130 YS and others, para. 44
131 YS and others, para. 45
132 YS and others, para. 46
133 See Part III.4
134 YS and others, para. 35
and hence also rejects the WP136 standards of “relating to” in purpose and effect. This is even more evident considering that the Court follows and explicitly refers to AG Sharpston’s opinion. While not in the paragraph cited by the Court, AG Sharpston is clear that “[i]n her opinion, only information relating to facts about an individual can be personal data. Except for the fact that it exists, a legal analysis is not such a fact.”\(^{135}\) While the Court did not adopt this distinction between the facts and analysis, which would restrict the meaning of “information relating to” even further, its narrow interpretation is likely to limit the range of situations within the scope of data protection law.

At the same time, YS and others does not shut the door for the broader interpretation of “information relating to” consistent with WP 136. In fact, the impact of YS and others can be restricted to similar situations only, i.e. where legal analysis is involved, and not affect situations which, while involving information used for assessment, are somewhat different from the facts in YS and others, as it appears to be the case in Nowak.

b. **Nowak**

The Court will soon revisit the meaning of “information relating to” in Nowak. Similar to YS and others, in a case before the Irish Supreme Court Mr Nowak requested access to his examination script based on a right to access to personal data. In its reference for a preliminary ruling the Irish court essentially asks whether the exam script containing the candidate’s answers may constitute personal data.\(^{136}\) The Court has not ruled on the case yet, but the opinion of AG Kokott was delivered on 20 July 2017.

The AG’s argument links to the two-fold argument in YS and others. First, the AG examines what kind of a relationship between information and the individual is relevant in the context of the examination answers and the examiner’s corrections on the answers. In both contexts the AG’s reasoning is consistent with the WP29 approach to consider information as personal data when it is processed with a purpose to evaluate the status or behaviour of an individual.

Although the answers are “formulated in abstract terms or relate to hypothetical situations”\(^ {137}\), “the script is a documentary record that that individual has taken part in a given examination and how he performed”.\(^ {138}\) While “the extent of the link between an examination candidate and his performance in an examination” varies,\(^ {139}\) “in every case, the aim of an examination […] is to identify and record the performance of a particular individual, i.e. the examination candidate. Every examination aims to determine the strictly personal and individual performance of an examination candidate.”\(^ {140}\) Therefore, contrary to the position of the Irish Data Protection Commissioner, the personal data in the script “is not confined to the examination result, the mark achieved or even points scored for certain parts of an examination.”\(^ {141}\)

A similar assessment is made regarding the examiner’s corrections on the examination script. The AG explicitly notes that Nowak resembles the case of YS and others,\(^ {142}\) and “at the first glance” the Court’s finding concerning legal analysis can be transposed to the examiner’s corrections, namely that it is not information relating to the examination candidate, but “at most information about the assessment”.\(^ {143}\) However, “the purpose of comments is the evaluation of the examination performance and thus they relate indirectly to the examination candidate. … [C]omments on an examination script are typically inseparable from the script

---

\(^{135}\) Opinion of AG Sharpston in YS and others, para. 56 (emphasis added)

\(^{136}\) Opinion of AG Kokott in Nowak, para. 2

\(^{137}\) Opinion of AG Kokott in Nowak, para. 19

\(^{138}\) Opinion of AG Kokott in Nowak, para. 21

\(^{139}\) Opinion of AG Kokott in Nowak, para. 23

\(^{140}\) Opinion of AG Kokott in Nowak, para. 24

\(^{141}\) Opinion of AG Kokott in Nowak, para. 27

\(^{142}\) Opinion of AG Kokott in Nowak, para. 58

\(^{143}\) Opinion of AG Kokott in Nowak, para. 59
Itself. Precisely because of that close link between the examination script and any corrections made on it, the latter also are personal data of the examination candidate [...].” 144 Consequently, the AG adheres to the WP29’s understanding of “information relating to an individual” in terms of purpose twice in one case, with regard to the examination answers and the examiner’s corrections.

Finally, AG Kokott considers the relationship between the concept of personal data and the purpose of the right of access which is the second pillar of the Court’s position in YS and others. The AG does not expressly reject the reasoning in YS and others as she notes that the examination candidate “has a legitimate interest, based on the protection of his private life” to be able to object to the processing of his or her examination script outside the examination procedure.145 Yet, her further statements are unequivocal: the purpose of the right of access under DPD does not “preclude the classification of an examination script as personal data”,146 “the issue of right of access is only secondary” to the interpretation of the concept of ‘personal data’,147 and “the classification of information as personal data cannot be dependent on whether there are specific provisions about access to this information which might apply in addition [...]. [N]either can problems [...] with the right of rectification be decisive in determining whether there exists personal data.”148

In other words, AG Kokott gracefully incorporates YS and others in her reasoning to the effect that the broad interpretation of “the information relating to an individual” in terms of purpose stands. Should the Court follow her opinion, it would seriously limit the restrictive potential of YS and others, and allow for the further use of the wide interpretation of “relate to” under WP136.

To sum up, the Court generally supports the broad interpretation of identifiability in Breyer, but adopts a restrictive view on what ‘information relating to’ an individual means in YS and others, suggesting that information only relates to an individual when it is about him or her. Yet YS and others does not reject a broad interpretation unequivocally and leaves a window for including information that relates to an individual in purpose or effect, like the AG suggests in Nowak. As to what ‘information’ in the definition of personal data means, the Court has not yet had a chance to express itself on this matter. Historically, the Court has considered the material scope of the DPD as very wide. It has not been inclined to narrow it down because the legal rights and obligations that follow appear difficult to comply with, ‘disproportionate’ or ‘unreasonable’, regarding those matters as secondary to the matter of scope. As a result, the broad interpretation of personal data in WP 136 still stands as a primary guideline for compliance with the data protection law. To restate the devil’s advocate’s argument, weather can still be plausibly considered personal data in Europe. The next part argues that this situation is not sustainable much longer.

V Discussion: the unbearable burden of the data protection law?

Should the European data protection law faithfully follow the broad interpretation of personal data as WP29 offered in 2007, and should the Court in Nowak adopt a broad view on how information relates to people to warrant the application of data protection law? The Court has faced this choice in Breyer regarding the identifiability criterion, is currently facing it in Nowak regarding the meaning of ‘relating to’ and is bound to face it in the near future regarding the meaning of information, for instance, in cases involving human tissue samples.

144 Opinion of AG Kokott in Nowak, para. 61-63
145 Opinion of AG Kokott in Nowak, para. 26
146 Opinion of AG Kokott in Nowak, para. 31
147 Opinion of AG Kokott in Nowak, para. 32
148 Opinion of AG Kokott in Nowak, para. 34
This Part briefly considers the ultimate consequence of maintaining the broad concept of personal data in the context of high intensity of compliance under the GDPR, i.e. eventual collapse of the data protection law.

In addition to the wide interpretation of personal data and a broadly defined material scope, two other factors are key to the ‘system overload’: broad personal scope of the data protection law and the high intensity of data protection compliance under the GDPR.

As to the personal scope, both WP29 and the Court took a broad approach to the concept of the data controller thereby widening considerably the range of actors subject to data protection obligations. Controller is the natural or legal person which alone or jointly with others determines the purposes and means of the processing of personal data, and is the primary bearer of the data protection burden. While a natural strategy of avoiding the data protection obligations was to argue that an entity at hand does not have control over data processing and is a mere processor, WP29 explained that the legally relevant control stems not only from the formal arrangements, such as law and contract, but also from the factual influence one has over data processing.

This assessment allows for the drawing of external conclusions, assigning the role and responsibilities of controller to one or more parties. This might be particularly helpful in complicated environments, often making use of new information technologies, where relevant actors are often inclined to see themselves as "facilitators" and not as responsible controllers.

In another opinion, in the context of the social network providers (‘SNS’), WP29 reasons that these are data controllers, since ‘they provide the means for the processing of user data and provide all the “basic” services related to user management (e.g. registration and deletion of accounts).’ An additional factor determining the status of the SNS as a controller is that SNS providers ‘also determine the use that may be made of user data for advertising and marketing purposes - including advertising provided by third parties.’

In Google Spain, the Court adopted a similarly broad approach to controllership in relation to the search engine providers and personal data uploaded on the third party websites, arguing that a narrow interpretation would be contrary to the DPD’s objective ‘to ensure … effective and complete protection of data subjects’, even in a situation where personal data is uploaded on a third party website and the controller has no knowledge of the data. Notably, as decisive for establishing the factual influence over data processing, the Court cited the fact that ‘processing [by the search engine] … can be distinguished from and is additional to that carried out by publishers of websites.’ This and WP29’s standards of the factual influence amounting to controllership could be applied to any entity with a business model consisting of providing somehow structured or new data processing services with a predetermined user interface, which makes it ‘distinguished from and additional to’ what their clients would do with the data otherwise.

---

150 Art. 2(d) DPD; Art. 4(7) GDPR
152 WP169, 11
155 Google Spain, para. 34
156 Google Spain, paras. 24, 34
157 Google Spain, para. 35
These entities would then be fully bound by the full scale of data protection obligations, which brings us to the point of high intensity of compliance. The current data protection regime under DPD already establishes both ex post and ex ante controller’s obligations (e.g. to define a purpose and a suitable legal ground, assess legality, fairness and necessity of data processing before it commences, or obligations to correct or erase data after). The GDPR relies on the culture of on-going compliance even more heavily; think inter alia of the principle of accountability which requires not only to comply with the GDPR but also demonstrate the compliance; or the data protection impact assessment that implies on-going assessment of risks of data processing, and taking risk mitigating measures before, during and after the data processing takes place.

Broad material and personal scope of the European data protection law combined with the high intensity of compliance under the GDPR, to exaggerate it ever so slightly, make data protection apply to nearly anyone processing nearly any information at nearly any time, and the threat of serious sanctions omnipresent. Going back to Lon Fuller and his Morality of Law, complying with and enforcing such a law in a meaningful way is impossible, and something – enforcement or compliance – will have to give. Facing the threat of effective and deterring sanctions, the controllers, instead of engaging in a meaningful assessment of fairness and necessity, will be pushed to create formal appearance of compliance by using ‘compliance surrogates’, like compliance roadmaps, ‘accountability tools,’ trust marks and certification schemes. The latter will likely be provided by private entities - law firms, consultancies and audit firms – who are in the business of selling reassurances more than safeguarding data protection interests, and whose products are bought to validate and broadcast compliance to the shareholders, data subjects and the data protection authorities. Alternatively, or in addition, enforcement of the data protection rights and obligations will likely become selective, determined by priority lists and shortcuts that the data protection authorities would develop to cope with the workload, including the ‘compliance surrogates’. The selective enforcement and pretend compliance will reinforce each other, amounting to a ridicule of the data protection law.

Another problem that has already manifested itself but will likely be further exacerbated is that a broad definition of personal data is conducive of considering a growing number of information technology-related problems, such as behavioural advertising, profiling, and data portability, as data protection problems. This risks to force the same data protection mechanisms on a set of very different information-related problems, what one may call ‘pouring new wine in old bottles’, and distort even further the ‘identify’ and goals of the data protection law.

Altogether, the broad concept of personal data would turn the data protection law in an uneconomical exercise of regulating everything, and deprive its protection of meaning.

158 Koops criticizes GDPR for creating equal data protection regimes for different data processing situations (BJ Koops, ‘The trouble with European data protection law’ International Data Privacy Law, Volume 4, Issue 4, 1 November 2014, Pages 250–261)
159 Art. 5(2) GDPR
160 Art. 35 GDPR et seq.
161 Think of the administrative fines in Art 83(4) GDPR of “up to 10 000 000 EUR, or in the case of an undertaking, up to 2 % of the total worldwide annual turnover of the preceding financial year, whichever is higher”.
VI  Conclusion: what to do about personal data?

This contribution examined the notion of personal data in the EU data protection law and concluded that its flexible definition in the DPD and the GDPR, its broad interpretation by WP29 and indecisive case-law of the Court of Justice make it possible to construct a plausible argument that everything is personal data. The rapid progress of (de)identification technologies and failure of anonymisation certainly play an important role in this argument. However, the main contribution of this article was to argue that the other two elements of the definition, i.e. ‘information’ and ‘relating to a natural person’, are problematic as well. Since the legal definition of personal data is not based on any articulated understanding of information, and we are currently able and willing to harness data from literally everything, everything can be argued to be or contain information. Furthermore, when increasing amounts of data are gathered in real time from increasingly smart environments that assess or affect us, and when we cannot possibly know how the autonomous computers draw meaning from data, we should always reasonably assume that – in the computer’s ‘mind’ – any information is likely to relate to a person, since we cannot eliminate this possibility with certainty.

While I must admit I have a heartfelt sympathy for the protective rationale behind the broad interpretation of personal data, I also believe that the good intentions will backfire. As this contribution argued, the system of legal protection based on such an all-encompassing notion is not going to be sustainable for long. So what other options do we have and what should we do with the concept of personal data?

One option would be to choose for a narrower meaning of personal data. However, this option has its own limitations. While the downside of having a too broad definition is stretching the scope of the data protection law too far, the risk of opting for a (too) narrow definition is incomplete protection of rights and interests that the data protection law was intended to provide. As the Court acknowledges, a broad concept of personal data and a very wide material scope serve the protective objectives of the data protection law. Granted, what those objectives are is subject of debate,¹⁶⁵ and the wording of the GDPR does not provide much clarification, referring to “the protection of the fundamental rights and freedoms of natural persons with regard to the processing of personal data”.¹⁶⁶ At the same, cutting back on the range of situations to which data protection law should apply cannot be done simply because a broader interpretation is unpractical. It requires careful thought, both as to what the protective objectives are, and where the boundaries of ‘personal data’ should lie to achieve those objectives. As discussion unfolds so far, is it not clear why the protection should be triggered by processing of information that is about a person, and not when the information is processed with the purpose to affect or assess a person, while the latter describes a significant part of the data processing situations which raise concerns. Likewise, it is doubtful that a database should be subject to the data protection regime when its owner has sufficient tools to identify people, and why it should be exempted when the database owner and the tools of identification are just a few transactions away from each other. In the age of software as a service, one does not need to be a ‘digital giant’ with own identification capacities, when one

¹⁶⁵ G Gonzalez Fuster, S Gutwirth ‘Opening up personal data protection: A conceptual controversy’ *Computer Law & Security Review*, Volume 29, 2013, pages 631-539; N Purtova ‘Default entitlements in personal data in the proposed Regulation: Informational self-determination off the table … and back on again?’ *Computer Law & Security Review*, Volume 30, Issue 1, February 2014, Pages 6-24, describing two conflicting reading of the rationale of data protection (“One such choice is the definition of the data protection regime as a tool to ensure individual informational self-determination and individual's control over his data on the one hand, or as an instance of ‘fair information practices’, i.e. a set of procedural rules channelling, not restricting, personal data flow and accommodating the competing interests of others in collecting, using and otherwise processing these data on the other hand.”)
¹⁶⁶ e.g. Recital 10 GDPR
can easily ‘borrow’ those on a pay-per-use basis. Think of the powerful data analytics tools such as IBM’s Watson\textsuperscript{167} becoming increasingly available to a growing range of users.

Another option would be to preserve the broad interpretation of personal data but reduce the intensity of compliance obligations, e.g. by matching the intensity to risks. Koops considered an option of different sets of obligations for different kinds of personal data an improvement on the compliance regime under the GDPR.\textsuperscript{168} Schwartz and Solove similarly proposed to treat information with varying degrees of identifiability differently.\textsuperscript{169} The GDPR arguably implements this idea, \textit{inter alia}, in the qualified obligation to perform data protection impact assessment under Article 35, and Article 11 exemption from the data protection obligations which require identification when such identification is not necessary for the purpose of processing. However, this option raises the same issue as the first one: how do we know where to draw the boundaries between the compliance regimes of different intensity, and how do we know a particular configuration will meet the protective objectives of the data protection law?

I believe that there is a third option which too deserves a careful consideration, \textit{i.e.} to abandon the concept of personal data as a cornerstone of data protection altogether, and seek remedies to the ‘information-induced harms’ without a sentimental attachment to this familiar proxy.

Whichever option it is, it is essential for success of the future legal protection against ‘information-induced harms’ that it is grounded on a better understanding of information and its relationship to people, well-understood and articulated problems it intends to solve and the fully mapped mechanics of the information-induced harms it intends to avert.

**Acknowledgements**

This contribution reports on the results of the project “Understanding information for legal protection of people against information-induced harms” (‘INFO-LEG’). This project has received funding from the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation programme (grant agreement No 716971). The paper reflects only the author’s view and the ERC is not responsible for any use that may be made of the information it contains. The funding source had no involvement in study design, in the collection, analysis and interpretation of data, in the writing of the report, and in the decision to submit the article for publication. I am grateful to Prof. Ronald Leenes and Prof. Bert-Jaap Koops for their helpful comments on the ideas developed in this paper.

\textsuperscript{167} https://www.ibm.com/watson/
\textsuperscript{168} BJ Koops, ‘The trouble with European data protection law’ International Data Privacy Law, Volume 4, Issue 4, 1 November 2014, Pages 250–261