Editorial

The current legal regime distinguishes between different types and categories of data. In general, the more personal, private and sensitive data are, the higher the level of protection provided. Among others, the legal regime differentiates between non-personal data and personal data, between metadata and content data and between non-sensitive and sensitive personal data. But there are at least three reasons why these legal categories may become redundant in the age of Big Data, artificial intelligence and increasing computational power.

First, categorising data only works when the status of the data is relatively stable, while in the current and future technological environment, it is likely that their nature will be highly volatile. A dataset that contains ordinary personal data may be linked and enriched with another dataset and transformed into a set that contains sensitive data; the data may then be aggregated or striped from their identifiers and become non-personal data; subsequently, the data may be deanonymised or integrated into another dataset containing personal data. These subsequent steps may happen in a split second. For example, when discussing the groups and categories in Big Data processes, it has been suggested that

> in the big data era, groups are increasingly fluid, not only through their changing membership, but also because of the changing criteria for the group itself. A group, the criteria for grouping people and the membership of a group might change in a split second. The purpose for which the group is designed may also change from day to day to adapt to new insights gained from data analytics, and groups may be formed and dissolved through the push of a button.[1]

This means that it becomes increasingly difficult to work with and uphold the various categories used in the law. The question is not only what falls under the definition of 'personal data', 'metadata', 'anonymous data' or 'sensitive personal data'; the point is that even although it might theoretically be possible to determine the status of a datapoint at every specific moment in time, such would be undoable in practical terms and defeat its purpose in legal terms, because applying a level of protection to a dataset at a specific moment in time is fruitless if its status is changed within a split second, potentially even a number of times. To draw a comparison. The question is not whether a caravan can, under specific conditions, be considered a home deserving protection under the right to privacy, Article 8 European Convention on Human Rights (ECHR). The question is whether it makes sense to work with a concept of home, as distinguished from non-homes, when a specific building is to be considered a home at one moment in time, a business premises the next sec-

DOI: 10.21552/edpl/2019/3/3

[1] L Taylor, L Floridi and B van der Sloot (eds), *Group Privacy* (Springer 2017) 284.

ond, then a sex shop, a hospital the next, a home again the next second, and so forth.

Second, categorising data only works when it is possible to determine with relative certainty in which category data fall, while this will be ever more difficult because the sensitivity of the data is less and less a quality of the data and more and more a result of the efforts invested by parties having access to the data. The definitions used in the current legal framework include a prospective element. For example, the definition of personal data contains reference to 'identifiable information', which means that data that at this moment in time do not identify anyone, but may do so in the future, will be considered personal data nevertheless when identification would cost relatively little effort. The other way around, in order to answer the question whether data should be considered anonymous, account should be had of the efforts and investments needed to deanonymise the data.[2]

Big Data has a number of important consequences for this constellation. Not only is it possible to change the status of data and datasets within a split second, due to the massive computational power and artificial intelligence, undoing these changes is also increasingly easy and cheap. Already in 2010, Paul Ohm conducted a study on anonymisation techniques and discussed three cases in which organisations had made public databases which had been stripped from all identifiers; in each case, third parties, such as academics and journalist, where able to re-identify the people in that database by combining those data with other data. Because of the increased technological powers to harvest indirect identifiable data and to combine existing databases with other open data sources, Ohm was convinced that in order to truly make a dataset anonymous, it has to be stripped from almost all data, hence arriving at the conclusion: 'Data can be either useful or perfectly anonymous but never both.'[3]

With the push to create an open data environment, in which datasets are published and made available for re-use,[4] still other datasets are available upon purchase and a high number of born digital data are generated on public websites, open discussion fora and social network sites, enriching, merging and combining existing datasets becomes increasingly easy. The fact that technologies are ever more potent and the costs for operating algorithms have dwindled means that both data and data-driven technologies are democratised.

Two things may be tentatively stipulated. First, given the democratisation of technologies and the minimal investment needed, it is increasingly likely that whenever a data-

[2] Article 29 Data Protection Working Party, Opinion 4/2007 on the concept of personal data (20 June 2017) 01248/07/EN WP 136 <https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2007/wp136_en.pdf> accessed 21 September 2019. Article 29 Data Protection Working Party, Opinion 05/2014 on Anonymisation Techniques, 0829/14/EN, WP216, 10 April 2014 <https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2014/wp216_en.pdf> accessed 21 September 2019.

[3] P Ohm, 'Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization' (2010) 57 UCLA Law Review 1701, 1704.

[4] See inter alia, Directive 2003/98/EC of the European Parliament and of the Council of 17 November 2003 on the re-use of public sector information.

base is shared, published or available upon request or payment, there will be a party that will combine those data with other data, enrich them with data scraped from the internet or merge them into an existing dataset. Thus, although there is no certainty, it is increasingly likely that if an anonymised dataset is made available, there will be some party around the world that will de-anonymise it or combine the data with other data in order to create personal profiles; that when a set of personal data is shared, there will be some party that will use those data to create a dataset with sensitive personal data; etc. Second, there will be other parties that have access to those data but will not use the data, use them as they are made available or even de-identify a database containing personal data. Who will do what is unclear beforehand.

Applying the current legal categories strictly might mean that indeed almost all data should be seen as personal data and potentially as sensitive personal data. In addition, because data are increasingly available, shared and made public, the same database will have multiple legal statuses at the same time. To draw from the analogy of the protection of the home again, the difficulty is not only, as described with the first argument, that the status of a building can change in a split second from a home to an office building to a fitness club to a private sex club to a home again. In addition, when determining whether a building should deserve the protection of a home, its future use should be taken into account; and while it is unknown whether the building will be used in the future as a home is unclear, it is increasingly likely that it will, though by whom is uncertain. Furthermore, the same building may have multiple functions for multiple parties at the same time, being a home to some, a restaurant to others, etc.

Third, the underlying rationale for providing different regimes of protection to different categories of data is that the more directly data or datasets are linked to an individual and the more sensitive the data are, the higher the level of protection provided. To give an example, one of the first legal instruments to introduce the category of sensitive personal data was the Council of Europe's 1981 Convention. This introduction was elucidated in the explanatory memorandum in the following way:

> While the risk that data processing is harmful to persons generally depends not on the contents of the data but on the context in which they are used, there are exceptional cases where the processing of certain categories of data is as such likely to lead to encroachments on individual rights and interests. Categories of data which in all member States are considered to be especially sensitive are listed in this article.[5]

This underlying rationale may become redundant over time. To provide an example, metadata can be just as revealing as content data, not just because they can reveal the content,[6] such as when a person visits a website with a xxx domain extension or when

[5] Explanatory Report to the Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data [1981].

[6] B Greschbach, 'The Devil is in the Metadata – New Privacy Challenges in Decentralised Online Social Networks' <http://www.nada.kth.se/~gkreitz/metadata/sesocMetaPrivacy.pdf> accessed 21 September 2019.

a person sends a letter to the national cancer institute, but also because they reveal other information that may be even more sensitive than content data. The type of videos a person watches on a porn site may reveal one thing, the fact that that person either visits a porn site once a year or twice a day may reveal more. What a person says to her mother over the telephone may reveal one thing, the fact that a person either spends two hours a day over the telephone or calls her mother once a year on her birthday may say more.

Not surprisingly, companies and governmental organisations increasingly rely on gathering metadata instead of content communication, both because processing these types of data is subject to less restrictive rules and regulations and because the analysis of these types of data often yields more valuable results than the analysis of content communication data, among others because fewer datapoints are needed and because the datapoints are less ambiguous. In order to have an algorithm analyse content communication data, the program should be relatively well apt to understand natural languages used within specific contexts. Far easier is it to create a heat map of where people go, how long they stay in specific places and who else is there, or of which sites they visit, on what items they click, how long they stay on a specific page; etc.

Reference can also be made to the legal differentiation between personal and non-personal data, such as aggregated data. Increasingly, data-analytics programs operate on anonymised and aggregated data or data that never were personal data. The correlations and group profiles found in Big Data may have as relevant determinant personal identifiers, but are often based on non-personal datapoints, such as zip codes. Obviously, when such categories are used to the disadvantage of specific individuals, one may argue that data profiles should be considered personal data again. The classic reference here is to *redlining,* in which banks' policy on giving out loans was based on zip code areas and that policy disadvantaged people living in neighbourhoods with a large African-American community. When a specific person is denied a loan on the basis of such a profile, it could be argued that this involves processing personal data.

Still, under such an approach, it is possible to design and make policies that affect groups of people on the basis of general information that were never personal data and may not have an effect on specific individuals, but on large groups or everyone living in society. For example, suppose an algorithm produces the result that one of the most effective ways to combat nightlife violence in a city is to spray tangerine smell between 22.00-04.00 in nightlife areas, because this makes people less aggressive. No personal data are processed, though such policies may have a high impact on people's lives.

In addition, because data protection regimes rely on the connection of the data to individuals and individual interests, two parts of the Big Data process are left unregulated. On the one hand, the gathering of non-personal or aggregated data is not regulated and on the other hand, the analysis of data, finding correlations and creating group profiles, is left unregulated, because Big Data analysis typically revolves around computation on aggregated datasets. This holds true for the human rights regime in general. Referring to the example of redlining, the core of the problem is not that this particular black person is disadvantaged by the policy of the bank, but that the algorithm, the data or both are biased in a way that discriminatory policies emerge. Working with a biased dataset or a biased algorithm is currently not prohibited or sanctioned, because analysing biased data or using biased algorithms as such does not harm any specific individual.

To refer to the metaphor of the home yet again, the reason for giving the home a special status was that within the private sphere, private matters were discussed, intimate actions took place and personal items were stored. If we are moving towards a world in which intimate actions take place irrespective of the physical domain, in which private discussions take place in open fora and in which personal items are stored in the cloud, then the question is whether the rationale behind the distinction between the private and the public domain, between the home and the non-home is still valid. The same holds true for the data categories in law. If processing metadata can be just as or even more revealing than processing content data, if non-sensitive personal data can be put together in a way that it gives a highly intimate picture of a person's life, if non-personal data can be used in ways that have far greater impact on the lives of ordinary citizens than the processing of sensitive personal data, the question is whether the underlying rationale for the categorisations should be upheld.

If these arguments hold true, two conclusions could be drawn. First, basing the level of regulatory protection on the status and nature of data is not the best way forward. Second, given the fact that non-personal data may be changed to sensitive data in a split second and that processing non-personal data can have a bigger impact on persons' lives than the processing of sensitive personal data, as long as the legal regulation *is* based on the status of data, it should provide for a basic framework for the protection of citizens' interests vis-a-vis the processing of non-personal data. This conclusion contrasts sharply with the approach taken by the European Union in 2018, when it adopted a Regulation on the transfer of non-personal data, which only aims at stimulating cross-border data processing, without providing any form of protection to citizens. Article 1 of that Regulation specifies:

> This Regulation aims to ensure the free flow of data other than personal data within the Union by laying down rules relating to data localisation requirements, the availability of data to competent authorities and the porting of data for professional users.[7]

The material provisions of the Regulation do not aim at restricting or laying down conditions for the processing or transfer of non-personal data, but in contrast, prohibit any type of restriction or limitation in national laws on the availability, transfer and processing of non-personal data.

[7] Regulation (EU) 2018/1807 of the European Parliament and of the Council of 14 November 2018 on a framework for the free flow of non-personal data in the European Union, art 1.

Given the previous, the EU might want to amend its approach and also provide protection to the interests of citizens when non-personal data are processed. The principles contained in the General Data Protection Regulation could serve as a source of inspiration. Although its material scope is determined by the identifiability of the data, many of the material principles in the GDPR do not so much aim to protect individual interests of specific data subjects, but lay down general duties of care and standards for good data governance by data controllers and can hence be transposed easily to the processing non-personal data.

For example, if an organisation collects more non-personal data than it needs for its specified purpose, given that these data may be converted in sensitive personal data and given that even the use of non-personal data can have a high impact on the lives of citizens, a data minimisation principle could be applied to processing non-personal data all the same. Having a specific purpose for gathering non-personal or aggregated data and limiting the use of the data to that specific purpose seems a basic requirement in the age of Big Data. Given that increasingly, decisions are made on the basis of non-personal data and aggregated datasets are used to design policies, it seems vital to ensure that those aggregated data are correct, complete and up to date. In addition, given the fact that having and processing non-personal and aggregated data potentially provides organisations with just as much power as processing personal data, requirements to ensure transparency seems vital. In addition, as the impact of data processing operations based on non-personal data can be significant, an impact assessment, also taking into account broader and societal interests, may be regarded as quintessential in the age of Big Data, which also holds true for the requirement to appoint a data protection officer. An obligation to ensure that the non-personal data are processed safely and securely, taking adequate technical and organisational security measures, having a data protection policy and embedding those principles in the technical infrastructure of an organisation by design or default ensures that non-personal data do not fall into the hands of unauthorised third parties. Finally, like the current General Data Protection Regulation does, a Regulation on the processing of non-personal data should contain a rule specifying that transferring non-personal data to other jurisdictions should be prohibited, unless similar rules are applied to the processing of non-personal data in that non-EU country or within that non-EU based organisation.

The two opinions in this edition continue to discuss the question of the regulation of non-personal and aggregated data. I'm proud that Aaron Fluitt, Aloni Cohen, Micah Altman, Kobbi Nissim, Salome Viljoen and Alexandra Wood, who were the stars of the last Privacy Law Scholars Conference in Berkeley, share with EDPL readers their thoughts on what they call data protection's compositions problem. They show how different data(sets) can be put together to arrive at a detailed personal profile and discuss, among other things, the implications for the level of protection provided to the processing and publication of datasets that in themselves do not or only marginally have an effect on citizens, but when put together, might. I'm equally proud that Dara Hallinan, whose doctoral thesis 'Feeding Biobanks with Genetic Data', is a must read for anyone interested in genetic data and biobanking, writes on a novel development in Germany, where there is discussion about applying the data protection framework to cameras that are installed, but are not switched on. Should the data protection framework apply to the non-processing of personal data as well?

Then there are five articles, whose authors virtually need no introduction. Chirstopher Docksey and Hielke Hijmans provide the reader with an overview of the recent trends in the case law of the Court of Justice; Paul de Hert and Juraj Sajfert discuss why the EU, in its recent laws and policies, has not made reference to Big Data and suggest that some actors in the policy field considered Big Data too dangerous while others have simply ignored the phenomenon; Bart Custers, Helena Vrabec and Michael Friedewald assess the legal and ethical impact of data reuse; Zohar Efroni, Jakob Metzger, Lena Mischau and Marie Schirmbeck delve into the topic of Privacy Icons; finally, Mark Leiser and Bart Custers evaluate consent in the Law Enforcement Directive.

The reports section led by Mark Cole, as always, deserves special mention. It contains one report on a European country, written by Ioannis Iglezakis on the judgement of a Greek Court on Messenger Messages and Facebook Photographs as Means of Evidence, and one report, written by Ashit Kumar Srivastava, on Data Protection Law in India. Our special GDPR Implementation Series focuses on Cyprus in a report written by Christiana Markou and revisits Germany where Christina Etteldorf informs us about the second round of adaptation legislation. The Practitioner's Corner, contains two reports on blockchain, one written by Jörn Erbguth and the other written by Rosanna Mannan, Rahul Sethuram and Lauryn Younge. In addition, there is a practitioner's report on data erasure written by Matthias Enzmann, Annika Selzer and Dominik Spychalski. We have a handful of very interesting case notes in the case note section led by Maja Brkan and Tijmen Wisman. Patrick Van Eecke and Anne-Gabrielle Haie evaluate the Advocate General Opinion on the *Planet49* case; Claudia Quelle discusses the Advocate General Opinion on *GC and Others v CNIL;* Mara Paun takes up the judgment of the ECJ in the Joined Cases T-639/15 to T-666/15 *Maria Psara and Others v Parliament* and T-94/16 *Gavin Sheridan v Parliament*; and finally, Marc Rotenberg and Bilyana Petkova analyse the recent Census 2020 case of *New York v Department of Commerce*. Lina Jasmontaite-Zaniewicz has written an insightful book review of Wolff's *You'll See This Message When It Is Too Late*, in the book review section led by Gloria González Fuster. Finally, as part of EDPL's cooperation with the 41st International Conference of Data Protection and Privacy Commissioners we are featuring interviews with privacy commissioners Andrea Jelinek (EDPB, Austria), Angelene Falk (Australia) and Stephen Wong (Hong Kong).

A final note, we are sad to inform you that Judith Rauhofer has stepped down as board member of EDPL. Judith was invaluable in starting the journal, conceiving its initial path and laying contacts with board members, authors and readership. We will always be grateful for Judith's invaluable work, intelligence and good spirits. At the same time, we are happy that Hielke Hijmans will accede to the board. Hielke perhaps needs no introduction. He was Head of Unit Policy and Consultations of the EDPS, he wrote the seminal book *The European Union as Guardian of Internet Privacy*, penned several articles in the Common Market Law Review and is now the Director at Belgian Data Protection Authority. We are proud to have him on board.

For those interested in submitting an article, report, case note or book review, please e-mail our executive editor Nelly Stratieva (<stratieva@lexxion.eu>) and keep in mind the following deadlines:

- Issue 2019/4: 1 October 2019 (Young Scholars Award);
- Issue 2020/1: 15 January 2020;
- Issue 2020/2: 15 April 2020;
- Issue 2020/3: 15 July 2020.

I hope you enjoy reading this edition of the European Data Protection Law Review!

Bart van der Sloot
Tilburg Institute for Law, Technology, and Society (TILT)
Tilburg University, Netherlands