German Law Journal (2021), **22**, pp. 1583–1612
doi:10.1017/glj.2021.79

CAMBRIDGE UNIVERSITY PRESS

ARTICLE

The Difficulty of Defining Sensitive Data—The Concept of Sensitive Data in the EU Data Protection Framework

Paul Quinn[1] and Gianclaudio Malgieri[2,*]

[1]Law, Science, Technology & Society, Vrije Universiteit Brussel, Brussels, Belgium and [2]EDHEC Augmented Law Institute, Lille, France
Corresponding author: paul.quinn@vub.be

(Received 14 May 2020; accepted 04 January 2021)

Abstract
The concept of sensitive data has been a mainstay of data protection for a number of decades. The concept itself is used to denote several categories of data for which processing is deemed to pose a higher risk for data subjects than other forms of data. Such risks are often perceived in terms of an elevated probability of discrimination, or related harms, to vulnerable groups in society. As a result, data protection frameworks have traditionally foreseen a higher burden for the processing of sensitive data than other forms of data. The *sui generis* protection of sensitive data—stronger than the protection of non-sensitive personal data—can also seemingly be a necessity from a fundamental rights-based perspective, as indicated by human rights jurisprudence. This Article seeks to analyze the continued relevance of sensitive data in both contemporary and potential future contexts. Such an exercise is important for two main reasons. First, the legal regime responsible for the regulation of the use of personal data has evolved considerably since the concept of sensitive data was first used. This has been exemplified by the creation of the EU's General Data Protection Regulation (GDPR) in Europe. It has introduced a number of requirements relating to sensitive data that are likely to represent added burdens for controllers who want to process personal data. Second, the very nature of personal data is changing. Increases in computing power, more complex algorithms, and the availability of ever more potentially complimentary data online mean that more and more data can be considered of a sensitive nature. This creates various risks going forward, including an inflation effect whereby the concept loses its value, as well as the possibility that data controllers may increasingly seek to circumvent compliance with the requirements placed upon the use of sensitive data. This Article analyzes how such developments are likely to influence the concept of sensitive data and, in particular, its ability to protect vulnerable groups from harm. The authors propose a possible interpretative solution: A hybrid approach where a purpose-based definition acquires a bigger role in deciding whether data is sensitive, combined with a context-based 'backstop' based on reasonable foreseeability.

Keywords: Sensitive Data; Data Protection; GDPR; Privacy; Health Data

A. Introduction

The concept of sensitive—or special categories of personal—[1] data has long been central to data protection frameworks. In recent times, the nature and use of sensitive data has been changing at a rapid pace. This article aims to analyze the evolution of the concept of sensitive data through past, contemporary, and potential future contexts. In particular, this Article will ask whether the

[1]This is the description used within the GDPR. *See, e.g.*, Regulation 2016/679 of the European Parliament and of Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), O.J. (L 119), art. 9 [hereinafter GDPR].

*The online version of this article has been updated since original publication. A notice detailing the change has also been published.

concept of sensitive data is and will remain fit for purpose. As section B will discuss, the EU's General Data Protection Regulation appears to claim that sensitive data must be regulated to avoid the risk of discrimination against vulnerable groups and individuals.[2] Such a vison seemingly regards sensitive data as a "means to an end." Importantly, however, elsewhere, including in pronouncements by the European Court of Human Rights (ECtHR), the need to maintain separate rules for the use of sensitive personal data is seen as "an end in of itself," falling under the fundamental right to privacy in Article 8 of the European Convention of Human Rights (ECHR).

Sensitive data as a concept within data protection has been evolving in both a *de jure* and a *de facto* sense. In terms of *de jure* changes, the concept has been expanded to cover new categories which have, to a large extent, arguably been fixed in law by the GDPR. In terms of a *de facto* evolution, increases in computing power, the availability of big data through *inter alia* the development of the Internet of Things (IoT)[3] and an ever-increasing level of interconnectivity—with the consequent increase in potentially complimentary data—mean that more and more personal data can be viewed as being of a sensitive nature. It seems likely that this trend will continue and intensify in the future. These changes raise the question as to what value the concept still has and whether it will retain its value in the future. This question is addressed in sections C and D?

In order to assess this, this Article will look at these changes in the context of recent innovations in the European data protection framework, in particular in light of recent changes made by the GDPR. These innovations include a reduction in the difference between explicit and non-explicit consent, discussed in section E, the need to appoint a Data Protection Officer (DPO), and a requirement to conduct a Data Protection Impact Assessment, discussed in section F, in many instances where sensitive data is used. As discussed in section G, these measures will arguably play a role in reducing the potential for resulting harms; but, if we consider the changing nature of sensitive data, such measures may also result in a number of negative consequences themselves. As section G discusses, these consequences include the risk of an "inflation effect," whereby the value of data protection efforts are called into question and a potential large increase in the burden on potential data controllers. In some contexts, this may encourage potential data controllers to make efforts to circumvent the requirements placed upon sensitive data. Such issues may call into question the value of the concept of sensitive data, especially in the long term where a number of these *de facto* changes are likely to intensify.

In order to moderate such problems, we suggest that it may be important to reconsider how personal data is defined. At present, the approach is mostly contextual in nature, whereby the question of sensitivity is judged objectively—taking into account the broader context in question. Relatively little attention is paid to the intentions of the data controller. If such an approach were to be continued in the future, the problems discussed in this Article are likely to be exacerbated. In order to avoid this, the authors of this article suggest employing a hybrid purpose-and-context-based definition, as stated in section H. Under such a vision, the aim of the data controller would be of central importance. Where it is intended to process data to draw sensitive conclusions or produce data that could reveal sensitive aspects, the data should be described as sensitive. However, where this is not the case, an objective contextbased "backstop" should be employed. This would involve a moderate analysis of the data in question in order to ascertain if it was reasonably foreseeable that the data in question could reveal sensitive information about data subjects. The aim of such a construction is to prevent a weakening of the value of sensitive data and with it the fundamental right-based need to protect sensitive data with higher *sui generis* safeguards.

[2]In recital 71 of the GDPR increased risks of discrimination are used as a justification for the concept of sensitive—or special—categories of data.

[3]*See* Omer Tene & Jules Polonetsky, *Big Data for All: Privacy and User Control in the Age of Analytics*, 11 NW. J. TECH. & INTELL. PROP., 239, 257 (2013).

https://doi.org/10.1017/glj.2021.79 Published online by Cambridge University Press

B. Sensitive Data in Law

I. "Means to an End" Justifications for the Existence of the Concept

Justifications for the concept of sensitive—or special categories of personal—data appear in the first international legal formulations of data protection, where one can find justification for a specific protection for sensitive data. Common themes for the existence of sensitive data often revolve around the need to prevent harmful forms of discrimination or related phenomena. The U.N., for example, issued Guidelines for the Regulation of Computerized Personal Data Files in 1990, justifying a further protection for sensitive data because such data are "likely to give rise to *unlawful or arbitrary discrimination.*"[4] Such a vision clearly views sensitive data as a "means to an end," or, in other words, being required to reduce the possibility of distinctive harms such as discrimination.

Similarly, almost two decades later, the GDPR[5] also provides a justification for the augmentation of requirements surrounding sensitive data. In particular, recital 51 explains that such data are, "by their *nature*, particularly sensitive in relation to fundamental rights and freedoms" and, thus, they "merit specific protection as the context of their processing could create *significant risks to the fundamental rights and freedoms.*"[6] In addition, recital 71, when addressing the right not to be subject to automated decisions with legal or similarly significant effects on individuals, explains that one significant concern is the possibility of discrimination based on sensitive data: "[D]iscriminatory effects on natural persons on the basis of racial or ethnic origin, political opinion, religion or beliefs, trade union membership, genetic or health status or sexual orientation, or that result in measures having such an effect." [7] In sum, the GDPR provides a specific protection for some types of personal data because their processing could produce significant risks to fundamental rights and freedoms, which includes—but is not limited to—discriminatory effects.[8]

A much more detailed justification was provided by the Council of Europe (CoE) Modernized Convention 108 on Automatic Processing of Personal Data in its Explanatory Report.[9] In general, it also acknowledges that sensitive data should be more protected because its processing "may lead to encroachments on interests, rights and freedoms."[10] The Explanatory Report argues that this "can for instance be the case where there is a potential risk of discrimination or injury to an individual's dignity or physical integrity, where the data subject's most intimate sphere, such as his or her sex life or sexual orientation, is being affected, or where processing of data could affect the presumption of innocence."[11] Accordingly, "in order to prevent adverse effects for the data subject" such processing "should only be permitted where appropriate safeguards, which complement the other protective provisions of the Convention, are provided for by law."[12] In sum, the CoE Convention Report does not mention only discrimination, but adopts a wider approach. It recognizes that the processing of sensitive data is more likely to have adverse effects on data subjects, in particular discrimination, but also injury to dignity or physical integrity, thus affecting their most intimate sphere, their presumption of innocence, etc.

[4]G.A. Res. 45/95, para. 5 (Dec. 14, 1990) (emphasis added).

[5]This was unlike its predecessor Directive 95/46/EC which did not give any justification/rationale for the existence of the concept of sensitive data.

[6]GDPR, *supra* note 1, at 51.

[7]GDPR, *supra* note 1.

[8]On the concept of "risks to rights," including discrimination, see Niels van Dijk, Raphaël Gellert, & Kjetil Rommetveit, *A Risk to a Right? Beyond Data Protection Risk Assessments*, 32 COMPUT. L. & SEC. REV. 286, 286–306 (2016), https://doi.org/10.1016/j.clsr.2015.12.017.

[9]*Explanatory Report on No. 223 of the Council of Eur. Treaty Series—Protocol Amending the Convention for the Protection of Individuals with Regard to Automatic Processing of Personal Data,* COUNCIL OF EUR., (2018), https://rm.coe.int/cets-223-explanatory-report-to-the-protocol-amending-the-convention-fo/16808ac91a.

[10]*Id.* at para. 55.

[11]*Id.*

[12]*Id.*

https://doi.org/10.1017/glj.2021.79 Published online by Cambridge University Press

Interestingly, the proposed e-privacy Regulation—in the version of the Commission Proposal—shows a broad justification for the protection of "sensitive information." Recital 2 states: "[T]he content of electronic communications may reveal highly sensitive information about the natural persons involved in the communication, from personal experiences and emotions to medical conditions, sexual preferences and political views, the disclosure of which could result in personal and social harm, economic loss or embarrassment."[13] Here, the term "discrimination" is not even mentioned, while the broad perspective—persona, social, economic harm, or embarrassment—is preferred.

The rationales put forward here are important, especially in the context of the ever-changing nature of sensitive data, which is discussed in section 8. The evolving nature of sensitive nature in *de facto* terms means that there is a constant need to ask whether sensitive data as a legal concept is able to prevent the kind of harms that have been discussed here —given that the need to prevent such harms is often presented as a justification for the existence of such measures.

II. The Protection of Sensitive Data as a Fundamental Right—as an "End in Itself"

As alluded to in recital 51 of the GDPR, discussed above, the need to protect personal data—and sensitive data in particular— from improper use also finds its grounding in fundamental rights. Privacy in general and data protection in particular are identified as fundamental rights in the EU's Charter of Fundamental Rights and Freedoms (ECFR).[14] The European Court of Justice has made the link between the need to protect sensitive data and fundamental rights in a number of cases.[15] In the recent case of *GC and Others v. CNIL,*[16] which concerned the application of the GDPR to search engines, the court stated:

> [A]n interpretation of Article 8(1) and (5) of Directive 95/46 or Article 9(1) and Article 10 of Regulation 2016/679 that excluded a priori and generally the activity of a search engine from the specific requirements laid down by those provisions for processing relating to the special categories of data referred to there would run counter to the purpose of those provisions, namely to ensure enhanced protection as regards such processing, which, because of the particular sensitivity of the data, is liable to constitute, as also follows from recital 33 of that directive and recital 51 of that regulation, a particularly serious interference with the fundamental rights to privacy and the protection of personal data, guaranteed by Articles 7 and 8 of the Charter.[17]

Perhaps of even more significance is Article 8 of the European Convention of Human Rights (ECHR) which mandates the protection of 'private and family life.' The ECHR is of general binding application on signatory states—unlike the ECFR which is limited in scope to the exercise of EU law. The case law related to Article 8 ECHR has been applied by the European Court of Human Rights (ECtHR) to ensure that individual privacy is respected in a wide array of contexts. This importantly includes an obligation to safeguard personal data in general and sensitive data in particular.

In a selection of seminal cases the ECtHR has demanded *inter alia* that a legal basis must exist for the processing of sensitive forms of data, that it should be outlined clearly in law and in a way

[13]Proposal for a Regulation of the European Parliament and of the Council concerning the respect for private life and the protection of personal data in electronic communications and repealing Directive 2002/58/EC (Regulation on Privacy and Electronic Communications). *See* GDPR, *supra* note 1, Recital 2.

[14]E.U. Charter of Fundamental Rights 326/391, arts. 7–8, 2012 O.J. (55).

[15]G. Gonzalez Fuster & R. Gellert, *The Fundamental Right of Data Protection in the European Union: In Search of an Uncharted Right*, 26 Int'l Rev. of L., Computs. & Techn. 73, 73–82 (2012).

[16]ECJ, Case C-136/17, GC and Others v. Commision nationale de l'informatique et des libertés (CNIL), ECLI:EU:C:2019:773, https://curia.europa.eu/juris/liste.jsf?num=C-136/17.

[17]Request for a preliminary ruling under Article 267 TFEU from the Conseil d'État (Council of State, France), made by decision of 24 February 2017, received at the Court on 15 March 2017, in the proceedings, Case C-136/17.

https://doi.org/10.1017/glj.2021.79 Published online by Cambridge University Press

that is foreseeable for data subjects.[18] In the case of *Z v. Finland* the court referred specifically to the need to have special and suitable legal frameworks in order to protect sensitive forms of data such as health data.[19] Unlike the "means to an end" vision of sensitive data outlined above, the ECtHR therefore appears to regard the existence of law regulating sensitive data to be an "end in of itself." This is significant because it implies that the existence of only a general framework to protect all personal data would be insufficient.[20] It is arguably necessary to have specific rules and requirements that safeguard the use of sensitive forms of data that exist in addition to those tailored to regulating the use of personal data in general. In cases such as *Z v. Finland*,[22] the authors of this Article would argue that the court therefore recognized the insufficiency of a single regime of data protection to protect the fundamental rights of individuals. This requirement is of central importance when analyzing the GDPR's approach to regulating the use of sensitive data. The requirements it poses with regards to the use of sensitive data should not merely be viewed as an extra or optional choice made by the EU, but rather, the creation of legislation that is needed to ensure that the fundamental rights of data subjects are protected. This arguably means that were such legislation not to function adequately—in other words, in creating a specific framework to protect sensitive personal data—the GDPR could no longer be considered as being able to protect the fundamental rights of individuals in terms of the processing of their personal data. This fundamental-rights-based requirement is accordingly something that must be taken into account when assessing continued fitness for purposes of the GDPR's approach to sensitive data, including, as the authors discuss in section D, in the context of an evolving world of personal data generation and use.

III. The Evolving Contours of Sensitive Data in Data Protection Law

1. Pre-GDPR

The concept of sensitive data is a bedrock of modern data protection. Its place within data protection law has not always been certain and discussions as to how the concept should be recognized continue to this day. Whilst few would argue with the abstract notion that some data is "more sensitive" than other data, it was historically difficult to find consensus on what types of data should be considered sensitive and how treaties and legislation should be worded to protect it.[22]

The concept of special categories or sensitive data was first proposed on the international stage by the Organization for Economic Co-Operation and Development (OECD) in its Guidelines on the Protection of Privacy and Trans-Border Flows of Personal Data.[23] These non-binding guidelines recommended that Member States of the OECD introduce the concept of sensitive data into national law dealing with the protection of data. Prior to this, Sweden and the German state of Hesse had already incorporated the concept into national and state law whilst many other states had not.[24] The OECD guidelines, though significant in terms of recommending the concept of sensitive data, were by no means definitive on the subject. The concept was not elaborated, and no effort was made to precisely outline which types of data should be considered as sensitive.

[18] *See, e.g.*, S. and Marper v. The U.K., App. Nos. 30562/04 & 30566/04 (Dec. 4, 2008), http://hudoc.echr.coe.int/eng?i=001-90051; Gaskin v. U.K., 12 Eur. Ct. H.R. 36 (1989), http://hudoc.echr.coe.int/eng?i=001-57491; Malone v. U.K., App. No. 8691/79 (Aug. 4, 1984), http://hudoc.echr.coe.int/rus?i=001-57533.

[19] Z. v. Finland, App. No. 22009/93 (Feb. 25, 1997), http://hudoc.echr.coe.int/rus?i=001-58033.

[20] L.H. v. Lat., App. No. 52019/07 (July 29, 2014), http://hudoc.echr.coe.int/eng?i=001-142673 (emphasizing the importance of protecting sensitive data).

[21] Z. v. Finland, *supra* note 20.

[22] *See, e.g.*, Karen McCullagh, *Data Sensitivity: Proposals for Resolving the Conundrum*, 2 J. Int'l Com. L. & Tech. 190, 190–201 (2007).

[23] *OECD Guidelines on the Protection of Privacy and Transborder Flows of Personal Data*, Organization for Economic Co-Operation and Development [OECD], http://www.oecd.org/document/18/0,2340,en_2649_34255_1815186_1_1_1_1,00.html.

[24] McCullagh, *supra* note 23.

https://doi.org/10.1017/glj.2021.79 Published online by Cambridge University Press

Indeed, the guidelines themselves state: "[I]t is probably not possible to define a set of data which are universally regarded as being sensitive."[25]

Whilst rather nebulous, the concept of sensitive itself has survived long after the OECD guidelines.[26] Another important step occurred with the creation in 1981 of the Council of Europe's Convention for the Protection of Individuals with Regard to Automatic Processing of Personal Data.[27] Unlike the OECD guidelines, it was considered binding upon signatories— though it did not require direct incorporation into national legal systems. It specified categories of data that were actually to be considered sensitive. These were listed as being personal data related to "racial origin," "political opinions or religious or other beliefs," as well as personal data concerning "health" or "sexual life."[28] Interestingly, however, these categories were not intended to be exhaustive. The explanatory report to the Convention explained that it was open to signatory states to create other categories of sensitive data in national law.[29]

The European Union Data Protection Directive 95/46/EC went further, clearly specifying the categories of data that should be considered as sensitive, outlining seven categories of data that warranted extra protection. Article 8 of the Data Protection Directive (DPD) stated that: "Member States shall prohibit the processing of personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, trade-union membership, and the processing of data concerning health or sex life."[30]

These categories were to find protection in all Member States, though the precise form that such protections should take was to a large extent left up to the jurisdiction in question. The result of such a formulation can be viewed from different angles in terms of the concept of sensitive data.[31] On the one hand, the creation of mandatory special categories of personal data meant that all Member States that were subject to Directive 95/46/EC, were required to create law protecting certain defined categories of sensitive data. This created certainty at least with regards to a minimum number of categories that would have to be protected. On the other hand, given the use of a Directive as the legislative form and the lack of any explicit mention otherwise, the possibility existed for Member States to add further categories of sensitive data. This occurred in a number of Member States where further categories were included, for example, genetic data as a specific category.[32] The choice of a directive therefore created problems in terms of heterogeneity—of legal systems—across Europe. This created problems for those who wanted inter alia to use sensitive data on a pan-European basis.[33] The result was considerable uncertainty for potential data controllers that wished to operate a cross European frontiers, for instance, for purposes of eHealth or in large scientific research consortia.[34]

[25]OECD Guidelines, *supra* note 24, at para. 19(a).

[26]*The Evolving Privacy Landscape: 30 Years After the OECD Guidelines*, Organization for Economic Co-Operation and Development [OECD], http://www.oecd.org/sti/ieconomy/49710223.pdf.

[27]Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data, Jan. 28, 1981, C.E.T.S. No. 108. http://conventions.coe.int/Treaty/EN/Reports/HTML/108.htm.

[28]*Id.* at art. 6.

[29]Council of Eur., Explanatory Report on No. 108 of the Council of Eur. Treaty Series—Convention for the Protection of Individuals with Regard to Automatic Processing of Personal Data, ¶ 9 (1981), https://rm.coe.int/16800ca434.

[30]Council Directive 95/46, art. 9, 1995 O.J. (L 281) (EC) (emphasis added).

[31]McCullagh, *supra* note 23; Dara Hallinan, Michael Friedewald, & Paul De Hert, *Genetic Data and the Data Protection Regulation: Anonymity, Multiple Subjects, Sensitivity and a Prohibitionary Logic Regarding Genetic Data?*, 4 Comput. L. & Sec. Rev. 317, 317–29 (2013).

[32]Article 29 Data Protection Working Party, *Working Document on Genetic Data*, (European Commission, Working Paper 203, 2004).

[33]Paul Quinn, Ann-Katrin Habbig, Eugenio Mantovani & Paul De Hert, *The Data Protection and Medical Device Frameworks? Obstacles to the Deployment of mHealth Across Europe?*, 20 Eur. J. of Health L. 185 (2013).

[34]Menno Mostert, Annelien L. Bredenoord, Monique C.I.H. Biessart & Johannes J. M. van Delden, *Big Data in Medical Research and EU Data Protection Law: Challenges to the Consent or Anonymise Approach*, 24 Eur. J. Hum. Genetics 956 (2016).

https://doi.org/10.1017/glj.2021.79 Published online by Cambridge University Press

The primary manner in which Directive 95/46/EC imposed extra requirements upon those who wished to process sensitive data was to create additional barriers before such processing could be permitted. This was largely achieved through the creation of specific legal bases that were applicable if sensitive data were to be processed. In general, these legal bases were more restrictive and less available—for example, only capable of applying within narrowly defined circumstances—than the legal bases that were available for those who wish to process non sensitive data.[35] The most prominent of these legal bases was that of "explicit consent," which as section D outlines, foresaw a more onerous form of consent than that which was available as a legal basis for the processing of non-sensitive forms of data. This would often entail a considerable extra burden given that in many cases, including, for example, the use of health data, where explicit consent was often understood in national law as entailing more formality, for instance, involving written, signed consent forms.[36] This requirement acted to serve an important barrier function, that effectively made it more difficult to process sensitive data.[37] Other possibilities for the processing of sensitive data existed for: Carrying out obligations or rights in the field of employment law, where such processing is in the vital interests of data subjects; when in it is in the public interests; when data are processed by non-profit entities with a political philosophical, religious, or trade-union aim; or where such data are manifestly made public by the data subject or are necessary for legal claims.[38] Each of these could only be utilized in a limited range of contexts and was subject to a range of conditions, for example, often requiring the existence of a specific legislative framework in Member State legal systems.[39]

2. Post GDPR

The GDPR contains several important adaptions and innovations concerning sensitive data. Most obviously, it has within Article 9(1) both confirmed a closed list of "special categories of personal data" and enlarged the scope of protection adding three more special categories of personal data: "[G]enetic data," "biometric data for the purpose of uniquely identifying a natural person," and "data concerning a natural person's ... sexual orientation."[40] The use of a regulation has to a certain extent ended the question of which categories of data should be sensitive in nature. Member States can no longer create further categories of sensitive data, as they could under Directive 95/46/EC, though the GDPR now includes the most common ones that Member States had themselves added, such as genetic data.[41] This has, to a certain degree, reduced the problems that were created by the heterogeneous legal landscape that was permitted under Directive 95/46/EC. This general harmonization has, however, been limited with regards to "genetic data, biometric data or data concerning health."[42] For these types of data Member States are permitted to "maintain or introduce further conditions, including limitations" concerning processing. Whilst this exception does not cover all of the forms of sensitive data described within Article 9, it does nonetheless apply to important categories that are central to many critical areas including security, healthcare, and scientific research.[43] In such areas Member States will be able to maintain and augment often complex webs

[35]Council Directive 95/46, supra note 31, at art. 8 (outlining these legal bases).

[36]Eugenio Mantovani & Paul Quinn, *mHealth and Data Protection—The Letter and the Spirit of Consent Legal Requirements*, 28 INT'L REV. L. COMPUT. & TECH. 222 (2013).

[37]*Id.*

[38]*See* Council Directive 95/46, supra note 31, at art. 8.

[39]Mahsa Shabani & Pascal Borry, *Rules for Processing Genetic Data for Research Purposes in View of the New EU General Data Protection Regulation*, 26 EUR. J. GENETICS 149 (2018).

[40]The original list of special categories of personal data in the GDPR Commission Proposal just added "genetic data" to the list of sensitive data already recognized at Article 8 DPD.

[41]Renate Gertz, *Is It 'Me' or 'We'? Genetic Relations and the Meaning of 'Personal Data' under the Data Protection Directive*, 11 EUR. J. HEALTH L. 231 (2004).

[42]GDPR, *supra* note 1, at art. 9(4).

[43]Paul Quinn & Liam Quinn, *Big Genetic Data and Its Big Data Protection Challenges*, 34 COMPUT. L. & SEC. REV. 1000 (2018).

https://doi.org/10.1017/glj.2021.79 Published online by Cambridge University Press

of national legislation on the processing of such data, maintaining problems in terms of heterogeneity for those who wish to use such data on a pan-European basis.

The GDPR has also amended the legal bases for processing special categories of data, by adding grounds relating to purposes of "substantial public interest," "preventive or occupational medicine," "public health," and for "research [or] ... archiving" purposes. It has also rephrased some of the other legal grounds for sensitive data processing.[44] These alterations represent *inter alia* a clarification of what was largely understood as being possible under the more limited description of legal bases under Directive 95/46/EC, as reflected in case law or opinions of the Article 29 Working Party. In addition, a further novelty is to restrain automated decision-making using sensitive data only where explicit consent has been obtained.[45]

Perhaps more importantly, the GDPR has added further administrative requirements that will often apply even after the "barrier protection," in other words, the existence of more restrictive legal bases. These requirements include, where the processing of special categories of data is "on a large scale," that data controller shall be obliged to appoint a Data Protection Office[46] and to perform a Data Protection Impact Assessment.[47] These requirements, which are further discussed in section F, represent an important evolution in terms of what it means to be the controller or processer of personal data. In particular, they have arguably acted to expand the divide separating sensitive and non-sensitive personal data, especially in terms of the potential burdens associated with processing the former. In sum, in the past, sensitive data was a type of data for which there was an added barrier, once that barrier had been overcome, processing such data attracted no further burdens than non-sensitive data. With the advent of the GDPR, however, this has changed, with creation of extra requirements that will apply long after this barrier has been overcome. These changes are important given the changing nature of sensitive data, discussed in section D, and must be taken into account when discussing the value of the concept and its continued relevance.

C. How to Define Sensitive Data?

I. Context and Purpose Based Definitions

As described above, the GDPR has introduced a number of new requirements that are likely to apply to the processing of sensitive data in many instances. In addition to the regulation of such data, however, the very nature of such data is itself in a state of full evolution. This change is taking place in a world where concepts such as the Internet of Things (IoT) and "Big Data" have become common.[48] These phenomena entail the continuous creation of enormous amounts of personal data.[49] Taken with never ending increases in computing power and the increasing ease of sharing

[44]Article 9(2) of the GDPR foresees a number of legal bases for processing sensitive data, including: (g); "processing is necessary for reasons of substantial public interest, on the basis of Union or Member State law which shall be proportionate to the aim pursued, respect the essence of the right to data protection and provide for suitable and specific measures to safeguard the fundamental rights and the interests of the data subject"; (h), "processing is necessary for the purposes of preventive or occupational medicine, for the assessment of the working capacity of the employee, medical diagnosis, the provision of health or social care or treatment or the management of health or social care systems and services on the basis of Union or Member State law or pursuant to contract with a health professional and subject to the conditions and safeguards referred to in paragraph 3"; (j) "processing is necessary for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes in accordance with Article 89(1) based on Union or Member State law which shall be proportionate to the aim pursued, respect the essence of the right to data protection and provide for suitable and specific measures to safeguard the fundamental rights and the interests of the data subject."

[45]Such processing is also subject to additional special safeguards adopted by the data controller. GDPR, *supra* note 1, at art. 22(4).

[46]*Id.* at art. 37(1)(c).

[47]*Id.* at art. 35(3)(b).

[48]These concepts are dicussed in detail in Section D.

[49]Paul Quinn, *The Anonymisation of Research Data — A Pyric Victory for Privacy That Should Not Be Pushed Too Hard by the EU Data Protection Framework?*, 24 EUR. J. HEALTH L. 1 (2017).

https://doi.org/10.1017/glj.2021.79 Published online by Cambridge University Press

and combining disparate datasets, more and more data is arguably becoming of a sensitive nature. How much is dependent upon the definition of sensitive data that is used. In particular, the potential scope of sensitive data can vary greatly depending on whether a context or purpose-based definition is used.

The contextual approach—originally adopted, for example, in Germany and Austria—[50] views the question of whether personal data is sensitive or not in primarily objective terms. Any personal information can, depending on the circumstances of the processing, be "sensitive." Accordingly, as Simitis argued two decades ago, all personal data should be assessed against the background of the *context* that determines their processing, as determined by several factors. In other words, the specific interests of the controller, as well as of the potential recipients of the data, the aims for which the data are collected, the conditions of the processing and its possible consequences for the persons involved. All these elements might help determine the sensitivity of personal data processing. Whilst a contextual understanding of what sensitive data is may well take into account the—subjective—intentions of any data controller, it is likely to go beyond that and consider a range of other—objective—factors also.

The contextual approach can be contrasted with the purposeful approach, which primarily focuses on the intentions of the data controller. It essentially looks at the intention of the data controller and asks whether the controller intends to draw conclusions from the processing of particular data that could be regarded as being sensitive in nature.[51] These intentions are in general determinative in deciding whether the data that is being used is sensitive or not. Where the controller in question has no intention of creating or using data in a way that could be considered sensitive, a purpose-based definition would find that no sensitive data is involved. By contrast, a context-based approach is less concerned with the intentions of the controller and more preoccupied with the objective nature of the data itself. A context-based approach would suggest that sensitive data is being processed where, given the overall context, it would be possible to draw a conclusion from the data that might be sensitive in nature.

In determining what this context is, there may be a number of important factors to take into account. First, it is necessary to consider what other data may be available to a data controller. This is important because the combination of various datasets may increase the likelihood that conclusions of a sensitive nature can be reached, even where this may not be apparent when looking at particular datasets in isolation.[52] In the increasingly interconnected online world, this may entail taking into account not only other data that may be physically in a controller's possession but also data that it may have access to elsewhere, such as data that may be freely available online. A second factor is the technical abilities of the data controller, or other potential data controllers. This will include the computing or analytical power or the technical know-how available to data controllers. Given that such factors are in a constant state of evolution, and that access to potentially complimentary datasets is ever increasing, the particular context of an instance of data processing is always changing. Given this—as section D discusses in more detail—data processing that may not have been considered sensitive in the past, may well be considered sensitive in the future.

The contextual and purpose-based approaches can be conceptualized as being distinct and in contrast to each other, but the reality is that elements of one approach may be infused with that of another. For example, a legal text may in general be understood as employing a contextual approach with certain elements that may also be purpose-based. Elements of one approach can be blended with another to either moderate the effect of one approach or ensure that it does

[50] *See* Bundesdatenschutzgesetz [Federal Data Protection Act], Dec. 20, 1990, BGBL. I S. 2954, §28, 35 (Ger.)(demonstrating that sensitive data is not an a priori category of personal data, but for each legal circumstance—such as legal basis, right to rectification—special rules are provided for sensitive scenarios: Healthcare data processing, rectifying political or religious beliefs data).

[51] Link to article 29 paper on health data and other papers.

[52] *See* Gianclaudio Malgieri & Giovanni Comandé, *Sensitive-by-Distance: Quasi-Health Data in the Alogrithmic Era*, 3 INFO., COMMC'N & TECH. L. 1 (2017). *See also* Bart Custers & Helena Ursic, *Big Data and Data Reuse: A Taxonomy of Data Reuse for Balancing Big Data Benefits and Personal Data Protection*, INT'L DATA PRIV. L. 1, 8 (2016).

https://doi.org/10.1017/glj.2021.79 Published online by Cambridge University Press

not miss certain instances of processing that should clearly be considered as sensitive in character. In terms of the former, one could imagine the use of a context-based approach that would also require elements of purpose or intention. For example it might be required that, for data to be considered of a sensitive nature, a potential data controller would have to be aware that there was a potential that the data under its control could be processed in a way to produce conclusions that would be of a sensitive nature, even if it were not intending to carry out such processing.[53] Such a compromise would act to reduce the effect of a purely contextual approach and to introduce an element of intention without restricting it as much as a purely intentional approach would do. Such an understanding could be used to moderate the potentially explosive growth in the volume of sensitive data that may occur in the future—discussed in section G—if a primarily context-based understanding of sensitive will be maintained.

In other situations, however, one approach can be added to another to ensure that important gaps are not created by exclusively depending on one approach alone. Imagine for instance a controller that collected large amounts of personal data concerning certain behavioral characteristics in the vague hope that some form of innovative data processing may be available in the future that would permit conclusions to be drawn concerning the health status of data subjects. Using a strictly contextual definition of sensitive data might mean such processing could not be considered as the processing of sensitive data, given that it might not be possible at present to draw such conclusions, because the required technological or analytic processes might not yet be available, as an example. Such a manner of defining personal data may be insufficient given the likelihood that future technological evolutions will render such data sensitive in nature—even though it might not be at present. Alternatively, new forms of potentially complimentary data might become available—allowing potentially sensitive conclusions to be drawn. Given this, a definition that ignores purpose may not always be suitable given that a data controller could assemble data with the hope that unknown future evolutions will allow sensitive conclusions to be drawn—that cannot at present.

In such contexts, the addition of an element of purpose to an otherwise context-based definition may serve to widen the scope of sensitive data in a way that would protect against a number of likely risks in terms of the harms that were discussed in section B, such as discrimination and related harms. Such a situation could be contrasted with a solely purpose-based definition where the intentions of the data controller were the central factor in deciding upon the sensitivity of the data. Whilst such an approach would undoubtedly prevent an overly extensive coverage of sensitive data it would arguably increase the risks of ill thought through or negligent processing of personal data. This could include cases where the controller itself had no intention to derive sensitive conclusions from personal data but nonetheless processed it in such a way that would create risks *vis-à-vis* third parties that might have access to the data and might be able to draw such conclusions.[54]

1. The GDPR: Still Overwhelmingly Contextual in Approach

Predominantly, context-based approaches have historically enjoyed precedence over purpose-based approaches. This was, for example, the case with the EU Data Protection Directive—95/46/EC—which employed prima facie a heavily context-based approach. In its general prohibition on the processing of sensitive personal data, it states:

[53]See the reasonableness test in Vaclav Janecek & Gianclaudio Malgieri, *Data Extra Commercium, in* DATA AS COUNTER-PERFORMANCE—CONTRACT LAW 2.0? 14-15 (Sebastian Lohsse, Reiner Schulze, & Dirk Staudenmayer eds., 2019) ("[a] party is responsible ... if and when, given the current state of knowledge and technology that is normally available to a person in a similar position, it is reasonably likely that the relevant data reveal personal [sensitive] information.").

[54]On the dynamic nature of personal data and sensitive data, see Vaclav Janececk & Gianclaudio Malgieri, *Commercialization of Data and the Dynamically Limited Alienability Rule*, 21 GERMAN L. J. 924 (2020).

https://doi.org/10.1017/glj.2021.79 Published online by Cambridge University Press

> Member States shall prohibit the processing of personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, trade-union membership, and the processing of data concerning health or sex life.

This formulation appears to leave little room for the role of purpose or intent, appearing totally focused on an objective, context-based definition of what could constitute personal data.[55] Problems concerning an approach to sensitive data that is too context-based have been raised over the years by various institutions and scholars. As Poullet and Dinant discussed in a report for the Council of Europe,[56] the overtly context-based definition—of sensitive data—within the directive risked missing some attempts by controllers to arrive at conclusions that might be sensitive in nature. Notably, the authors stated:

> [T]he extremely broad definition of sensitive data . . . makes it absolutely necessary to abandon the approach based on a definition of the actual nature of data in favor of a purpose-based approach . . . This approach would make it possible to consider the actual processing of data as sensitive rather that the data itself, even if no sensitive data were involved.[57]

As McCullagh pointed out, the OECD seemingly opted for such an approach in its 1980 guidelines.[58] Furthermore, the Explanatory Report of the Council of Europe's Modernized Convention 108, for example, argues that some specific types of data processing may entail a particular risk for data subjects independently of the context of the processing and mentions as emblematic examples genetic data and data related to criminal proceedings.[59] According to the report, what matters is the purpose of the processing, not the context, because "processing of sensitive data has the *potential* to adversely affect data subjects' rights when it is processed for specific information it reveals."[60]

Interestingly, the GDPR, unlike its predecessor Directive 95/46/EC, seems to take into account the idea that a purely context-based definition alone may not always be sufficient in terms of the recognition of what sensitive data is. Its definition seems to see a role for both a context-based and a purposeful understanding of what sensitive data is. Strangely, however, the GDPR rather than employing a consistent approach for all forms of personal data, appears to call for a contextual understanding of most categories of personal data and a purposeful understanding of only a few categories of data. It states in Article 9:

> Processing of personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership, and the processing of genetic data, biometric data for the purpose of uniquely identifying a natural person, data concerning health or data concerning a natural person's sex life or sexual orientation shall be prohibited.[61]

[55]The Article 29 working Party Opinion on Sensitive Data is notable for not even discussing this point. Given this there is little evidence to suggest that Article 8 should be read in a way that differs from its literal meaning which is seemingly very context-based. *See* Article 29 Data Protection Working Party, *Advice Paper on Special Categories of Data ("sensitive data)*, (European Commission, Working Paper 444105, 2011).

[56]*Report on the Application of Data Protection Principles to the Worldwide Telecommunications Networks—Information Self-Determination in the Internet Era, Thoughts on Convention No. 108 for the Purposes of the Future Work of the Consultative Committee,* COUNCIL OF EUR. (2004), https://rm.coe.int/168068416a.

[57]*Id.* at 43.

[58]McCullagh, *supra* note 23.

[59]*Explanatory Report to the Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data*, COUNCIL OF EUR., §57 (1981), https://rm.coe.int/16800ca434.

[60]*Id.* at § 60.

[61]GDPR, *supra* note 1, at art. 9

https://doi.org/10.1017/glj.2021.79 Published online by Cambridge University Press

A simple reading of this definition of personal data would seem to imply that the GDPR proposes a context-based definition for all types of data except biometric data for which it proposes a purpose-based definition. For all of the other categories of sensitive data invoked in Article 9, the GDPR uses the terms "data revealing" or "data concerning." This seems to denote a largely context-based definition and sees little role for the intention of potential data controllers to play. Biometric data therefore forms an exception in terms of the vision of sensitive data under the GDPR.

The purpose of this exception appears to be to exclude data such as photographs and visual images from being automatically considered as sensitive data.[62] This is because if this were not the case—and a purely contextual definition of sensitive data were to be applied— collections of photographs of identifiable individuals could potentially be considered biometric data even where a data controller had no intention of applying biometric processes to them.[63] This would in many instances mean that the possession of photographs and related images of people could be considered biometric and thus sensitive in nature even where there was no intention to use them in such a way. In place of this, the insertion of the words "for the purpose of" with regards to biometric data means that biometric data will only be sensitive where there is a clear intention to apply biometric processes to images or other material that can potentially be used to identify individuals.

This can notably be contrasted with health data, which the GDPR seemingly indicates will continue to be identified according to a primarily context-based definition. As section D.II discusses, the role of intent seems less important with regards to health data. This means that the collection of apparently innocuous data that may not intuitively relate to health can in certain context be considered as health data even where there is no intention to process data in such a way. As section D also discusses, evolutions in computing power, the availability of potentially complimentary big data and new analytic processes means that the likelihood of such events is only going to increase in the future.

II. Likely Problems with a Purpose Based Definition

1. Difficulties in Objectively Demonstrating Purpose

Whilst the foregoing section may have given the reader the impression that the use of a context-based definition—of sensitive data—is likely to be problematic going forward, it is at the same time important to realize that any alternative purpose-based definition would present its own—and possibly greater —problems. Perhaps the most important of these is the difficulty of proving what exactly the intentions of the data controller are.[64] Whilst a data controller may claim a certain purpose, it may, in reality, be difficult to demonstrate that what was claimed was indeed the true purpose behind the data processing operation intended by the controller. Relying on declared purposes is likely, in many instances, to leave the door open to abuses in allowing situations where the data controller can simply declare that it has no intention to process sensitive data, even where the opposite may be the case. In order to avoid this, a purposeful definition of sensitive data would almost certainly have to contain some contextual or objective elements. This would likely involve a requirement to analyze the context in question, the background of the controller, and possibly the potential for commercial or other gain in order to discern whether the declared intention of the data controller appears to be objectively verifiable. The need for such a requirement arguably makes a purely subjective understanding of purpose infeasible and means that a number of the problems associated with the changing *de facto* nature of sensitive data would still apply to any conceivable use of a purpose-based definition. In other words, the ease with which sensitive conclusions can be drawn from personal data going forward. This would mean that some of the

[62]Catherine Jasserand, *Legal Nature of Biometric Data: From 'Generic' Personal Data to Sensitive Data*, 2 EUR. DATA PROT. L. REV. 297 (2016).

[63]GDPR, *supra* note 1, at Recital 35.

[64]Nikolaus Forgó, Stefanie Hänold & Benjamin Schütze, *The Principle of Purpose Limitation and Big Data*, *in* NEW TECHNOLOGY, BIG DATA AND THE LAW, 17–42 (Marcelo Corrales, Mark Fenwick, & Nikolaus Forgó eds., 2017), https://doi.org/10.1007/978-981-10-5038-1_2.

https://doi.org/10.1017/glj.2021.79 Published online by Cambridge University Press

burdens associated with the current use of a largely contextual understanding of what sensitive data is would remain even if a purposeful conception was to be adopted in its place.

Another problem that could occur, even where a data controller was being entirely honest in its intentions not to derive sensitive conclusions, is the negligent processing of data that was *de facto* sensitive in nature— at least under a context-based understanding. This could occur where a data controller processed, retained, or made data available in an ill-considered manner where it was feasible to draw sensitive conclusions, even where this was not the intention of the data controller. In such contexts, it is important to remember that just because the data controller in question does not have any intentions to derive sensitive conclusions, this may not be the case concerning third parties who may have access to the data that is produced. Given this, the ill-considered or negligent processing of what is, in reality, sensitive data, could have malign consequences that were not envisaged by the data controller.

3. The Dynamic Nature of Purposes Under the GDPR

Another problem in addressing sensitive data under a purpose-based approach is the intrinsically dynamic of the concept of purpose. Specifically, such a concept may change from one moment to the next.[65] Whilst in general, the GDPR states that personal data can be collected only for "specified, explicit and legitimate purposes and not further processed in a manner that is incompatible with those purposes"[66] there are important exceptions to this requirement. For example, further processing for archiving purposes, in the public interest, for scientific or historical research purposes or statistical purposes cannot be considered incompatible with the initial purpose, if such processing respect specific safeguards as prescribed at Article 89(1), such as pseudonymization, data storage limitation, specific transparency measures.[67] Recital 33 further clarifies that it is often not possible to "fully identify the purpose of personal data processing for scientific research purposes at the time of data collection. Therefore, data subjects should be allowed to give their consent to certain areas of scientific research ... to the extent allowed by the intended purpose."[68] Accordingly, in cases of data processing for research purposes, it is not always necessary to fully determine and explicitly describe the exact purpose in question.

In addition, Article 6(4) allows processing for a purpose other than that for which the personal data have been collected, if the processing is not based on consent or Union or Member States law, if the data controller successfully performs an assessment of the new purpose under certain criteria— link between the initial and the new purposes, context, possible consequences, existing safeguards.[69]

There are therefore a number of instances where it is possible for the purpose to be modified during the data processing: In case of research or statistics and in any case in which the data processing is not based on consent or a Union or Member State law and the data controller can prove that the new purpose is compatible with the initial one. Both are potentially of very broad application.

This presents a further problem for using a purpose-based definition of sensitive data: The definition of sensitive data on the purposes of data processing in a number of contexts be misleading and inaccurate. In some cases, the purposes might not be fully identified at the beginning and, in many cases, new purposes may evolve during data processing. This means that if the definition of sensitive data were to be fully purpose-based, the "nature" of personal data, such as whether they are sensitive or not, could change as soon as new purposes are determined. Where this happens, a new legal base would need to be found, for instance, within Article 9(2). This would not only make planning on the part of data controllers difficult but would

[65] *Id.*
[66] GDPR, *supra* note 1, at art. 5(1)(b).
[67] *See id.*
[68] *Id.* at Recital 33
[69] *See* Article 29 Data Protection Working Party, *Opinion 03/2013 on Purpose Limitation*, (Working Paper 203, 2013).

https://doi.org/10.1017/glj.2021.79 Published online by Cambridge University Press

also make it difficult for data subjects to understand the real, and future, implications of their data being processed.[70]

D. The Changing *de facto* Nature of Sensitive Data

I. More Data is Likely to Mean More Sensitive Data

Whilst there may be logical reasons for giving the definition of biometric data a purposebased perspective, the question could be raised as to why this did not occur with other forms of sensitive data, some of which are likely to face similar issues to biometric data, such as the likely unintentional collection and processing of sensitive data that would occur where a solely context-based definition is used. This includes not only health but also other aspects such as political opinion and sexual orientation. The changing nature of personal data means that, in reality, it may not be very difficult for a data controller to use datasets within its possession to arrive at conclusions that may be sensitive in nature. Although the decision to make biometric data an exception in this regard is understandable, there is no clear reason why such an exception has been made only for biometric data because it is likely that conclusions that might be of sensitive nature—in terms of other categories of sensitive data—may be drawn from potentially any large dataset.

Another puzzling issue is that no effort appears to have been made to combine purposeful and context-based elements in the definition of sensitive data despite the advantages that this could produce.[71] Rather, all forms of sensitive data are defined contextually, with the exception of biometric data which is defined exclusively in purposeful terms. This does not permit the moderating effect that such a combination would allow,[72] whereby the extreme effects of a more polarized definition—tending exclusively towards a purposeful or context-based perspective— are reduced. Accordingly, and without the seeming existence of a purposeful element in their respective definitions, it appears that most types of sensitive data—data that are not biometric in nature—will be defined in an extremely broad fashion, encompassing a potentially enormous amount of personal data.

One potential issue arises as a result of the ever-increasing amount of data that is likely to fall within the net of sensitive data. Changes in terms of computing power and online interconnectivity mean that more and more data samples are likely to fall within the definition of personal data, in other words, that it can be linked directly or indirectly to particular individuals.[73] Indeed, the guidance by the Article 29 Working Party about the boundary between personal and non-personal data—before the GDPR came into force—indicated clearly that the types of data that can be considered "personal" may go far beyond that which is immediately intuitive.[74] In invoking the concept of "reasonably likely" invoked both in Directive 95/46/EC and the GDPR, the working party emphasized that it may often be difficult to state that various data sets are anonymous.[75] This is due largely because of increasing computational capacity and advancement in data mining technologies. This allows more forms of analysis that can potentially identify individuals from data that may intuitively seem to be non-personal in nature.[76]

An analogous argument can be made for the case of sensitive data. That is because the same factors—data mining technologies, availability of data—mean that it is becoming more and more likely that data that might not intuitively appear to be sensitive data is indeed sensitive data. The increasing ability to combine various datasets that may not be sensitive data and perform complex

[70]On the dynamic nature of personal data and sensitive data, see Janececk & Malgieri, *supra* note 56.

[71]As dicussed in section C.III.

[72]This is dicussed in section H.

[73]GDPR, *supra* note 1, at art. 4.

[74]*See* Article 29 Data Protection Working Party, *Opinion 04/2007 on the Concept of Personal Data* 6-12 (Working Paper 136, 2013).

[75]*Id.* at 15–17

[76]Quinn, *supra* note 51.

https://doi.org/10.1017/glj.2021.79 Published online by Cambridge University Press

analysis on them may mean that, together, such data can be considered sensitive in nature.[77] Looking at datasets in isolation to discern whether they are of a sensitive nature is becoming increasingly unacceptable. The continued application of a strictly contextual understanding of what most forms of sensitive data are will mean data controllers will have to consider access not only to data that is in their own direct possession; but, because of a world of online connectivity, access to potentially large amounts of complimentary data. The availability of such data, together with the availability of ever more powerful analytical tools means that inferences or conclusions can be drawn—often even from intuitively innocuous data—that can in fact make them of a sensitive nature.

This situation is by no means a static one. Each of the elements described here are in a constant state of evolution. This is only likely to intensify the problems that exist in determining whether data is of a sensitive nature or not.[78] The amount of data that is publicly available is only likely to increase by orders of magnitude in the future. Developments such as IoT mean that an enormous amount of data on all matter of things is continuously being created. This increases the possibility that, through a combination of various datasets, data controllers may be able to arrive at sensitive conclusions, conclusions that would not be possible with individual datasets. This must be considered in addition to the enormous explosion of all forms of data that are available online on a shared basis. This ranges from social media postings to databases of family histories to the results of scientific research on an enormous range of issues.[79] The pace of the creation of such data is ever increasing, augmenting in turn the possibility that various innocuous sets of data can be combined to form inferences that would amount to sensitive data.[80] A problem that is now much more complex than it was a number of years ago—in other words, trying to determine when exactly personal data is sensitive data—will become more prominent in the coming years. Computing power is only going to increase further. With this comes the possibility to create ever more powerful algorithms that are able to deduce relationships between various sets of data that may not have been possible before. These developments all mean that it will become increasingly likely that controllers of a particular dataset, will be able to make inferences that themselves represent sensitive data, through comparison and analysis with data available elsewhere. Such developments will provide an important justification to question the use of a strictly contextual understanding of what most forms of sensitive data are.

II. An Emblematic Example: Medical Information v. Health Status

Health data is one of the most problematic examples of "special categories of personal data." It is defined at Article 4(15) as "personal data related to the physical or mental health of a natural person, including the provision of health care services, which reveal information about his or her health status."[81] This has the potential to include a potentially enormous range of data, depending on how much and how accurate the degree of information revealed about an individual's needs to be. As one of the authors of this work has previously stated, determining the "degree

[77]Gianclaudio Malgieri & Giovanni Comandé, *Sensitive-by-Distance: Quasi-Health Data in the Algorithmic Era*, 26 INFO. & COMMC'N TECH. L. 229, 229–49 (2017).

[78]*Id.*

[79]Quinn & Quinn, *supra* note 45. *See generally* Tene & Polonetzky, *supra* note 3.

[80]Giovanni Comandè & Giulia Schneider, *Regulatory Challenges of Data Mining Practices: The Case of the Never-Ending Lifecycles of "Health Data"*, 25 EUR. J. OF HEALTH L., 284, 284–307 (2018).

[81]*See also* GDPR, *supra* note 1, at Recital 35 ("[p]ersonal data concerning health should include all data pertaining to the health status of a data subject which reveal information relating to the past, current or future physical or mental health status of the data subject. This includes information about the natural person collected in the course of the registration for, or the provision of, health care services as referred to in Directive 2011/24/EU of the European Parliament and of the Council (9) to that natural person; a number, symbol or particular assigned to a natural person to uniquely identify the natural person for health purposes; information derived from the testing or examination of a body part or bodily substance, including from genetic data and biological samples; and any information on, for example, a disease, disability, disease risk, medical history, clinical treatment or the physiological or biomedical state of the data subject independent of its source, for example from a physician or other health professional, a hospital, a medical device or an in vitro diagnostic test.").

https://doi.org/10.1017/glj.2021.79 Published online by Cambridge University Press

of *revelation*" can be a complex affair.[82] Trying to set a clear test for the definition of health data, the Article 29 Working Party summarized three cases in which personal data should be considered health data:

1. The data are inherently/clearly medical data;
2. The data are raw sensor data that can be used in itself or in combination with other data to draw a conclusion about the actual health status or health risk of a person;

Conclusions are drawn about a person's health status or health risk, (irrespective of whether these conclusions are accurate or inaccurate, legitimate or illegitimate, or otherwise adequate or inadequate.)

Whilst these concepts may seem clear they are not. The second example in particular can be used to illustrate the problems that a purely context-based definition of sensitive data could produce. In the world of IoT and big data, an extremely large range of data could be thought to fall under the concept of "sensor data.'" It will become more and more feasible for disparate sources of data to be combined in order to allow conclusions to be drawn about individuals and their health status. Going forward, the increasing penetration of IoT, including in domains such as eHealth, will mean that more and more data will be considered "raw sensor data." This will occur in a big data world where the capacity to find a particular piece of information using other data will be greatly expanded.[83] At the same time, drawing conclusions from raw data is a common trend for many digital service providers.[84] This is also particularly true in the field of biomedical data processing.[85] As one of the authors of this Article identified in a previous article there are two important elements that can be used to determine the likelihood of any particular data being sensitive in nature:

1. The "*intrinsic sensitiveness,*" of a particular data set;
2. The *computational distance*: In other words, the effort that would be needed to draw sensitive conclusions from various data that might not be prima facie intrinsically sensitive.[86]

The intrinsic sensitiveness of data concerns the content of that information: Data about health status—such as diagnosis, blood pressure, and blood readings—are inherently "sensitive," whereas data about, for instance, food consumption, daily exercise, air pollution of one's city —are not intrinsically "sensitive," but might be considered sensitive because the computational distance between such data and an inference revealing a sensitive aspect, such as health status, is relatively small. Even though it is a static parameter, intrinsic sensitiveness is not binary; instead, it is a spectrum of different shades. Some data might concern an explicit health status description, for instance, a diagnosis; while others might be biometric data unrelated to health conditions, such as height; or be strongly related to health even if not directly describing a health status, for example, sleep-wake schedules and appointments on a calendar. Further, others might be unrelated to health at all, such as quantity of time spent on social networks; or—as in a recent case in Italy—being the beneficiary of a periodic payment from the State in accordance with law regulating reimbursement for victims or parents of victims of injuries caused by vaccination.[87]

[82]Malgieri & Comandé, *supra* note 79. On how "reveal" should be interpreted, see Bryce Goodman, *A Step Towards Accountable Algorithms?: Algorithmic Discrimination and the European Union General Data Protection*, 29th Conference on Neural Information Processing Systems 2–3 (2016).

[83]*See* Tene & Polonetsky, *supra* note 3.

[84]*Id.*

[85]Effy Vayena & Urs Gasser, *Strictly Biomedical? Sketching the Ethics of the Big Data Ecosystem in* Biomedicine in The Ethics of Biomedical Big Data 20–24 (Brendt D. Mittelstadt & Luciano Floridi eds., 2016).

[86]Malgieri & Comandé, *supra* note 79. On how "reveal" should be interpreted, see Goodman, supra note 84, at 2–3.

[87]*See* Cass., 3 Febbraio 2017, n. 3456 (It.).

https://doi.org/10.1017/glj.2021.79 Published online by Cambridge University Press

Conversely, computational distance concerns the level of scientific, economic, and technological effort required when combined with other—personal or non-personal data—to infer sensitive data from apparently non-sensitive information. For example, the computational distance between daily diet and health status is small, while the computational distance between hours spent on social media and future health conditions is larger. Continuing developments in the *de facto* nature of both personal data and sensitive data mean that the computational distance required to arrive at sensitive conclusions is being dramatically reduced. This is becoming particularly apparent with regards to health data. Enormous reservoirs of personal data are being created in diverse domains such as social media or the monitoring of IoT in the form of wearables for fitness purposes. Such data can be in theory combined with a range of other data in order to form conclusions that could be considered sensitive in nature, such as inferences about health status. The online and interconnected world means that such data is more likely to be available to potential data controllers. Potentially complimentary data does not only include additional datasets that are in the possession of a data controller but a range of complimentary available data from other sources including social media, genealogical databases, data concerning movement, and potentially electronic health records. Computational distance is further reduced by ever-increasing computer power and the development of novel analytic techniques, including *inter alia* forms of deep learning that allow the potential for a range of sensitive conclusions concerning health status to be drawn even where this may not be the intention at the outset.[88]

The likely increase in the quantity of sensitive data that it will entail means that it is important to consider whether approaches such as the GDPR will be in a position to provide for adequate and meaningful regulation of the use of such data, especially where a contextbased form of definition continues to be used. Given the constant evolution of the nature of data in general, it is arguably important to constantly enquire as to the ability of data protection frameworks such as the GDPR to prevent the types of harm that sensitive data have been traditionally associated with, for example, harms associated with discriminatory contexts. Section F will look at the main features of the GDPR that are intended to apply when sensitive data is processed and discuss what their value is likely to be in a world where more and more data is likely to become sensitive in nature.

E. Consent as a Weakened Barrier to the Use of Sensitive Data?

I. Two Forms of Consent for Two Types of Personal Data

Informed consent is perhaps the best-known legal basis for the processing of personal data. The notion of informed consent, however, predates the concept of data protection itself. The intuitive idea of consent is unproblematic: It is an important ethical principle,[89] familiar to all. This instinctual understanding is demonstrated through any number of daily deals, deeds, or transactions. In some more formal contexts, it is common practice to engrave informed consent in documents and legal requirements, for instance, in marriage. Regardless of the form, any act of giving, refusing, or withdrawing consent, can be represented as an act that is freely embarked upon. Discourses on consent have flourished in the last four decades; many fields, from finance to education, but with particular intensity in the medical field, have grown accustomed to requiring formal forms of consent, employing procedural ways of seeking, giving, recording, and respecting informed consent.[90] Traditionally the processing of personal health data, for example, was seen as something that

[88]Jane Kaye, *The Tension between Data Sharing and the Protection of Privacy in Genomics Research*, 13 Ann. Rev. Genomics and Hum. Genetics 415 (2012).

[89]*See* Onora O'Neill, Autonomy & Tr. in Bioethics, (2002).

[90]*See generally* Mantovani & Quinn, *supra* note 38; Amy L. McGuire & Laura M. Beskow, *Informed Consent in Genomics and Gentic Research*, 11 Annual Rev. of Genomics and Hum. Genetics (2010); Paula Boddington, Liam Curren, Jane Kaye, Nadja Kanellopoulou, Karen Melham, Heather Gowans, Naomi Hawkins, *Consent Forms in Genomics: The Difference between Law and Practice*, 18 The Eur. J. of Health L. 491 (2011).

https://doi.org/10.1017/glj.2021.79 Published online by Cambridge University Press

posed a high degree of risk for those involved, including a risk of discrimination and related phenomena.[91] For these reasons the use of consent—as a legal basis—for the processing of personal health data has often been more formalized than the type of consent required for the processing of data in other areas which may have regarded as posing a lower level risk.[92] Such a division has been clearly recognized in data protection since the creation of Directive 95/46/EC. This directive foresaw two different regimes relating to consent for personal data that was not sensitive in nature and personal data which was sensitive in nature. The first foresaw a looser set of requirements relating to the need for a: *"[F]reely given specific and informed indication of his wishes by which the data subject signifies his agreement to personal data relating to him being processed."*[93]

This could be compared to consent for the processing of sensitive data which was deemed to require "explicit consent."[94] The former foresaw forms of passive consent where actions could be understood as implicitly representing consent.[95] These include entering an area with signs—explaining personal data could be processed—or continuing to visit a web site after being warned that there may be related processing of personal data. For sensitive data, however, such passive forms of consent were not permitted. Explicit consent was understood as requiring that consent be given in a discrete and separate act, distinguishable from other activities that might be taking place. Such consent equated to a definitive sign that the data subject had been properly informed and understood the consequences of consent.[96] Unlike "regular" consent—for non-sensitive data—there was no option of such consent being passive or simply implied from other actions. Explicit consent could only be affirmative in nature.[97] In addition, for explicit consent, there is a need to for some form of record to be kept of this affirmative action and its significance.

Again, health data can be used as an illustrative example of how more rigorous requirements for consent concerning the processing of health data were required. Such requirements traditionally manifested themselves as an informed consent form in healthcare contexts. For many years, the signed consent form was seen as the gold standard of informed explicit consent. National law in many States accordingly demanded that consent for the use of certain forms of sensitive data—in the medical treatment context, for instance,—required written consent.[98] Although this was not required by Directive 95/46/EC, the fact that it foresaw two different levels of consent and that it was a directive, leaving the choice of transposing law up to Member States, seemed to encourage or at least permit such a division.

II. A Barrier Function Weakened by the GDPR?

The more burdensome form of consent for sensitive data arguably allowed the creation of a "sensitive data barrier." Whilst consent is not the only legal basis for the processing of sensitive data, it is one of the most important, and for many data controllers that are not able to avail themselves of the other options—for example, actors outside the traditional health care or scientific research setting— it may be the only basis available.[99] The "consent barrier" therefore represented an

[91]Philip Hunter, *The Big Health Data Sale: As the Trade of Personal Health and Medical Data Expands, It Becomes Necessary to Improve Legal Frameworks for Protecting Patient Anonymity, Handling Consent and Ensuring the Quality of Data*, 8 EMBO Rep 1103 (2016).

[92]Article 8

[93]Council Directive 95/46, supra note 31, at art. 2(H).

[94]*Id.* at art. 8(2)A.

[95]Article 29 Data Protection Working Party, *Guidelines on Consent under Regulation 2016/679*, 30 (Working Paper 259 rev.01, 2013).

[96]Deryck Beyleveld, *An Overview of Directive 95/46/Ec in Relation to Medical Research*, *in* The Data Protection Directive and Medical Research Across Europe (2004).

[97]Article 29 Data Protection Working Party, *supra* note 97, at 12.

[98]Article 29 Data Protection Working Party, *Working Document on the Processing of Personal Data Relating to Health in Electronic Health Records* (Working Paper 131, 2007).

[99]For more on the potential legal bases that were available under Directive 95/46/EC, see Quinn, *supra* note 51.

https://doi.org/10.1017/glj.2021.79 Published online by Cambridge University Press

important extra hurdle that those who wanted to process sensitive data had to overcome. In the healthcare context, this often meant that those wishing to create and process medical data could not do so unless they had obtained a more formalized form of consent than would be required for the processing of non-sensitive personal data—,the details of which would often be spelt out by national law.[100] This barrier represented more than a simple piece of paper and entailed a number of important efforts that were not to be underestimated. Participants had to be identified and contacted. They had to be informed of what was to happen. Importantly, potential controllers also had to gather potential signees physically in a certain location so that they could sign the consent form. This might entail a physical appointment with patients where a health professional presented them with the necessary material and then kept the signed form for their records.[101] Where the intention was to process the health data of numerous individuals—such as in large healthcare or scientific research projects—such requirements could represent an enormous impediment that may, in many cases, make potential controllers or processors think twice about processing such data in the first place.[102] In many instances, the problems associated with obtaining explicit consent arguably disincentivized the unnecessary of processing of *inter alia* health data.

With the development of more complex ICT systems, an increased ability to share data—especially in the health sector with the development of phenomena such as eHealth and mHealth—the traditional barrier of written consent became to be seen as more of a problem.[103] In place of an accepted golden standard, the need for written consent came increasingly to be seen as a barrier to the development of novel forms of health care.[104] These novel forms were built on the digitization of health care and have in certain contexts allowed acts of medicine to be conducted outside the usual setting of a medical institution. This includes the ability to transfer medical records between distant institutions or to engage in the remote monitoring or treatment of patients in various contexts. In such changing contexts, together with the increasing need to make consent as granular as possible, the need for static forms of written consent became less evident—and even perceived as an impediment to progress. In order to facilitate technological developments in areas such as eHealth/mHealth, requirements in national law for consent—for the processing of health data—have been largely relaxed. This has allowed the increasing use of electronic and remote consent processes.

This process has arguably been facilitated by the GDPR which has blurred the line between consent for the processing of non-sensitive and sensitive data. This can be seen in two regards. First, the GDPR confirms that explicit consent—for sensitive data—need not be written.[105] This confirms the changes that have been taking place in Member State law in recent years, in particular concerning consent related to the use of health or medical data. This has been enforced by the choice of regulation as legislative instrument, bringing about a greater degree of harmonization across Europe.[106] Second, the requirements pertaining to consent for non-sensitive data have been bolstered. Passive, implied consent is no longer permitted.[107] Rather, data subjects must give an indication— an affirmatory act—signifying that they provide consent for the processing of their data. This dilution of the formerly hard distinction between the two consents means that in many contexts the real difference between the two comes down to the notion of formality. This is exemplified with digital forms of consent and particularly in the context of online forms for the use of health data. The changes discussed here arguably mean that there is little real difference in terms

[100]*See* GDPR, *supra* note 1, at art. 9(4).

[101]Amy McGuire & Laura Beskow, *Informed Consent in Genomics and Genetic Research*, Annual Rev. Genomics and Hum. Genetics (2010).

[102]Pam Carter, Graema T. Laurie & Mary Dixon-Woods, *The Social Licence for Research: Why care.data Ran into Trouble*, 41 J. Med. Ethics 404 (2015).

[103]James G. Anderson, *Social, Ethical and Legal Barriers to e-Health*, 76 Int'l J. Med. Informatics 480 (2006).

[104]Article 29 Data Protection Working Party, *Opinion 15/2011 on the Definition of Consent* (Working Paper 187, 2007).

[105]Article 29 Data Protection Working Party, *supra* note 97, at 18 (clarifying requirements of the EU GDPR)

[106]Exceptions to harmonization are, however, notably permitted for some forms of sensitive by GDPR art. 9(4).

[107]GDPR, *supra* note 1, at art. 7(1).

https://doi.org/10.1017/glj.2021.79 Published online by Cambridge University Press

of the ability of consent to act as a barrier function to the processing of sensitive data.[108] The remaining difference—relating to the notion of the consent being explicit—is that the consent for the processing sensitive data amounts to a discrete and clear acknowledgement that the action being undertaken is indeed an act of consent for the processing of sensitive data, together with the legal ramifications of such an act. Whilst requiring separate and clear affirmatory acts to signify consent may sound like an impediment of some significance, the reality is that in the modern electronic and online context, this difference is becoming less significant. Efforts to comply with such requirements can often be complied with through pop-up prompts—linked to terms and conditions—asking for confirmation that consent is indeed intended. In contexts such as eHealth, where granular consent is increasingly being encouraged[109] and consent has become a frequent exercise, there is a risk that such requirements become an exercise in "ticking the boxes," an inconsequential burden that differs little from the type of consent that would be needed for non-sensitive data.[110]

It is, however, important not to consider the "barrier function" in isolation, especially given that the GDPR includes a range of novel measures that pertain to the potential processing of sensitive data. These are discussed further in the sections below.

F. The GDPR and New Requirements on Sensitive Data

I. Data Protection Impact Assessment

One of the novel requirements of the GDPR is the need to perform a "Data Protection Impact Assessment"[111] (DIPA) in a number of circumstances where the proposed processing may "*represent a high risk to the rights and freedoms of natural persons.*"[112] The required content of such assessments is currently the subject of debate both within and beyond academic circles.[113] The GDPR does not exhaustively describe all the situations where a data protection impact assessment is required but does describe certain occasions where it shall be required, including situations that require "processing on a large scale of special categories of data."[114] As section D discussed, in the era of big data, the processing of data, including sensitive forms of data such as health data will increasingly become the norm. This arguably means that more and more processing of sensitive data will *inter alia* meet the criterion of "processing on a large scale of special categories of data" and, thus, warrant a DPIA.

In terms of what exactly may be required concerning the form such an impact assessment should take or what substance it should have, there is currently much uncertainty, though some guidance has been created in order to aid potential data controllers.[115] These range from harms on the individual level, such as privacy harms, to wider and more diffused harms produced at the societal level, including the types of harms that may be traditionally associated with the improper use of sensitive data and which were discussed in section B.I. What is certain is that the GDPR is demanding that data controllers consider issues that go beyond those one might have traditionally associated with data protection. A consideration of all such harms and the measures needed to

[108]Bart W. Schermer, Bart Custers & Simone van der Hof, *The Crisis of Consent: How Stronger Legal Protection May Lead to Weaker Consent in Data Protection*, 16 ETHICS & INFO. TECH. 171 (2014).

[109]Sandra Wachter, *Normative Challenges of Identification in the Internet of Things: Privacy, Profiling, Discrimination, and the GDPR*, 34 COMPUT. L. & SEC. REV. (2018).

[110]Mantovani & Quinn, *supra* note 38.

[111]GDPR, *supra* note 1, at art. 35.

[112]*Id.* at art. 35(a).

[113]For discussion on the potential breadth of Article 35, see Dariusz Kloza, Niels Van Dijk, Raphaël Gellert, István Böröcz, Alessia Tanas, Eugenio Mantovani & Paul Quinn, *Data Protection Impact Assessments in the European Union: Complementing the New Legal Framework Towards a More Robust Protection of Individuals*, POLICY BRIEF (2017).

[114]GDPR, *supra* note 1, at art. 35.

[115]Article 29 Data Protection Working Party, *Guidelines on Data Protection Impact Assesment (DIPA) and Determining Whether Processing is "likely to result in a high risk" for the Purposes of Regulation 2016/679* (Working Paper 248 rev.01, 2017).

https://doi.org/10.1017/glj.2021.79 Published online by Cambridge University Press

mitigate them—if this is indeed what Article 35 is demanding—[116] may often be a considerable exercise demanding a truly multidisciplinary perspective from diverse disciplines such as ethics, law, and sociology.[117] Again, the example of health data can be used to illustrate the potential breadth of issues that could be involved and would likely merit consideration within a DPIA. Health data has a relevance that goes far beyond medical treatment, which is itself of a highly sensitive nature. The improper use of such data can thus have a range of consequences not only in the healthcare sector but in domains far beyond, including discrimination in employment, insurance, and a range of other areas. In addition, the GDPR's use of the term "a risk to rights and freedoms" means it may be necessary to go far beyond traditionally considered harms such as discrimination and include other more complex issues than the protection of fundamental rights and the prevention of phenomena such as stigmatization and marginalization.[118] The requirement to mobilize and use expertise in such areas may be onerous for many data controllers, especially when they are smaller entities or individuals. Given the potential effort required, the need to perform a DPIA will likely form a deterrent to the potential processing of sensitive data in certain contexts, if this duty is not accompanied by financial or organizational support.

Whilst forming a potential extra burden on data controllers, where DPIAs are performed they should result in a greater consideration of the harms that the use of such data can produce and therefore represent an added protection for data subjects.[119] The *raison d'être* behind DPIAs was to make data controllers more responsible for considering the potential negative externalities of their processing decisions. This involves requiring data controllers to go beyond the consideration of obvious risks such as data breaches and the harms that could stem therefrom. In doing so, it is hoped that they will reduce the risk of discrimination and related harms that are often given a reason for the existence of data protection regulation itself. Given the wide potential scope for DPIAs that appears to be envisaged within the GDPR one might hope that they may play a role in preventing harms not only fundamental rights—for example, privacy and discrimination—but also the wide array of related harms that can be produced from improper use of data.[120] If executed properly and thoroughly, this may well be the case, but there are also risks that DPIAs may not be so effective, especially given the potential for the volume of sensitive data to increase greatly and the associated extra burdens that may be placed upon data controllers.[121]

II. Data Protection Officers

The GDPR envisages that in a number of instances it will be necessary for controllers to appoint a Data Protection Officer (DPO).[122] One of these is described as: "[W]here the core activities of the controller or the processor consist of processing on a large scale of special categories of data ...."

[116]For discussion on the potential breadth of Article 35, see Kloza, *supra* note 116.

[117]Alessandro Mantelero, *AI and Big Data: A Blueprint for a Human Rights, Social and Ethical Impact Assessment*, 34 COMPUT. L. & SEC. REV. 754, 754–72 (2018); David Wright & Charles D. Raab, *Constructing a Surveillance Impact Assessment*, 28 COMPUT. L. & SEC. REV. 613, 613–26 (2012).

[118]Kloza, *supra* note 116. *See also* Quinn & Quinn, *supra* note 45. On the link between 'risks to rights and freedoms' and impacts on individuals, see Niels van Dijk, Raphaël Gellert, & Rommetveit, *A Risk to a Right?*, 32 COMPUT. L. & SEC. REV 286, 304 (2016). *See also* Katerina Demetzou, *Data Protection Impact Assessment: A Tool for Accountability and the Unclarified Concept of "High Risk" in the General Data Protection Regulation*, 14 COMPUT. L. & SEC. REV (2019).

[119]*See* GDPR, *supra* note 1, at Recital 75.

[120]*See* Niels van Dijk, Raphaël Gellert & Kjetil Rommetveit, *A Risk to a Right? Beyond Data Protection Risk Assessments*, 32 COMPUT. L. & SEC. REV. 286, 304 (2016); *See also* Katerina Demetzou, *Data Protection Impact Assessment: A Tool for Accountability and the Unclarified Concept of "High Risk" in the General Data Protection Regulation*, 35 COMPUT. L. & SEC. REV. 105342 (2019).

[121]*See* section G.I.

[122]GDPR, *supra* note 1, at art. 37.

https://doi.org/10.1017/glj.2021.79 Published online by Cambridge University Press

In short, the role of the DPO is to help ensure adherence to data protection and other forms of law when certain forms of data are processed in certain contexts.[123] This involves both giving advice on the law—and how to implement it—and verifying whether such implementation has occurred.[124] Other requirements involve giving advice on the implementation of DPIAs, as discussed above. This function can therefore play an important role in ensuring compliance with the requirements of the GDPR and protecting data subjects,[125] *inter alia* in contexts where sensitive data is produced. The existence of this function will also represent an added administrative burden for controllers that wish to process sensitive data. Such individuals should be adequately trained—not only in the rigors of the GDPR but also all other potentially applicable law—and should have high level of decisional independence.[126] Such requirements will entail the commitment of resources and may represent important burdens for smaller data controllers that have limited personal and resources. Their role may entail a high level of legal knowledge, especially in instances where GDPR permits Member States to maintain their own additional law on concerning the use of certain types of personal data.[127] In such instances, a likely solution may be to access external expertise, often at considerable cost.[128]

III. The Possibility for Extra Protection in Member State Law

Another important difference between non-sensitive and some forms of sensitive personal data is Member States may maintain divergent laws in a number of instances. This is a result of Article 9(4) of the GDPR which states: "Member States may maintain or introduce further conditions, including limitations, with regard to the processing of genetic data, biometric data or data concerning health."[129] Whilst this does not cover all forms of sensitive data described by the GDPR, it does cover some very important types, most notably health data, which, for the reasons described above in Section D, is likely to cover an ever-broader range of personal data going forward, especially if a context-based definition continues to be deployed. The consequence of Article 9(4) of the GDPR is that Member States do not need to harmonize their law concerning the processing of such data. Member States will therefore be able to maintain—and add to—the complex and diverse array of laws they already have concerning for example the use of medical files in particular or health data in general. Such laws are exacting and their variation across Europe makes the utilization of health data on a cross-border basis problematic. The demanding nature of such laws and the variation in form therefore represents a considerable added burden for parties that wish to process such types of data.[130] This will form a particular difficulty for DPOs that will be tasked with taking such laws into account[131] given that they will have to be aware of not only the requirements of the GDPR on such issues, but also a potentially complex web of applicable Member State law.

[123]For further discussion of these requirements, see Article 29 Data Protection Working Party, *Guidelines on Data Protection Officers ('DPOs')*, (Working Paper 243, 2016).

[124]Atana Yordanov, *Nature and Ideal Steps of the Data Protection Impact Assessment under the General Data Protection*, 486 EUR. DATA PROT. L. REV. (2017).

[125]*See, e.g.*, Miguel Recio, *Data Protection Officer: The Key Figure to Ensure Data Protection and Accountability*, 3 EUR. DATA PROT. L. REV. 114 (2017).

[126]*See* GDPR, *supra* note 1, at art. 38(1)

[127]*See* section F.III.

[128]Quinn & Quinn, *supra* note 45.

[129]GDPR, *supra* note 1, at art. 9(4).

[130]Article 37(5) of the GDPR demands inter alia that the DPO have a knowledge and understanding of all potentially applicable legal frameworks.

[131]*See* section F.II.

https://doi.org/10.1017/glj.2021.79 Published online by Cambridge University Press

IV. Stricter Conditions for Automated Processing.

Automated processing is likely to form a central plank of many novel uses of data, in contexts as diverse as advertising, insurance, and health care. Developments such as those discussed in section D will allow powerful analytical techniques to discover relationships and correlations between various data and allow decisions to be made that may be seen as bringing about an advantageous result. Such processes also, however, threaten to bring about or exacerbate problems associated with discrimination and associated phenomena.[132] This is because they are able to create and act upon harmful stereotypes in ways that may produce harms for vulnerable or sensitive groups. As a result, the use of automated decision making has given rise to concerns, especially in certain areas that have been traditionally considered as being associated with risks for vulnerable and marginalized groups. The phenomenon of "machine bias" is becoming perceived of as an ever more important risk in the future.[133] From the perspective of this Article, it is notable that the risks most often associated with automated processing are broadly similar to the reasons often put forward for the existence of sensitive data.[134] Automated decision-making is accordingly often described as a practice that will exacerbate the risks associated with the processing of sensitive data in general.

Given these risks, it seems appropriate that the GDPR foresees stricter conditions for the automated decision-making data processing based on sensitive data. In particular, Article 22(4) restricts the potential legal bases for such processing for controllers wishing to process sensitive data in this fashion. This leaves data controllers with the sole option of obtaining explicit consent unless legislation exists under national law permitting such processing where it is for reasons of substantial public interest.[135] In many cases, the latter option will not be available, meaning that data controllers will be forced to obtain explicit consent. Given that explicit consent must be informed and unambiguous[136] this should reduce the risks of harms occurring, given that individuals from at-risk backgrounds may arguably be in the best position to appreciate such risks and therefore not give consent where this is appropriate. Whilst this will undoubtedly be true in certain instances, the use of explicit consent itself is not a panacea for concerns surrounding automated processing for a number of reasons. First, as discussed in section E, the changing nature of consent in the electronic online age and the way in which this has rendered the concept of explicit consent a less onerous burden than was the case in earlier times. This means that consent may not be such an important barrier as it might seem at first glance. A second problem is the difficulties data controllers may have in adequately explaining what processing of data is occurring and what are the potential outcomes. This is because, given the nature of many automated forms of data processing, it may be difficult to understand in advance what is occurring. This may be particularly true with forms of machine learning where computer programs effectively decide based on complex forms of analysis of sample data what processes are to be applied.[137]

G. The Potential Effects and Risks of More Sensitive Data

The changing *de facto* nature of sensitive data, taken together with the continued use of a potentially expansive context-based definition of sensitive data, will mean that an ever-increasing proportion of data will become sensitive data. As the paragraphs below will discuss, problems may

[132]Sandra Wachter and Brent Mittelstadt, *A Right to Reasonable Inferences: Re-Thinking Data Protection Law in the Age of Big Data and AI*, 2019(2) COLUM. BUS. L. REV. 2 (2019).

[133]Danton S. Char, Nigam H. Shah & David Magnus, Implementing Machine Learning in Health Care - Addressing Ethical Challenges, 378 THE NEW ENGLAND J. MED., no. 11 (2018).

[134]*See* section B.

[135]GDPR, *supra* note 1, at art. 22(4).

[136]*See* section E.

[137]Andrew D. Selbst & Solon Barocas, *The Intuitive Appeal of Explainable Machines*, 87 FORDHAM L. REV. 1085 (2018): 1085; Sandra Wachter, Brent Mittelstadt, & Chris Russell, *Counterfactual Explanations without Opening the Black Box: Automated Decisions and the GDPR*, 31 HARV. J. L. & TECH. 841 (2018), http://arxiv.org/abs/1711.00399.

https://doi.org/10.1017/glj.2021.79 Published online by Cambridge University Press

arise in a world where more and more data is of a sensitive nature not only because these requirements may entail an added administrative burden for potential controllers, but also because in certain circumstances the increased prevalence of sensitive data may bring about a devaluation —or inflation—of the concept. Furthermore, it is possible that in certain circumstances such changes could also induce potential data controllers to attempt to circumvent the concept all together. As section B.II discussed, these problems are not only raising questions in an academic sense but also more serious questions about the ability of the GDPR to meet the fundamental rights requirement that an effective form of regulation exists for the use of sensitive data.

I. More Administrative Burdens on Controllers

Perhaps the most obvious issue relates to the fact that the use of sensitive data means that controllers will have to endure a higher regulatory burden. If the proportion of personal data that is sensitive data is higher, this burden may apply in instances where the data in question might not hitherto have been considered of a sensitive nature. Such burdens would include a more frequent need to utilize legal bases associated with sensitive data, including, as section E discussed, explicit consent, the more frequent need to appoint a DPO, conduct a Data Protection Impact Assessment, and implement other administrative requirements requested by national legislations pertaining to sensitive data.[138],[139]

Another important burden that is likely to apply more frequently, given the changing backdrop above, may often apply even where the data in question turns out not to be sensitive at all. This is due to the need to perform a threshold analysis to discern what data one is in possession of. This is needed to discern both whether one is dealing with personal data and, if so, whether that data is sensitive. In order to do this, it is necessary to analyze the data in question in the context of the other data that might be available to the controller—including from publicly available sources—and the processing methods that could be applied.[140] Performing such an exercise in order to discern the possible presence of sensitive data will become a more frequent and more demanding exercise. With each year that passes, and the increasing use of large—often big—datasets, the likelihood that data could be both personal and sensitive in nature will increase. The fact that the presence of sensitive data will become less and less intuitive will mean that exercises of determination—of the existence of sensitive data—will have to occur more often. Increases in both the quantity of potential complimentary available data and computing power—or "computational capacity"—will also make such a determination a more demanding task. More possibilities for combination with other alternative sources will have to be taken into account.[141] The increasing ability to deduce relationships brought about by forms of big data analysis as deep learning will also make the exercise of interpreting when data may or may not be sensitive more difficult.[142] Such techniques create a range of problems in terms of foreseeability.[143] Given that they can result in finding correlations and relationships that were completely unexpected, they may make predicting the presence of sensitive data increasingly difficult.[144] This will be particularly true where data protection frameworks such as the GDPR foresee a primarily context-based definition for certain types of sensitive data.[145] The result of this may be the need to adopt an ever more

[138]*See* section F.

[139]*See* section I.II.

[140]Khaled El Emam & Cecilia Álvarez, *A Critical Appraisal of the Article 29 Working Party Opinion 05/2014 on Data Anonymization Techniques*, 5 Int'l Data Priv. L. 73 (2015).

[141]Malgieri & Comandé, *supra* note 79.

[142]Comandé & Schneider, *supra* note 82.

[143]Quinn & Quinn, *supra* note 45.

[144]Reza Shorki and Vitaly Schmatikov, "Privacy-Preserving Deep Learning," *CCS '15 Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security* (2015).

[145]*See* section C.III.

https://doi.org/10.1017/glj.2021.79 Published online by Cambridge University Press

cautionary approach towards potential personal data, assuming that it may often be of a sensitive nature also even where this may not intuitively appear to be the case and, thus, increasing the administrative burden attached to it.

II. Sensitive Data Inflation

A further risk is that the concept of sensitive data itself becomes devalued over time. If an ever-greater proportion of personal data is likely to be sensitive data, one might begin to question the value of sensitive data. Whereas in the past sensitive data represented a small fraction of personal data for which the need for a higher regulatory burden may have been easier to understand, this may change in a future where an ever-greater proportion of personal data is sensitive in nature. One can arguably say that this is already the situation, for example, with regards to the use of health data given the increased acceptance of digital forms of consent.[146] Whilst facilitating various forms of mHealth processes that would not have been possible, such forms of consent arguably represent a far lower barrier to the utilization of sensitive data in comparison with more traditional requirements of consent for the processing of medical data, for example, signed written consent forms. We would argue that the change in perception of what is required in terms of consent for the use of health data can, in part, arguably be attributed to an inflation effect that has been produced as a result of the ever-increasing quantity of health data that is continuously being created and the innovation of processes designed to make use of it. This has in some ways arguably reduced the perception of the importance of the concept, allowing for a softening of consent as a barrier function to its use.[147]

As section F discussed, the GDPR in introducing a new range of requirements for the processing of *inter alia* sensitive data—especially on a large scale—arguably provides some balance to the situation that has arisen *vis-á-vis* the changing nature of explicit consent. Requirements such as the need to appoint DPOs or to conduct a DPIA will act to reduce some of the risks that are associated with the use of sensitive data and which may have arguably been increased by the weakening of the "consent barrier." There exists, however, the risk that, with time and the potential "inflation" of sensitive data, these tools will also become weakened. In a world where more and more data are sensitive data, one might ask whether instruments such as DPIAs will retain the same level of effectiveness. Given the potential ubiquity of sensitive data, will the effort expected of and put into such requirements remain the same? If such exercises become the norm, arguably there is a risk that they will be reduced to tick box exercises, for instance, through software that automatically conducts or structures large parts of such assessments.[148] Such a risk will increase where too few resources are deployed for an ever-increasing number of assessments to make them meaningful. This would reduce their ability to prevent the sort of harms that are associated with the use of sensitive data—including *inter alia* risks such as discrimination and associated effects. At present, these exercises are not seen as the norm, but as undertakings that must be engaged in when the data processing is seen to bring higher risks.[149] This arguably provides such exercises with a certain level of gravity that might be reduced if they become the norm.

[146]About the limits of consent see generally Gabriella Fortuna-Zanfir, *Forgetting about Consent. Why the Focus Should Be on 'Suitable Safeguards' in Data Protection Law, in* RELOADING DATA PROTECTION, 237–55 (Serge Gutwirth, Ronald Leenes, & Paul De Hert eds., 2014); Bart W. Schermer, Bart Custers & Simone van der Hof, *The Crisis of Consent: How Stronger Legal Protection May Lead to Weaker Consent in Data Protection*, 16 ETHICS AND INFO. TECH. 171 (2014); Bart Custers, *Consent and Privacy*, *in* THE ROUTLEDGE HANDBOOK OF THE ETHICS OF CONSENT (2019), https://papers.ssrn.com/abstract=3383465.

[147]*See* section E.

[148]Kloza, *supra* note 116.

[149]*See, e.g.*, Kloza, *supra* note 116. *See also* Miguel Recio, *Data Protection Officer: The Key Figure to Ensure Data Protection and Accountability*, 3 EUR. DATA PROT. L. REV. 114 (2017).

https://doi.org/10.1017/glj.2021.79 Published online by Cambridge University Press

III. The Risk of Sensitive Data Protection Circumvention

Given the relative nature of sensitive data and the increasing potency of data mining techniques, it seems clear that risks of circumvention of sensitive data protection rules and safeguards are high.

In particular, data controllers might avoid processing formally sensitive data by substituting them with "proxies."[150] The classic example is not processing information related to the "race" of individuals, but just processing data related to postal codes. This may be possible because in several cities, data about neighborhood can be an effective proxy for inferring ethnic origins.[151] This represents an important disadvantage of an overly contextual definition of sensitive data:[152] Even where data cannot apparently be used to draw firm objective conclusions about individuals, it may be possible to make probabilistic correlations that are sufficient for commercial or other purposes. In the example quoted here, for example, the fact an individual comes from a certain area may only mean that there is a fifty percent chance that they belong to a particular ethnic minority. Whilst this may not be sufficient to conclude, with any reasonable degree of certainty, membership of such a group, it may in many instances be sufficient for purposes such as marketing or insurance for which low probabilistic determination may still have significant commercial importance. Similarly, imagine a situation where a data controller discovers that people having yellow cars are more likely to buy products popular amongst gay people. In such situations, data controllers interested in exploiting the commercial opportunities of targeted advertising to gay consumers will not need to process any data related to sexual orientation, but rather be able to use a proxy that has low—but sufficient—correlation.[153]

Importantly, where data is of a low probabilistic nature, potential controllers may arguably be able to process it whilst claiming that it does not represent sensitive data from a context-based perspective. This is because on an individual level the data in question may be too uncertain to draw a sensitive inference—even if on a macro level it may be of commercial use, such as for advertising purposes. Unfortunately, it is possible that whilst not being sensitive data, from a purely context-based perspective at least, its use may have the potential to produce the same adverse effects that could be associated with the use of sensitive data.[154] This situation arguably creates an incentive for certain data controllers to circumvent the use of objectively clear forms of sensitive data—from a context-based perspective—in favor of forms of data that will allow similar conclusions to be reached. This incentive is arguably increased by the changing *de facto* nature of sensitive data and the increased administrative burdens for data controllers who opt to process sensitive data, discussed in Section D.

H. The Way Forward?

The previous sections have outlined several of the problems that the concept of sensitive data may be likely to experience in the future. Most of these stem from the fact that the amount of and use of sensitive data is likely to increase enormously in the future. The extent to which such problems will manifest themselves is not set in stone but will be determined by how the concept of sensitive data itself is understood and interpreted. Of particular importance will be the extent to which the concept will be understood from an objective context-based viewpoint or, alternatively, a more subjective purpose-based perspective. At present, as section D.III discussed, the definition of

[150]Solon Barocas & Andrew D. Selbst, *Big Data's Disparate Impact*, 104 CALIF. L. REV. 671, 693 (2016).

[151]*See, e.g.*, GUIDO CALABRESI, IDEALS, BELIEFS, ATTITUDES, AND THE LAW: PRIVATE LAW PERSPECTIVES ON A PUBLIC LAW PROBLEM 35–36 (1995).

[152]Lokke Moerel & Corien Prins, *Privacy for the Homo Digitalis: Proposal for a New Regulatory Framework for Data Protection in the Light of Big Data and the Internet of Things*, SSRN Scholarly Paper (2016), 11, https://papers.ssrn.com/abstract=2784123.

[153]Comandè & Schneider, *supra* note 82; Sandra Wachter & Brent Mittelstadt, *A Right to Reasonable Inferences: Re-Thinking Data Protection Law in the Age of Big Data and AI*, 2019 COLUM. BUS. L. REV. 4 (2019).

[154]*See e.g.*, section B.

https://doi.org/10.1017/glj.2021.79 Published online by Cambridge University Press

sensitive data in the GDPR, for example, is heavily skewed towards a context-based perspective. This raises questions as to whether a future definition should be more purpose-based. This can be viewed from different perspectives. On the one hand, as section D above discussed, an understanding that is too context-based in nature will likely maximize the coverage of sensitive data, particularly in an environment where there is an ever-increasing prevalence of big data and interconnectivity. Whilst this may maximize coverage of the concept to include all forms of processing where there is a potential risk—related to the types of categories involved in sensitive data—the potential expansion of the concept in the coming years might likely bring with it a number of problems—as discussed in section G—related to increasing the burden on potential data controllers or the concept of "data protection inflation." On the other hand, an understanding of sensitive data that is too purpose based— in other words, turning on the intent of the data controller—is likely to minimize the coverage of the concept to an extent that may bring about its own problems. This will occur because it is possible for harms to result—related to the sensitive nature of data—even where the data in controller in question has no intention to process the data in a way that would reveal sensitive aspects. Current intent does not preclude future changes of intent or the actions of third parties that may be able to use the data in question—in ways that may not be entirely foreseeable at the present time. Such an understanding of the concept of sensitive data may therefore leave the door open to negligent use of personal data that could be thought of being capable of revealing aspects that may be sensitive in nature. In addition, there is the possibility that some data controllers may in certain instances be able to circumvent the notion of sensitive data where they only need to deduce relationships of a low probabilistic nature, for example in certain forms of advertising.[155]

Whilst the choice between a context or purpose-based understanding of the concept of personal data may appear black and white or diametrically opposed, the authors of this Article would argue that this need not necessarily be the case.

In particular, the authors of this Article would propose that a possible solution to these issues—both *de lege ferenda* and *de lege lata*—could be a hybrid approach. This approach would be mostly based on a purpose-based interpretation with a contextualbased "backstop." Indeed, it might be possible to combine elements of both understandings in a way that would moderate some of the harms discussed here, in other words, relating to an undesirable maximization or minimization of the concept of sensitive data. In discerning the presence of sensitive data, we would therefore propose a two-step process. As is outlined below this would depart with the question of whether there was an intent to use sensitive data on the part of the data controller.

I. Step I: Is there an intent to use the data in question for a sensitive purpose?

Determining an answer to this question will likely involve asking a series of important sub questions. Does the data controller intend to process sensitive data? Do they intend to use the data in question to arrive at conclusions—now or potentially in the future—that could be deemed of being of a sensitive nature?

Where the answer to such questions is "yes," based on the available evidence, such as the identity of controller, stated aims, its usual practices or commercial motives, then the data in question could be assumed as being sensitive personal data. In particular, this could include attempts to assemble large amounts of data which although not sensitive at present might one day in the future be considered so because of—for example—increases in scientific knowledge, computing power, or the increased availability of potentially compatible data.

II. Step II: Is There a Need to Use the Objective "Backstop?"

If the answer to the main question above was however "no," that would not necessarily be the end of the matter. This is because, to avoid some of the problems discussed above in section G.III, it is also

[155] *See* section G.III.

https://doi.org/10.1017/glj.2021.79 Published online by Cambridge University Press

necessary to determine that there is at least some level of objective analysis—irrelevant of the intentions of the data controller—concerning the nature of the data in question and its potential sensitivity. In such cases the context-based "back stop" should be applied. This involves asking the following question; regardless of the purpose of the data controller, is it reasonably foreseeable that in certain contexts the data in question could reveal sensitive aspects of data subjects or allow them to be inferred?

The authors of this Article would argue that such a hybrid view of what constitutes sensitive data can actually be supported looking at the wider way the GDPR refers to sensitive data. Unlike Article (9)(1), which seems to use mainly contextually orientated language, language that could be understood as being representative of a more purposeful understanding of what constitutes sensitive data, is used elsewhere. This is particularly true when one looks at the alternative legal bases to consent that are provided with in Article 9. This relates to the possibility to process data for reasons of "scientific research," "public health," and "substantial public interest."[156] Each of these potential legal grounds for the processing of sensitive data is defined in terms of a "purpose." This arguably demonstrates that the concept of purpose should by no means be seen as alien to the question of what constitutes sensitive data. Given this, the authors of this Article would argue that utilizing "purpose" in the question of "what is sensitive data" within the context of a hybrid definition as outlined above is by no means a "step too far."

One open issue concerning Step II is the threshold of the objective analysis. Given the risks of inflation and the creation of too many burdens identified in this Article, we would argue that the bar should not be set too low. It should therefore preclude distant theoretical possibilities that sensitive conclusions could be drawn from a particular dataset. Doing so would avoid many of the problems associated with employing solely a "maximalist" contextbased definition of personal data. At the same time, however, were the bar to be set too high, such as requiring it to be immediately obvious that the data in question was of a sensitive nature, then too much data that was in reality of a sensitive nature would be excluded. What is rather required is an intermediate level, in other words, where the bar would be set requiring a level of threshold analysis to discern whether or not it is reasonably foreseeable that the data in question could be considered as sensitive in nature.[157] This would include considering the computing power and analytical algorithms available and other potentially available complimentary sources.[158] It would also require a fair consideration of likely developments in the future.

The most difficult aspect of this proposal is admittedly determining the intensity of review that should be required. As with many areas of data protection, the particular context involved will be important in providing guidance. Going beyond mere intuition is obviously necessary but an exhaustive investigation of all theoretical possibilities would clearly be too. That effort would invoke too many burdens and would result in problems linked to the concept of "maximization" discussed above. Rather, such an analysis should look for what is reasonably foreseeable in the context which exists. Such a "backstop," if employed properly, should act to avoid many instances of negligent or poorly thought through processing that could put data subjects at risk.

I. Conclusion

The concept of sensitive data has represented an important pillar of the EU's and other— data protection frameworks for some time. Whilst the precise *raison d'être* of the concept has never been agreed upon, the justifications that have been put forward often relate to the need to prevent

[156]All of these grounds are outlined in article 9(2). One important exception article 9(2)(e) – "processing relates to personal data which are manifestly made public by the data subject" as this ground makes no reference to the purposes of the data controller.

[157]As regards a reasonableness test for sensitive data, see Vaclav Janecek & Gianclaudio Malgieri, *Data Extra Commercium*, *in* DATA AS COUNTER-PERFORMANCE—CONTRACT LAW 2.0? (S. Lohsse, R. Schulze and D. Staudenmayer eds., 2019) (Forthcoming), https://ssrn.com/abstract=3400620.

[158]Malgieri & Comandé, *supra* note 79.

https://doi.org/10.1017/glj.2021.79 Published online by Cambridge University Press

harms such as discrimination and stigmatization. There is furthermore an argument that the regulation of sensitive data with specific legal frameworks is a fundamental rights requirement, in other words linked to the right of privacy. The risk of harms in these areas is arguably greater for the processing of sensitive data. Irrespective of any basis justification, sensitive data is undergoing a transition which has intensified in recent years. This transition has been occurring both in the *de facto* real-world manifestation of sensitive data and the legal definition of the concept used within legislative instruments. In terms of the former, the nature of sensitive data is being transformed by evolutions linked to the never-ending augmentation of computing power and the increasing availability of various forms of big data. Such processes mean that the sensitive nature of a particular dataset may no longer be as intuitively obvious as it has been in the past. Such problems are becoming increasingly exacerbated with the increasing prevalence of big data and the increasing online availability complimentary data, including, *inter alia*, as a result of developments such as IoT. As a result, it is becoming more and more necessary to take precautionary measures with large datasets given that they may well contain sensitive data, even where this is not intuitively apparent.

In addition to this shift in the *de facto* nature of personal data, changes in the legal framework outlining how sensitive data can be used have been occurring. Whilst the primary difference between non-sensitive and sensitive data in earlier times was the existence of legal bases—for the processing of sensitive data—with more strenuous requirements, such as explicit consent, the emphasis has begun to shift towards a focus on greater administrative requirements, including the use of DPOs and DPIAs, for controllers that wish to process sensitive data. Accordingly, whilst the barrier of explicit consent may not be as insurmountable as it once was, there are a range of other requirements that will give data controllers pause for thought before deciding to process sensitive data. Some of these requirements are likely to be of an onerous nature, especially given the likelihood that more and more personal data will be sensitive data in the future.

Whilst the introduction of these new requirements arguably serves an important balancing function, going some ways to address issues such as the potential weakening of the "consent barrier," the authors have in this Article outlined a number of important risks that are likely to develop concerning the potential use of sensitive data. These stem, not only from the changes in the *de facto* nature of such forms of data, but also through uncertainty over how they will be defined in legal terms in the future. As this Article highlighted, sensitive data may be defined in a purposeful or a contextual manner. Thus far, EU law has generally adopted a contextual approach in judging whether data is sensitive or not. Originally, such an approach was straightforward given that it was often intuitively self-evident whether a particular dataset was of a sensitive nature or not; imagine, for instance, an electronic health record. With the rise of big data processing and the potential availability of almost limitless potentially complimentary data in an ever more interconnected world, the adoption of a purely contextual approach has become more problematic. This is because the sensitive nature of such data can no longer be judged intuitively. With big data, it may rather be very difficult to easily discern whether data is sensitive or not. Furthermore, upon investigation it may be likely that big data may often allow sensitive inferences to be drawn. Several risks arise from this. Most obvious is that many data controllers will be subjected to extra burdens where they have no intention of processing data in a way that can reveal sensitive information. The result of this could be the deterrence of certain forms of processing that have important economic, scientific, or social value, to avoid the risk that sensitive data is unintentionally processed. Other risks include an inflation of the concept of sensitive data whereby it becomes devalued, reducing the value of the administrative requirements attached to it to mere "tick box" exercises. Were this to happen, the likelihood of harmful outcomes for data subjects would seemingly increase.

The purpose of this Article was to analyze the definition of sensitive data in the context of problems that are likely to become more prevalent, in other words, risks of "inflation" or "circumvention." This exercise is necessary because, as Section B.II clarified, the *sui generis* protection of sensitive data—in other words, in addition to the protection of non-sensitive personal data—is

https://doi.org/10.1017/glj.2021.79 Published online by Cambridge University Press

seemingly demanded from a fundamental rights perspective. As the authors have argued in this Article, it is, as a consequence, necessary to "protect" the concept from the risks of either a too restrictive or too extensive interpretation. If the former were to occur—in other words, if only few specific types of data are sensitive—data controllers could easily process sensitive information using alternative proxies, circumventing the rules and safeguards designed to apply to sensitive data processing. If the latter were to occur; in other words, applying the label of sensitive data to any dataset revealing—even indirectly and implicitly—some sensitive information, the result could be that most, or maybe one day all, personal data falls under the scope of sensitive data. This problem is likely to intensify given the immense computational capacity of modern data mining algorithms and the big availability of other potentially compatible datasets, in the Big Data ecosphere, where any data might in principle reveal some sensitive aspects of data subjects. Were such a wide notion to be adopted, the concept of sensitive data would arguably become useless creating serious problems *viz-á-viz* the infringement of the fundamental rights requirement, discussed in section B, to protect sensitive data as a separate and specific form of personal data.

The first "minimalist" scenario described here could occur if a wholly purpose-based definition of sensitive data—in other words, determined solely by intention of the data controller—were to be used. The second "maximalist" scenario could occur if a wholly contextual-based definition of sensitive data—in other words, data are sensitive in any situation in which they might reveal sensitive aspects of the data subjects—were to be adopted.

The authors of this Article have accordingly argued that it may be necessary to rethink the approach that is used towards sensitive data. Whilst several options seem apparent none of them seem able to solve all of the problems discussed in this Article. Accordingly, this Article proposes a hybrid approach: A purpose-based interpretation of sensitive data, with a relevant context-based backstop. In other words, personal data should be considered sensitive IF the intention of the data controller is to process or discover sensitive information OR if it is reasonably foreseeable that, in a given context, the data in question can be used to reveal or to infer sensitive aspects of data subjects.

This formulation would have the advantage of not only seeing data as sensitive where there was an intention of processing sensitive data or a real risk of doing so, but would simultaneously avoid the label of sensitive data being applied where there was no intention to process sensitive data and where there was no reasonably foreseeable prospect that this could be the case. The authors of this Article would argue that it is only through such a formulation that a balance can be struck where the concept of sensitive data remains viable and a real level of protection is offered to data subjects who may be in a vulnerable position and at risk from discrimination and associated phenomena in line with their fundamental rights.

Cite this article: Quinn P (2021). The Difficulty of Defining Sensitive Data—The Concept of Sensitive Data in the EU Data Protection Framework. *German Law Journal* **22**, 1583–1612. https://doi.org/10.1017/glj.2021.79