Incompatible: The GDPR in the Age of Big Data

Tal Z. Zarsky[*]

I. INTRODUCTION AND ROAD MAP

In April 2016, after years of drafting and negotiations, the European Union (EU) finally passed the General Data Protection Regulation ("the GDPR" or "the Regulation").[1] Ever since, regulators, businesses and citizens in Europe and far beyond are grappling with the difficult task of establishing the legal regime which will follow from the GDPR's entry into force in May 2018.[2] The GDPR's impact will, most likely, be profound. It is perhaps the most comprehensive and forward looking piece of legislation to address the challenges facing data protection in the digital age. It replaces[3] the 1995 Data Protection Directive (DPD),[4] and is set to guide the EU throughout the next few decades.

[*] Vice Dean and Professor – University of Haifa. I thank Gaia Bernstein, Frederik Borgesius, Courtney Bowman, Mireille Hildebrandt, Bart van der Sloot, as well as Chelsea Ott, the members of the *Seton Hall Law Review* and the participants of "The New EU Data Protection Regulation: Transnational Enforcement and its Effects on U.S. Businesses" Symposium at Seton Hall Law School (September 2016) for their helpful comments.

[1] Regulation 2016/679 of the European Parliament and of the Council on the Protection of Natural Persons with Regard to the Processing of Personal Data and on the Free Advancement of Such Data, and repealing Directive 95/46/EC, 2016 O.J. L 119/1 [hereinafter the General Data Protection Regulations or GDPR].

[2] *Id.* art. 99, at 87–88.

[3] *Id.* art. 94(1), at 86.

[4] Council Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the Protection of Individuals with Regard to the Processing of Personal Data and on the Free Movement of Such Data, 1995 O.J. L 281/32 [hereinafter DPD].

The GDPR enters into force at a crucial time for the digital economy and ecosystem; one in which substantial risks to rights and liberties are emerging, while at the same time vast opportunities to create value, promote welfare, and enhance various social objectives are unfolding. Enacting complex regulation pertaining to a rapidly changing environment is always a challenging task and the data protection context is certainly one that is constantly in flux and is, therefore, no exception.

Among the challenges data protection law faces in the digital age, the emergence of Big Data is perhaps the greatest: this term refers to the practices of creating and analyzing vast datasets, which at times include personal information. Indeed, Big Data analysis carries both hope and potential harm to the individuals whose data is analyzed, as well as other individuals indirectly affected by such analyses. These novel developments call for both conceptual and practical changes in the current legal setting.[5] Unfortunately, the GDPR fails to properly address the surge in Big Data practices. The GDPR's provisions are—to borrow a key term used throughout EU data protection regulation—*incompatible* with the data environment that the availability of Big Data generates.[6] Such incompatibility is destined to render many of the GDPR's provisions quickly irrelevant. Alternatively, the GDPR's enactment could substantially alter the way Big Data analysis is conducted, transferring it to one that is suboptimal and inefficient. It will do so while stalling innovation in Europe and limiting utility to European citizens, while not necessarily providing such citizens with greater privacy protection.[7]

To defend and explain the provocative assertions noted above and the conclusions that derive from them, this article ("Article") proceeds as follows: after this brief introduction (Part I), Part II quickly defines Big Data and its relevance to EU data protection law. Part III addresses four central concepts of EU data protection law as manifested in the GDPR: Purpose Specification, Data Minimization, Automated Decisions and Special Categories. It thereafter proceeds to demonstrate that the treatment of every one of these concepts in the GDPR is lacking and in fact incompatible with the prospects of Big Data analysis. Many of these points have been made by

[5] Bart van der Sloot & Sascha van Schendel, *Ten Questions for Future Regulation of Big Data: A Comparative and Empirical Legal Study*, 7 JIPITEC 29 (2016), http://www.jipitec.eu/issues/jipitec-7-2-2016/4438 (illustrating which countries are considering novel regulation to approach the challenges of Big Data).

[6] *See, e.g.*, DPD, *supra* note 4, art. 6(1)(b), at 40; GDPR, *supra* note 1, art. 5(1)(b), at 35.

[7] See Tal Z. Zarsky, *The Privacy–Innovation Conundrum*, 19 LEWIS & CLARK L. REV. 115, 161 (2015) [hereinafter Zarsky, *The Privacy–Innovation Conundrum*], for a full analysis of this argument.

European commentators and even regulators.[8] However, this Article highlights the central concerns and argues that these problems are substantial, even incurable. It does so by selecting the most crucial critiques which the age of Big Data generates, carefully examining their components, and framing them in the most convincing manner. Part IV concludes by discussing the aggregated effect of such incompatibilities on regulated entities, the EU, and society in general.

This Article's focus is on legal analysis. It therefore does not provide a systematic discussion as to the reasons that led to the GDPR's incompatibility, which are most likely varied and range from the political to the economic.[9] Yet it is important to note that the incompatibility here discussed most likely resulted from a conscious policy decision, rather than from the EU decision-makers' misunderstanding of what lies ahead or mere regulatory neglect. As opposed to the conditions surrounding the enactment of the 1995 DPD, which pre-dated much of the Big Data revolution, a rich and vibrant discussion regarding the effects and implications of Big Data has unfolded in Europe over the last few years.[10] Indeed, the GDPR, in general, is premised on deep philosophical convictions regarding the extent to which the specific rights of both individuals and groups must be protected in the digital age.[11] Furthermore, as part of the GDPR's drafting process, firms engaging in Big Data (and their relevant associations) voiced their concerns

---

[8] *See* Ira S. Rubinstein, *Big Data: The End of Privacy or a New Beginning?*, 3 INT'L DATA PRIVACY L. 74, 74 (2013), https://academic.oup.com/idpl/article-lookup/doi/10.1093/idpl/ips036 (noting that the "Big Data tsunami" will "overwhelm" the EU's reform efforts). *See also Big Data – Privacy Principles Under Pressure*, DATATILSYNET 42–44 (Sept. 2013), https://www.datatilsynet.no/globalassets/global/04_planer_rapporter/big-data-engelsk-web.pdf (displaying a report of Norwegian Data Protection Authority).

[9] For such an analysis of the GDPR's article 22, see Sandra Wachter, Brent Mittelstadt & Luciano Floridi, *Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation*, SSRN 9–10 (2016), https://ssrn.com/abstract=2903469. In the context of purpose specification, see 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 51–52 (May 25, 2016), https://ssrn.com/abstract=2784123.

[10] *See, e.g.*, Paul de Hert & Hans Lammerant, *Predictive Profiling and Its Legal Limits: Effectiveness Gone Forever?*, *in* THE NETH. COUNCIL FOR GOV'T POLICY, WRR, EXPLORING THE BOUNDARIES OF BIG DATA 145 (Bart van der Sloot, Dennis Broeders & Erik Schrijvers eds., 2016); Antoinette Rouvroy, *"Of Data and Men": Fundamental Rights and Freedoms in a World of Big Data*, COUNCIL OF EUR., DIRECTORATE GEN. OF HUM. RTS. AND RULE OF L., at 11 (Jan. 11, 2016), https://rm.coe.int/CoERMPublicCommonSearchServices/DisplayDCTMContent?documentId=09000016806a6020.

[11] *See* Colin J. Bennett & Robin M. Bayley, *Privacy Protection in the Era of 'Big Data': Regulatory Challenges and Social Assessments*, *in* EXPLORING THE BOUNDARIES OF BIG DATA 205, 212 (2016) ("The GDPR contains a more accurate and faithful expression of the various policy instruments that currently comprise the 'governance of privacy' . . . It is rooted in the traditions of European data protection law, but it also borrows from policy innovations first introduced in countries outside Europe.").

regarding the impact of various contemplated GDPR provisions to the relevant decision-makers.[12]

The critical discussion to follow is therefore a call to all relevant regulators to account for the insights here discussed and reconsider the balance reached in the GDPR and the ideological convictions behind them, to the greatest extent. It does not merely write off the EU's political convictions and its reliance on fundamental rights which also include privacy and data protection. Rather, it strives to clearly articulate what the impact of the GDPR's implementation on Big Data practices would be, arguing that it would be substantial and highly problematic. Even through the GDPR's text is final, a critical discussion of its content is far from futile, even on the practical/policy level. There will be plenty of opportunities to engage in changes, as over the next few years, courts, national legislators and regulators will respond to, interpret and enforce the new regulation. Thus, the policy and legal position taken when drafting the GDPR might still shift and evolve. Therefore, the time to discuss these pressing matters is certainly now.

II. A BRIEF PRIMER ON BIG DATA AND THE LAW

Industry leaders' proclamations, as well as publications in the popular and academic press, continuously announce that the age of Big Data is upon us[13]—a tectonic change in the way data is collected, analyzed and applied in the digital era. When striving to define the "Big Data" concept, the professional literature refers to the four Vs: the **V**olume of data collected, the

[12] Jennifer Baker, *EU Data Protection Proposals Taken Word for Word from US Lobbyists*, TECHWORLD (Feb. 12, 2013), http://www.techworld.com/news/security/eu-data-protection-proposals-taken-word-for-word-from-us-lobbyists-3425637/; *see also*, *e.g.*, *DigitalEurope Comments on the Risk-Based Approach*, DIGITALEUROPE 6–10 (Aug. 28, 2013), http://www.digitaleurope.org/DocumentDownload.aspx?Command=Core_Download&EntryId=601. *See generally Big Data: A New World of Opportunities*, NESSI (Dec. 2012), http://www.nessi-europe.com/Files/Private/NESSI_WhitePaper_BigData.pdf (addressing the benefits of big data and how legal rules should be set accordingly).

[13] *See* Jonathan Shaw, *Why "Big Data" Is a Big Deal: Information Science Promises to Change the World*, HARV. MAG. (Mar.-Apr. 2014), http://harvardmagazine.com/2014/03/why-big-data-is-a-big-deal; Steve Lohr, *The Age of Big Data*, N.Y. TIMES (Feb. 11, 2012), http://www.nytimes.com/2012/02/12/sunday-review/big-datas-impact-in-the-world.html; Svetlana Sicular, *Gartner's Big Data Definition Consists of Three Parts, Not to Be Confused with Three "V"s*, FORBES (Mar. 27, 2013), http://www.forbes.com/sites/gartnergroup/2013/03/27/gartners-big-data-definition-consists-of-three-parts-not-to-be-confused-with-three-vs/#39afa0e13bf6l. *See also Big Data*, GARTNER, http://www.gartner.com/it-glossary/big-data/ (last visited May 15, 2017) (defining "Big Data"); Chris Forsyth, *For Big Data Analytics There's No Such Thing as Too Big: The Compelling Economics and Technology of Big Data Computing*, 4SYTHCOMM.COM (Mar. 2012), http://www.cisco.com/c/dam/en/us/solutions/data centervirtualization/big_data_wp.pdf?utm_source=datafloq&utm_medium=ref&utm_campaign=datafloq.

Variety of sources, the **V**elocity with which the analysis of the data can unfold, and the **V**eracity of the data which could (arguably) be achieved through the analytical process.[14]

To facilitate the discussion this Article sets forth, the term "Big Data" would be broadly applied[15] to address a fundamental change in the way data is *collected*, *stored*, and subsequently *used*—all a result of recent technological developments. In today's digital age, data is *collected* using multiple sensors as well as through various applications which record user's movements, communications and transactions.[16] It is *stored* using sophisticated mechanisms on distributed databases, the cost of which is constantly shrinking. Finally, it is *used* in advanced analytical processes and thereafter applied in a variety of contexts.[17]

Furthermore, Big Data often refers to specific advanced forms of data analyses, at times even machine-driven[18] and powered by data mining tools.[19] These automated processes, which are the focus of this Article's discussion, seek out correlations and clusters within the vast datasets, with analysts merely setting overall parameters and sifting through the results to set aside obvious errors.[20] Indeed, data-driven (as opposed to query-driven or human-driven) processes enable analysts to utilize the huge databases at their disposal, especially in instances in which the analysts actually do not know where to start looking.

While Big Data analyses which are driven by automated processes

---

[14] Sicular, *supra* note 13. Some literary sources even add a fifth "V" – that of "Value," yet this factor seems rather speculative and is thus best omitted. Bernard Marr, *Why only one of the 5 Vs of big data really matters*, IBM BIG DATA & ANALYTICS HUB (Mar. 19, 2015), http://www.ibmbigdatahub.com/blog/why-only-one-5-vs-big-data-really-matters.

[15] For a similar working definition, see Mireille Hildebrandt, *Location Data, Purpose Binding and Contextual Integrity: What's the Message?*, *in* PROTECTION OF INFORMATION AND THE RIGHT TO PRIVACY – A NEW EQUILIBRIUM? (Luciano Floridi ed., 2014), https://works.bepress.com/mireille_hildebrandt/54/ (using the term "Big Data Space").

[16] PRESIDENT'S COUNCIL OF ADVISORS ON SCI. & TECH., EXEC. OFFICE OF THE PRESIDENT, BIG DATA AND PRIVACY: A TECHNOLOGICAL PERSP. 22 (2014), https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/PCAST/pcast_big_data_and_privacy_-_may_2014.pdf [hereinafter WHITE HOUSE REPORT].

[17] See Tal Z. Zarsky, *Desperately Seeking Solutions: Using Implementation-Based Solutions for the Troubles of Information Privacy in the Age of Data Mining and the Internet Society*, 56 ME. L. REV. 13 (2004) [hereinafter Zarsky, *Desperately Seeking Solutions*], for a discussion of these steps.

[18] *See Big Data*, GARTNER, http://www.gartner.com/it-glossary/big-data/ (last visited May 15, 2017) (defining "Big Data" to include the enablement of "process automation").

[19] *See generally* Tal Z. Zarsky, *"Mine Your Own Business!": Making the Case for the Implications of the Data Mining of Personal Information in the Forum of Public Opinion*, 5 YALE J.L. & TECH. 1 (2003) [hereinafter Zarsky, *"Mine Your Own Business!"*]; WHITE HOUSE REPORT, *supra* note 16, at 24.

[20] Tal Z. Zarsky, *Transparent Predictions*, 2013 U. ILL. L. REV. 1504, 1517 (2013) [hereinafter Zarsky, *Transparent Predictions*].

introduce a wide variety of practices, this Article further focuses its scope on those which feature two specific attributes. *First*, it focuses on processes which rely on and produce, respectively, data and information of personal nature. Indeed, Big Data analysis might feature vast datasets related to astronomical sightings, chemical compounds, geological measurements and other similar sources that do not involve specific individuals, yet their analysis might, at one point, advance their lives. These initiatives are not part of the current inquiry. Yet many other sources *are* of personal nature, and their analysis does generate data protection concerns.[21] These involve, for instance, datasets related to individualized consumption, communications, and actions. Note, however, that the advanced analytical abilities here discussed render additional spheres of data to be potentially "personal," such as demographic and statistical data pertaining to larger groups which can now be possibly attributed to specific persons.[22] These latter processes are therefore rendered relevant to the current discussion as well.

Second, this Article addresses instances in which the results of the Big Data analyses are applied to specific individuals and thus affect them directly. Again, this often is not the case. The outputs of Big Data processes—even those pertaining to personal data—are often merely statistical findings related to aggregated data which are applied broadly by the relevant firm. Yet, in other instances, the results of Big Data analysis are indeed used in a unique interaction with a specific individual—either directly or indirectly. Among others, this aspect of the data flow is enabled by the emergence of personalized digital interfaces which allow for the tailoring of interactions with users on the basis of previously collected data, even in real time.[23] While these two latter aspects of Big Data pertain to a mere sliver of the overall realm of big data uses, they potentially generate both the most difficult policy questions and risks on the one hand, and the greatest social benefits and advances on the other. Therefore, they are worthy of this specific inquiry.

As noted, business gurus sing the praise of big data analysis, explaining how it not only generates efficiency and promotes welfare, but also provides society with rich knowledge which could better our lives in a variety of contexts.[24] However, as the fascination with Big Data and its prospect grew,

---

[21] van der Sloot & van Schendel, *supra* note 5, ¶ 113 (referring also to statements made by the EDPS).

[22] GDPR, *supra* note 1, art. 4(1), at 33 (broadly defining "personal data" to include instances in which a specific individual is identifiable). *See also id.*, rec. 26; Ira Rubinstein & Woodrow Hartzog, *Anonymization and Risk*, 91 WASH. L. REV. 703, 710–11 (2016) (regarding the risk of re-identification).

[23] For more on this data cycle, see Zarsky, *Desperately Seeking Solutions*, *supra* note 17.

[24] For a summary of publications and reports making this point, see Omer Tene & Jules

skepticism quickly followed. Critics often claim that the term "Big Data" consists of mostly hype. Therefore, stories of its success must be taken (at least) with a grain of salt.[25] If that is indeed true, one might claim that the in-depth analytical discussion to follow is unnecessary. But as now follows, the analytical discussion this Article addresses is nonetheless crucial.

The "hype" critique of Big Data features two distinct aspects. First, one can certainly argue that the "Big Data Revolution"[26] is at best a mere evolution. It relates to a long list of practices, benefits, and problems which were continuously unfolding for some time, and at one distinct point captured the media's—and thereafter the public's—interest.[27] Given this gradual change, this argument might state there is no real technological revolution which requires the urgent recalibration of policy objectives and legal rules.

At its core, this critique has much truth to it. Thus far, however, it has limited relevance to any policy debate. "Big Data" is currently generating great interest, leading many firms to accelerate their adaptation to the new data environment. Therefore, even if the process is merely evolutionary by nature, law and policy can no longer ignore the incremental changes that have brought about this new digital era. Rather, they must provide responses to this change—be it a "revolution," or a mere "evolution."

A second, harsher critique argues that Big Data refers to a promise that cannot be fulfilled.[28] One might argue that these processes provide vast benefits in *theory*. Yet in *practice* they are costly (both in actual costs and

Polonetsky, *Big Data for All: Privacy and User Control in the Age of Analytics*, 11 NW. J. TECH. & INTELL. PROP. 239, 243–51 (2013).

[25] See Steve Dodson, *Big Data, Big Hype?*, WIRED (Apr. 2014), https://www.wired.com/insights/2014/04/big-data-big-hype/ (last visited Apr. 10, 2017), for a review of these arguments. *See also* Moerel & Prins, *supra* note 9, at 14.

[26] For a prominent reference to this term, see ROB KITCHIN, THE DATA REVOLUTION: BIG DATA, OPEN DATA, DATA INFRASTRUCTURES AND THEIR CONSEQUENCES (2014). In the consulting context, see Peter Groves, Basel Kayyali, David Knott & Steve Van Kuiken, *The 'Big Data' Revolution in Healthcare*, MCKINSEY & CO. (Jan. 2013), http://www.mckinsey.com/~/media/mckinsey/industries/healthcare%20systems%20and%20services/our%20insights/the%20big%20data%20revolution%20in%20us%20health%20care/the_big_data_revolution_in_healthcare.ashx.

[27] For a discussion of this argument in the press, see Jason Hiner, Dan Kusnetzky & Andrew Brust, *Big Data: Revolution or evolution?*, ZDNET (Apr. 2, 2012), http://www.zdnet.com/article/big-data-revolution-or-evolution/; Samuel Arbesman, *Five myths about big data*, WASH. POST (Aug. 16, 2013), https://www.washingtonpost.com/opinions/five-myths-about-big-data/2013/08/15/64a0dd0a-e044-11e2-963a-72d740e88c12_story.html.

[28] *See* Bennett & Bayley, *supra* note 11, at 206. *See also* Svetlana Sicular, *Big Data is Falling into the Trough of Disillusionment*, GARTNER (Jan. 22, 2013), http://blogs.gartner.com/svetlana-sicular/big-data-is-falling-into-the-trough-of-disillusionment/; Paul Ohm, Response, *The Underwhelming Benefits of Big Data*, 161 U. PA. L. REV. ONLINE 339 (2013), https://www.pennlawreview.com/online/161-U-Pa-L-Rev-Online-339.pdf.

other damages to data subjects and their rights) and their results flawed.[29] The data used is often inaccurate and its transformation to different contexts leads to problematic outcomes. But here, too, this argument cannot lead to rejecting the analysis to follow. Indeed, the benefits of Big Data are often exaggerated.[30] This Article will accept as given the fact that such analyses can generate substantial social benefits, if exercised with sufficient caution. This assumption is not farfetched, and is already reflected in central policy documents, such as those released by the White House in 2014.[31] It is, though, conceded that actual Big Data practices are in their infancy and will take years to assess.[32]

Big Data analysis of personal information is therefore a substantial dynamic and here to stay. It is also quite clear that it both affects and is affected by the extent of data protection policy. On the one hand, these advanced forms of data analyses can compromise the individuals' privacy rights and the control citizens have over their personal data. Thus, the availability of these tools might require stricter enforcement of privacy laws to so limit privacy-related harms.[33] On the other hand, stringent data protection laws impede the flow of personal data, as well as the ways it could be analyzed and used. In other words, stricter data protection and privacy laws compromise the growth of the Big Data industry and the benefits to be derived from it.

The noted double-sided tension between data analytics and data protection has not escaped the GDPR's framers. Quite to the contrary, such tension is already evident in the GDPR's title. The GDPR is a regulation "on the protection of natural persons," and thus would surely strive to limit the risks big data analysis might generate. Conversely, the title also indicates such regulation is "with regard to . . . the free movement of such data" and therefore must also facilitate the analysis of data, to the benefit of society.[34] Indeed, the GDPR's framers were called to balance between the ability to engage in big data analysis to its fullest extent and the protection of privacy interests and rights. It is the precise nature of this balance and its flaws that

---

[29] Harvey Schachter, *Unearthing Big Myths about Big Data*, THE GLOBE & MAIL (May 3, 2015), http://www.theglobeandmail.com/report-on-business/careers/management/unearthing-big-myths-about-big-data/article24176687/.

[30] Bennet & Bayley, *supra* note 11, at 206.

[31] EXEC. OFFICE OF THE PRESIDENT, BIG DATA: SEIZING OPPORTUNITIES, PRESERVING VALUES 48–58 (2014), https://obamawhitehouse.archives.gov/sites/default/files/docs/big_data_privacy_report_5.1.14_final_print.pdf.

[32] van der Sloot & van Schendel, *supra* note 5, at 22–25, 37 (indicating that very few Big Data initiatives have already been launched, yet many are planned for the near future; at this time, however, the extent of opportunities is unclear).

[33] *See id.* at 7–8; *see generally* Zarsky, *"Mine Your Own Business!"*, *supra* note 19.

[34] GDPR, *supra* note 1, at 1.

this Article will now explore.

The acknowledgement of the existence of this apparent tension is not shared by all. One might argue against the existence of such tension, while asserting that appropriate data protection regulation in fact *promotes* the extent of data analysis and the benefits derived from it. This is, in fact, the EU's official position as reflected in some of its policy documents. For instance, in a "Fact Sheet" addressing this specific matter, the European Commission explains that the new regulation can *enhance* big data analysis, as it would promote "trust" and thus lead to greater engagement with these platforms, more data and greater benefits.[35] A similar notion was set forth in an Article 29 Working Party opinion addressing the growing use of Big Data.[36] These opinions echo similar statements made by EU leaders in the past regarding the advantages and benefits of data protection law in general.[37]

The argument linking greater data protection and enhanced Big Data abilities is flawed, and most likely merely represents wishful thinking. For the most part, users thus far have failed to demonstrate in their actual practices and behavior (as opposed, perhaps, to their mere responses to surveys and questionnaires) that they sufficiently care as to whether trust is or is not maintained. Therefore, there is no reason to believe that firms will strive to promote and achieve such trust in the ways they design or operate their data-rich businesses.[38] This conclusion leads to the clear understanding that enhanced data protection in fact potentially undermines the abilities to engage in Big Data in general, and that rules adopted in the GDPR enhance these problems on various fronts in particular, as the next chapter

---

[35] Rouvroy, *supra* note 10, at 1.

[36] *See* ARTICLE 29 DATA PROTECTION WORKING PARTY, STATEMENT ON STATEMENT OF THE WP29 ON THE IMPACT OF THE DEVELOPMENT OF BIG DATA ON THE PROTECTION OF INDIVIDUALS WITH REGARD TO THE PROCESSING OF THEIR PERSONAL DATA IN THE EU, at 2 (2014), http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2014/wp221_en.pdf.

[37] For an extensive discussion of this point, see Zarsky, *The Privacy–Innovation Conundrum*, *supra* note 7, at 130. *See also* Viviane Reding, *The European Data Protection Framework for the Twenty-First Century*, 2 INT'L DATA PRIVACY L. 119, 129 (2012). See also statements by various EU regulators as noted in van der Sloot & van Schendel, *supra* note 5, at 52.

[38] Zarsky, *The Privacy–Innovation Conundrum*, *supra* note 7 (including references therein). *But see* Neil M. Richards & Woodrow Hartzog, *Privacy's Trust Gap: A Review*, 126 YALE L.J. 1180 (2017), https://ssrn.com/abstract=2899760; Courtney Bowman & John Grant, *A Marketplace for Privacy: Incentives for Privacy Engineering and Innovation*, *in* THE CAMBRIDGE HANDBOOK FOR CONSUMER PRIVACY (Evan Selinger et al. eds., forthcoming 2017) (draft on file with author) (stating that firms might strive to provide consumer trust given their interest to attract talented employees who might demand that the firm adhere to high moral and ethical standards; the strength of this intriguing argument is yet to be seen but should certainly be tracked and studied).

demonstrates.

III. THE GDPR'S INCOMPATIBILITY – FOUR EXAMPLES

As explained, data protection regulation, in its essence, is in tension with Big Data practices. This Part will further demonstrate that the balance set forth in the GDPR is unacceptable and suboptimal; in some instances the Regulation will undermine the ability to exercise big data analysis. In others, the availability of Big Data technologies undermines some of the measures and distinctions the GDPR features.[39] This assertion will be demonstrated by referring to the key provisions of *purpose limitation*, *data minimization*, *special categories*, and *automated decisions*.

As demonstrated below, all the noted provisions are reiterations of legal concepts stated in the Data Protection Directive and elsewhere in European law. Nonetheless, addressing these issues at this juncture (just before the GDPR comes into force), and in concert, is essential. Furthermore, merely stating that these matters were already legislated, regulated, discussed, and therefore, accepted is an insufficient response. The fact that these concepts were repeated in the GDPR (with adding some important changes) is of great significance which requires a renewed examination of these concepts.[40]

When the EU Data Protection Directive (DPD) passed, much of what we are now discussing was the stuff of science fiction. Given the fact that the legislator's intention as to how we must deal with today's novel challenges was unclear, one might argue that the rules based on the DPD should be somewhat bent and circumvented. However, and as noted,[41] the negotiations surrounding the GDPR's enactment were carried out in an environment in which the feasibility and benefits of Big Data analytics were acknowledged. It is therefore fair to state that the EU community has provided a clear sign of approval to the noted policy ideas by reintroducing these concepts into the law in the Big Data Age: a sign that would here be questioned.

---

[39] An additional context in which Big Data technology *might* further undermine distinctions set out in the GDPR is the technology's ability to undermine the distinction between identifiable and non-identifiable data. Given the wealth of recent academic and regulatory research regarding this matter, this Article sets this matter aside. For a discussion of this element, see Bennett & Bayley, *supra* note 11, at 210. For more on this matter, see generally Rubinstein & Hartzog, *supra* note 22; Rubinstein, *supra* note 8, at 78. Yet another issue this Article sets aside and does not address is the way in which Big Data undermines the ability to obtain meaningful and informed consent for subsequent data usage. For more on this argument, see *id.* at 78. This issue is set aside as the notion of consent raises vast problems and questions generally, and therefore arguing that the emergence of Big Data at this point undermines this already shaky notion is quite difficult.

[40] Rubinstein, *supra* note 8, at 74 (noting that the "Big Data tsunami" will "overwhelm" the EU's reform efforts).

[41] *See supra* notes 10–12 and accompanying text.

Furthermore, while one can debate whether the GDPR provides rules substantially different than those already existing in the DPD, it is quite clear it provides key jurisdictional and procedural innovations which are bound to make a substantive difference. The GDPR now clearly pertains to a far broader scope of entities engaging in Big Data analysis of information pertaining to EU residents. Therefore, the analysis of the Regulation is of greater importance as it will impact far more entities on an international level.[42] In addition, the rules set in the GDPR must be reevaluated as they will be taken far more seriously than those set out in the DPD, given the noticeable fines which can follow from non-compliance.[43] Thus, if until now some firms might have been willing to take the calculated risks of ignoring data protection compliance requirements, the GDPR will be sure to capture these firms' attention and force them to adopt relevant changes. Such changes would be substantial because firms will be more inclined to opt for erring on the side of caution given the extensive fines.

In what follows, this Part will address every one of the four provisions noted, their meaning and their origin, explain how they are linked to Big Data analyses, the problems the GDPR will create as well as possible mitigating factors, loopholes, and work-arounds already existing in the law. Every segment will nonetheless conclude that the current GDPR provisions lead to problematic outcomes given the prospect of Big Data analytics.

A. *Purpose Limitation*

Article 5(1)(b) of the GDPR sets forth the fundamental notion that personal data must be collected for a "specific, explicit and legitimate" purpose and cannot be further "processed" (a term broadly defined)[44] in a way which is "incompatible" with such original purposes.[45]

As many commentators and reports point out, purpose specification is clearly at odds with the prospect of Big Data analyses.[46] Quite often

---

[42] GDPR, *supra* note 1, art. 3. *See also* Christopher Kuner, *The Internet and the Global Reach of EU Law*, *in* THE COLLECTED COURSES OF THE ACADEMY OF EUROPEAN LAW 25 (forthcoming 2017), https://ssrn.com/abstract=2890930.

[43] GDPR, *supra* note 1, art. 83.

[44] *See id.* art. 4(2).

[45] *See id.* art. 5(1)(b).

[46] *See, e.g.*, Mireille Hildebrandt, *Slaves to Big Data. Or Are We?*, IDP. REVISTA DE INTERNET, DERECHO Y POLÍTICA? 16, 17 (2013), http://works.bepress.com/mireille_hildebrandt/52; Moerel & Prins, *supra* note 9, at 7. *See* ARTICLE 29 DATA PROTECTION WORKING PARTY, *supra* note 36, at 2. *See also* Rouvroy, *supra* note 10, at 25 (noting that purpose specification is "inimical to the whole philosophy of Big Data"). This notion was even stated by the Article 29 Working Party itself (even though it was ultimately rejected). *See also* van der Sloot & van Schendel, *supra* note 5, at 38–39; *Big data, artificial intelligence, machine learning and data protection*, UK INFO. COMMISSIONER'S OFF. 37–39 (2017), https://ico.org.uk/media/for-organisations/documents/2013559/big-data-ai-ml-and-data-

analyzing Big Data involves methods and usage patterns which neither the entity collecting the data nor the data subject considered or even imagined at the time of collection. To comply with the purpose specification rule, entities striving to engage in Big Data analysis will need to inform their data subjects of the future forms of processing they will engage in (which must still be legitimate by nature) and closely monitor their practices to assure they did not exceed the permitted realm of analyses. Carrying out any one of these tasks might prove costly, difficult and even impossible. Alternatively, those engaged in Big Data analysis must strive to tailor their practices to fall within the noted exceptions to this rule, as addressed below. Trying to circumvent this limitation by initially defining a very broad and vague purpose for future uses would most likely not resolve this matter, as the stated purposes must also be "specific."[47] Furthermore, stating an unnecessarily broad purpose might even be considered as "illegitimate" and thus lead to unacceptable processing.[48]

Purpose limitation is one of the cornerstones of the EU's data protection regime. It was featured in the DPD.[49] Yet more importantly, this concept is clearly noted in article 8(2) of the European Charter.[50] Given purpose limitation's enshrinement within the EU's primary legal source, the GDPR's drafters had no choice but to incorporate it in full within the Regulation. Any step short of that would have risked the invalidation of the GDPR by the European Court of Justice.[51]

Beyond tradition and the constitutional mandate, there are several substantial justifications for upholding and embracing the purpose limitation principle, even in the age of Big Data.[52] On a theoretical level, assuring that data controllers respect the purpose limitation principle will allow data subjects to exercise at least some control over their personal information—control being a central justification for EU data protection.[53] Additional

protection.pdf [hereinafter UK ICO Report].

47 ARTICLE 29 DATA PROTECTION WORKING PARTY, OPINION 03/2013 ON PURPOSE LIMITATION (WP 203), at 17, 52 (Apr. 2, 2013), http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2013/wp203_en.pdf.

48 *See, e.g.*, Viktor Mayer-Schönberger & Yann Padova, *Regime Change? Enabling Big Data through Europe's New Data Protection Regulation*, 17 COLUM. SCI. & TECH. L. REV. 315, 326 (2016), http://stlr.org/download/volumes/volume17/SchonbergerPadova.pdf.

49 DPD, *supra* note 4, art. 6(1)(b).

50 Charter of Fundamental Rights of the European Union art. 8(2), 2012 O.J. C 326/391 [hereinafter Charter of Rights].

51 For a similar outcome in a related context, see Digital Rights Ireland v. Comm'n, Joined Cases C-293/12 & C-594/12, ECLI:EU:C:2014:238.

52 Hildebrandt finds that purpose specification is closely linked to the central notion of legality and even that of the "rule of law." *See* Hildebrandt, *supra* note 15, at 23.

53 The notion of control is further linked to fundamental principles such as individual autonomy and self-fulfillment. See discussion in Zarsky, *Transparent Predictions*, *supra* note 20, at 1541–42.

instrumental justifications, recently stated by the Article 29 Working Party, are that abiding by this principle promotes trust in data environments,[54] as well as competition. This latter argument further states that purpose-specification rules weaken the hold monopolies have in data markets, while allowing start-ups to enter and compete.

These justifications are not beyond debate. Responding on the theoretical level, one might argue that in the digital age, users have objectively surrendered much of their control over personal data. The state's active intervention in providing individuals with rights they did not necessarily demand might amount to paternalism and undermine autonomy, and the nature of this basic right must therefore be questioned.[55]

On the instrumental level, it is possible to promote trust and limit abuses by closely monitoring data uses, rather than blocking ex ante analyses.[56] In addition, one might seriously question to what extent purpose limitation can indeed promote competition. Quite to the contrary, it might act as an inhibitor of competition, as it limits the abilities of start-ups to gather information on secondary markets and use it to enter new realms of business.[57] Instead, abiding by the "purpose specification" principle assures that only the monopolies that already have access to clients and their data can remain active in data rich markets, as these entities can eventually receive proper authorization from the data subjects to proceed with the data analysis. Given these counter-arguments, we must move forward and reconsider the validity of this key principle in the Big Data age.

Returning to the doctrinal analysis, it must be noted that the purpose limitation principle, as stated in the GDPR, includes a specific feature which, in theory at least, might enable Big Data analysis nonetheless to thrive: the notion of *compatibility*. If subsequent processing goes beyond the originally specified purpose yet is nonetheless compatible with it, such processing is permitted. However, even after accounting for the non-trivial maneuvering space this element provides, one can confidently assert that the GDPR substantially hampers Big Data initiatives.

The key provisions to understanding the notion of "compatibility" in this context are articles 5(1)(b) and 6(4) of the GDPR. Article 5(1)(b) states that processing for "statistical purposes" would not be considered incompatible. Therefore, if Big Data analytics will fall within this category,

---

[54] Hildebrandt explains that putting a brake on the re-usage of personal data, while also limiting its collection, is one way of preventing the datafication of everything and the threats it entails. *See* MIREILLE HILDEBRANDT, SMART TECHNOLOGIES AND THE END(S) OF LAW: NOVEL ENTANGLEMENTS OF LAW AND TECHNOLOGY 205 (2015).

[55] For a somewhat more detailed discussion of this argument, see Zarsky, *Transparent Predictions*, *supra* note 20, at 1541–45.

[56] Zarsky, *Desperately Seeking Solutions*, *supra* note 17, at 33.

[57] See discussion in Zarsky, *The Privacy–Innovation Conundrum*, *supra* note 7, at 136.

they could proceed. The extent of this exception is further detailed in article 89(1),[58] which states that "appropriate safeguards" must, nonetheless, be applied. Such safeguards, which must assure data minimization and might apply forms of pseudonymization, are not defined by the Regulation, yet might be set by the relevant Member States at a later time.[59]

The greatest challenge to relying on the "statistical purposes" exception to execute and enable the Big Data processes is noted at the very end of recital 162. This recital provides additional explanatory language regarding the meaning of this exception.[60] Here, the recital notes that the term "statistical purposes" implies that the results of such processing are not used "in support of measures of decisions regarding any particular natural person." Yet the Big Data practices which this Article chooses to address are specifically those that directly impact individuals, as they are used to provide them with unique and specific treatment. On its face, such uses are only permitted when meeting the strict purpose specification rule as detailed above—a difficult task in the Big Data environment.[61]

Beyond the "statistical process" exception, article 6(4) explains how compatibility could be established while applying various safeguards.[62] The listed criteria are an extension of the exceptions available under the DPD,[63] but are nonetheless somewhat abstract and of great difficulty to establish in the Big Data context. For instance, article 6(4)(b) calls for considering the context in which the data was collected—a notion counter to that of Big Data, which calls for analyzing data in different and distant contexts.[64] Article 6(4)(c) calls for considering the "nature of the personal data"—yet another factor that is constantly in flux when applying Big Data measures.[65] Finally, article 6(4)(e) calls for the use of possible safeguards such as pseydonymization—a measure which can substantially undermine the quality of the data and the insights it can provide given the loss of identifiable data which adds to the process's precision and accuracy.[66]

In conclusion, the "purpose specification" requirement clearly clashes

---

58 GDPR, *supra* note 1, art. 89(1).

59 *Id.* art. 89(2).

60 *Id.* art. 162.

61 For a similar analysis, see Rouvroy, *supra* note 10, at 26 (analyzing the provisions of Convention 108 and the "statistical processes" exception it features, which also includes similar reservations which limit instances which pose harm to a data subject). But see Mayer-Schönberger & Padova, *supra* note 48, at 329, for an opposing view, finding that the "statistical" exception could enable Big Data analysis.

62 GDPR, *supra* note 1, art. 6(4).

63 Moerel & Prins, *supra* note 9, at 52–54.

64 GDPR, *supra* note 1, art. 6(4)(b).

65 *Id.* art. 6(4)(c).

66 *Id.* art. 6(4)(e).

with the prospect of Big Data analysis. It appears that the GDPR's drafters have taken some measures to ease this tension, yet the current outcome is still insufficient and generates uncertainty. Furthermore, the safeguards offered are complex, difficult to execute, and might undermine the utility of the entire process. However, it is possible that member states would be able to take steps to indeed mitigate this concern to the extent possible given the EU Charter.[67] The UK's ICO Report has indicated, for instance, that a "fairness"-based test as to subsequent uses should be applied (while considering the data subject's expectations); unfortunately, this option might generate substantial uncertainty as well.[68] Another option would be applying this principle and the limitation it poses narrowly. This is advisable, as the analytical justifications for upholding this principle are shaky, at best.[69]

B. *Data Minimization*

The "Data Minimization" principle is yet another cornerstone of EU data protection policy. Unlike "Purpose Specification," this principle is not explicitly noted in the EU Charter, an omission which provides European legislators and regulators with greater flexibility when defining its outer limits.[70] The right is articulated in article 5(1)(c) stating that data must be "limited to what is necessary in relation to the purposes for which they are processed."[71] This principle appears at several other junctures throughout the Regulation, most notably in article 25, where it is stated as a requirement when designing relevant systems.[72] The minimization principle pertains to several dimensions: it relates to the scope and categories of data initially collected. In addition, it also refers to the limited duration during which personal data may be retained and the requirement that such data be deleted after its intended use.[73]

The justifications for data minimization are both intuitive and instrumental. When the minimization principle is followed, data controllers have fewer opportunities to undermine the data protection rights of data

67 Mayer-Schönberger & Padova, *supra* note 48, at 329.

68 UK ICO Report, *supra* note 46, at 37–39.

69 Note, for instance, that a report published by the Norwegian Data Protection Authority noted the clash and recommends that the notion of purpose limitation rules be applied narrowly. *See* DATATILSYNET, *supra* note 8, at 47.

70 *See* Charter of Rights, *supra* note 50, 2012 O.J. C 326/391.

71 GDPR, *supra* note 1, art. 5(1)(C).

72 *Id.* art. 25(1).

73 For a clear statement linking data minimization to the limited duration of storage, see *id.*, rec. 39 (explaining this principle requires "ensuring that the period for which the personal data are stored is limited to a strict minimum," and that "time limits should be established by the controller"). For a demonstration of the importance of this principle in the ECJ's jurisprudence, see Digital Rights Ireland Ltd v. Minister for Communications, Marine and Natural Resources, Case C-293/12, [2014] E.C.R. 238.

subjects. Indeed, with less data, data controllers will be unable to go beyond consented usage or violate their users' privacy in other ways. An additional justification can relate to the realm of cyber-security. The longer a data controller holds personal information (especially in large quantities), the greater the risk that such data would be hacked by both internal and external entities. The fact that data controllers do not have sufficient incentives to apply optimal cyber-security measures most likely enhances this risk of data leakage. Data minimization requirements can minimize this risk. In addition, on a theoretical level, the mere holding of personal data by the controller may undermine the data subject's autonomy and generate anxiety.[74] Data minimization lessens these concerns as well. Note that an overall response to these concerns could state that *ex post*-based regulations could protect data subjects from harmful uses and insufficient protection, punishing data controllers for unacceptable actions and omissions after the fact.

Data minimization requirements were prominently featured in the DPD as well.[75] However, the GDPR's enactment has somewhat expanded the reach of this principle.[76] The DPD noted that personal data must not be "excessive in relation to the purposes" collected or further processed.[77] The GDPR states that personal data should be "*limited to what is necessary*."[78] Clearly the language currently used calls upon data controllers to apply even greater scrutiny to their data practices and minimize the data at their disposal.[79]

The clash between the data minimization principle and the practices of Big Data analysis is intuitive, and was noted by several commentators as well.[80] The rush towards Big Data provides firms with a clear incentive to

---

[74] For more on these specific concerns from the U.S. perspective, see generally Daniel J. Solove, *Privacy and Power: Computer Databases and Metaphors for Information Privacy*, 53 STAN. L. REV. 1393 (2001).

[75] DPD, *supra* note 4, art. 6(1)(c).

[76] Detlev Gabel & Tim Hickman, *Data Protection Principles - Unlocking the EU General Data Protection Regulation*, WHITE & CASE (July 22, 2016), http://www.whitecase.com/publications/article/chapter-6-data-protection-principles-unlocking-eu-general-data-protection.

[77] DPD, *supra* note 4, art. 6(1)(c). This provision features an additional requirement that personal data be adequate and relevant, which has not changed in the GDPR.

[78] GDPR, *supra* note 1, art. 5(1)(c).

[79] *But see* Joris van Hoboken, *From Collection to Use in Privacy Regulation? A Forward-Looking Comparison of European and US Frameworks for Personal Data Processing*, *in* EXPLORING THE BOUNDARIES OF BIG DATA 231, 238 (Bart van der Sloot et al. eds., 2016) (stating that whether this change will strengthen privacy protection is up to the future interpretation of this concept).

[80] Bennet & Bayley, *supra* note 11, at 210 ("The business model of Big Data is antithetical to these Principles [of data minimization][.]"); Rouvroy, *supra* note 10, at 14. *See also* van der Sloot & van Schendel, *supra* note 5, at 38–39; DATATILSYNET, *supra* note 8, at

collect and retain as much data as they can for as long as possible (while accounting for the non-trivial costs of data collection and analysis). The ongoing improvements in data science and related fields might generate the belief that tomorrow holds great promise in what we might find while analyzing existing data, and therefore we certainly must not dispose of the data we hold today. In theory, at least, with more data will come greater knowledge and thus greater benefit to the firms and potentially society in general. On the other hand, diligently enforcing the "data minimization" principle will limit the success of "Big Data" initiatives while undermining their utility, with perhaps only limited justification.

Here as well the GDPR offers exceptions and potential loopholes, which might enable some limited yet insufficient Big Data analysis. In situations which meet the definition of a "statistical purpose,"[81] the GDPR concedes that data minimization could be achieved by pseudonymization; applying technological and statistical safeguards which will not allow for the identification of the data subjects. Yet as explained above,[82] this exception does not apply to many of the Big Data analyses addressed in this Article, given Big Data's subsequent effects on specific individuals. In addition, removing identifiers to achieve pseudonymity can potentially undermine the quality of the results derived, as the data would be purposefully altered and the aggregation of different datasets would be rendered difficult.[83] All these measures can subsequently limit the utility and benefits of the Big Data analyses.

To conclude, in the age of Big Data, the GDPR's data minimization requirements should be reconsidered, and perhaps somewhat loosened, as they undermine the success of potential Big Data initiatives. While the privacy and security concerns noted are substantial, they could probably be resolved through *ex post* regulation, which regulates unacceptable uses and abuses while still enabling the rich data analysis here discussed.[84] Thus the justifications for maintaining the strict data minimization rule are somewhat undermined. In the view of all the above, the logic of applying even stricter rules in the GDPR must be questioned.

42; Rubinstein, *supra* note 8, at 78 (predicting this requirement will often be breached).

[81] *See* GDPR, *supra* note 1, art. 89(1), rec. 156.

[82] *See supra* Section III(1), notes 58–61 and accompanying text.

[83] Paul Ohm, *Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization*, 57 UCLA L. REV. 1701, 1704 (2010).

[84] For a similar argument, see Rouvroy, *supra* note 10, at 17. The GDPR indeed provides enhanced measures to promote data security. GDPR, *supra* note 1, arts. 32–34.

C. *Special Categories*

A cornerstone of EU data protection policy is the creation of a layered regime, in which some forms of data categories and datasets are treated differently from others. In the DPD, article 8(1) prohibited the processing of 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," while providing narrow exceptions.[85] This distinction was embraced by the GDPR. The GDPR's article 9 prohibits the processing of similar "special categories," while adding genetic data, biometric data for the purpose of uniquely identifying a natural person, and data related to sexual orientation to the list of special categories.[86]

Processing of such information is still possible, subject to "explicit" consent or situations in which specific exceptions apply.[87] The former requirement itself might not substantially set the "special categories" apart from other forms of information, as obtaining consent for Big Data analysis is a challenging task in any event (i.e. even for data which is private, yet does not belong to a special category).[88] In addition, the GDPR provides a long list of general and specific "necessity"-based exceptions which allow, at times, for the processing of such sensitive data (especially in the health-related context).[89] Additional provisions permit Member States to set additional rules delineating specific exceptions.[90]

Again, the justification for setting this higher level of protection for special categories is intuitive. The categories noted constitute those which almost all individuals consider the most private.[91] The spreading or leaking of data from within these categories is bound to cause individuals the most distress, and can potentially generate the greatest harms.[92] Given these concerns, the GDPR took additional steps to enhance the protection of several categories by expanding their definitions, most notably in the context of "health." Here, recital 35 states that "health data" should include various factors such as "disease risk" and "medical history." An Article 29 Working

85 DPD, *supra* note 4, art. 8(1).

86 GDPR, *supra* note 1, art. 4(13–15) (providing elaborate definitions to the terms "genetic data," "biometric data," and "data concerning health," respectively).

87 *Id.* art. 9(2)(A).

88 Obtaining "consent" is one way for enabling lawful processing in general. *See* GDPR, *supra* note 1, art. 6(1)(a). Under the GDPR, "consent" is strictly defined in article 4(11) to require freely given, specific, informed, and unambiguous indication. It is unclear how this might prove different from the requirement for "explicit" consent noted in article 9. For a similar position, see Moerel & Prins, *supra* note 9, at 57.

89 GDPR, *supra* note 1, art. 9(2)(b–j), rec. 54.

90 *Id.* art. 9(4).

91 *See* Paul Ohm, *Sensitive Information*, 88 S. CAL. L. REV. 1125, 1169 (2015).

92 *Id.* at 1131 ("[S]ensitive information can lead to significant forms of harm.").

Party opinion addressing this matter found that the proceeds of novel forms of data collection (through the use of Internet of Things applications) and data analysis, which reveal health-related information, fall within this special category as well and should receive greater protection.[93]

Enter Big Data. These new forms of enhanced analytics challenge the ability to draw a distinction between "special" and other categories. Namely, an analysis merely relying on and addressing "regular" categories can quite quickly end up pertaining to "special categories."[94] For instance, health data can be deduced from a variety of datasets, such as shopping databases, and therefore this category has quickly and sharply expanded.[95] Thus, over time and given Big Data analysis, "special categories" mushroom in size. It was said in a different context that "we feel that all data is credit data, we just don't know how to use it yet."[96] The same statement could probably be made regarding almost all the other special categories noted. Therefore, the need to distinguish between the processing of "regular" and "special" categories encumbers Big Data processes that might inadvertently shift from one category to another, every one of which requires the application of a different set of legal rules.

Beyond encumbering the process, Big Data potentially undermines the entire distinction between these categories.[97] If nearly all forms of data categories and data sets can produce special data, why even bother with this distinction, which is rendered almost artificial? Note that, unlike the analysis in other segments of this Article, here the GDPR is not only impeding upon the ability to carry out big data analyses, but the availability of such analyses is undermining the most basic distinctions the GDPR sets forth.

The enhanced protection of specific categories could nonetheless be justified in the Big Data age, even if drawing out the actual distinctions proves impossible. Such protection, arguably, has symbolic value; through

---

[93] ARTICLE 29 DATA PROTECTION WORKING PARTY, ANNEX – HEALTH DATA IN APPS AND DEVICES, at 1 (Feb. 5, 2015), http://ec.europa.eu/justice/data-protection/article-29/documentation/other document/files/2015/20150205_letter_art29wp_ec_health _data_after_plenary_annex_en.pdf.

[94] *See* Ohm, *supra* note 91, at 1170; Lokke Moerel, *GDPR conundrums: Processing special categories of data*, IAPP (Sept. 12, 2016), https://iapp.org/news/a/gdpr-conundrums-processing-special-categories-of-data/.

[95] Rouvroy, *supra* note 10, at 27.

[96] This was famously noted by Dr. Douglas Merrill, former Google CTO, who went on to found ZestFinance. Quentin Hardy, *Just the Facts. Yes, All of Them*, N.Y. TIMES (Mar. 24, 2012), http://www.nytimes.com/2012/03/25/business/factuals-gil-elbaz-wants-to-gather-the-data-universe.html.

[97] For a similar argument, see Moerel & Prins, *supra* 9, at 57. Even without referring to Big Data, some have argued that rather than examining sensitive categories, we must focus on sensitive contexts. Given the limitations of this specific project, this broader critique is not discussed in this Article. *See* Ohm, *supra* note 91, at 1145.

these rules, the law provides a clear signal that specific forms of data can generate substantial harm and therefore must be treated with greater care. On the other hand, the age of Big Data might be further undermining the justifications for introducing these distinctions, even on the symbolic level. According to Antoinette Rouvroy, there is a substantial difference between previous forms of discrimination along the lines of sensitive factors, and those unfolding in the Big Data age.[98] In this new era, newer forms of discrimination do not necessarily result from discriminatory intent, which the singling out of special categories aims to protect both actually and symbolically.[99] Rather, discrimination carried out today is data driven, often does not involve intent, and is not split along the simple clear lines of the noted special categories.[100]

In addition, the practice of establishing discriminatory factors is unstable and even unpredictable, as their negative impact might gradually grow, and the effect of their analysis is compounded over time.[101] Thus, it must be approached using newer tools, both theoretical and doctrinal. Therefore, even the symbolic emphasis on sensitive factors might be misplaced and unnecessary.

In conclusion, the rise of Big Data substantially undermines the logic and utility of applying a separate and expansive legal regime to "special categories" for various reasons. For starters, there are practical considerations. Attending to this distinction generates extensive and unnecessary regulatory costs. Regulators on both the continental and domestic level will be required to ponder over the question as to whether various datasets and analyses fall within the special categories noted. Courts will also need to weigh in on this unimportant question—which will be costly to all the parties involved. Beyond costs, the current legal regime will generate substantial uncertainty, which will again impede on firms striving to engage in Big Data analysis, with smaller entities suffering the greatest harms given their inability to seek out costly legal advice. Finally, given the symbolic justification to maintain the distinction here discussed between sensitive and other forms of data, it is important to emphasize other symbolic reasons to abandon the use of special categories. If almost all data might fall under the "special" category, the signal and message this regulatory framework provides regarding the higher level of privacy due to special categories is subsequently diluted. At the end of the day, there will be no real special treatment for these special categories as this stricter standard will

98 *See also* Rouvroy, *supra* note 10, at 16–17.

99 *Id.*

100 *Id.*

101 I thank Courtney Bowman for pointing out this observation. See Moerel, *supra* note 94, for "sensitive" in different contexts.

be applied across the board. This is not necessarily a good thing. With time and given practical needs, even information within this special category will be processed regularly. Thus, the entire aim of this regulatory regime which strives to provide special categories with special treatment is defeated and rendered meaningless—and the public will view all these forms of data as similar.[102] In view of all the above, the impact Big Data will have on this regulatory aspect must be accounted for at once. Even if some form of particular regime for special categories is to be considered, it must be applied narrowly.[103]

D. *Automated Decisions*

The GDPR's article 22 sets forth a specific legal rule governing decision-making processes, which are both fully automated and substantially impact individuals, such as credit applications or recruiting.[104] In this unique provision, EU law provides the individual with the right not to be subjected to these processes. A similar general rule that specifically singles out such processes does not exist in American law.[105]

This general rule includes several exceptions which nonetheless allow for such automated analyses to proceed—such as the data subject's explicit consent.[106] In addition, member states can authorize such a process, and it will also be permitted if deemed necessary to enter or perform a contract.[107] Such authorization and finding of necessity would most likely pertain to automated processes used to detect fraudulent activity, especially in financial contexts.[108] It is unclear whether additional exceptions will be provided to other categories of analysis.

Even when the noted exceptions apply, the data subject is provided with

---

[102] Moerel & Prins, *supra* note 9, at 11 (arguing that this distinction is no longer "meaningful").

[103] For a recommendation to abolish these categories in the GDPR, see Moerel, *supra* note 94.

[104] GDPR, *supra* note 1, art. 22, rec. 71.

[105] Possible exceptions are U.S. credit reporting laws which require at some junctures that individuals be informed of the main factors affecting their credit score. *See* 15 U.S.C. § 1681g(f)(1)(C) (2012). See also the discussion in Danielle Keats Citron & Frank Pasquale, *The Scored Society: Due Process for Automated Predictions*, 89 WASH. L. REV. 1, 17 (2014).

[106] GDPR, *supra* note 1, art. 22(2)(c).

[107] *Id.* art. 22(3). Article 22(4) further stipulates that the exceptions cannot be premised on the "special categories" noted in article 9(1) (and addressed in Part II(3) above). Given the noted expansion of these categories in the big data age, this might further inhibit automated processes. However, the GDPR still provides for a loophole, allowing for processing which meets the requirements of article 9(2)(a) or (g), which provide for the processing of such categories subject to explicit consent, or when there is a substantial public interest, subject to some additional safeguards. These loopholes are quite broad and therefore article 22(4) will probably not prove to be a substantial obstacle.

[108] GDPR, *supra* note 1, rec. 71.

several important rights when facing automated decisions: he or she has a right to "obtain human intervention" as well as the right to "contest the decision"[109] (and express his or her point of view). Furthermore, the GDPR provides data subjects with the right to access background data to enable such a contest. In several other instances, and while addressing the right of "access" throughout the Regulation, the GDPR clearly indicates[110] that when these automated processes unfold, the data subject must be informed of their existence, as well as provided with "meaningful information about the logic involved, as well as the significance and the envisaged consequences of such processing for the data subject."[111]

The requirements article 22 sets forth could nonetheless be sidestepped relatively easily by inserting human intervention into the process. In other words, once the process is not "solely" automated, this provision will not apply.[112] It is of course too early to establish how courts and regulators will rule on issues related to this matter, and whether they might render some forms of limited human intervention fictitious and negligible, thus nonetheless subjecting article 22 and its requirements to various processes.

The GDPR's drafters did not pull article 22 out of thin air. Indeed, a very similar rule was set forth in the DPD.[113] However, this rule was rarely applied, and in many member states was in fact a dead letter.[114] Furthermore, recent rulings in Germany further limited the rights the predecessors of article 22 provide by finding that firms can refrain from disclosing insights as to their automated internal processes, given their interest in protecting their trade secrets.[115]

Several justifications could be provided to explain article 22 and its predecessors. First, one can link this right to the notion of honor and respect; when faced with crucial decisions, a human should be treated with the

---

109 *Id.* art. 22(3).

110 *Id.* arts. 13(12)(f), 14(12)(g) & 15(1)(h).

111 *Id.* art. 13(2)(f). Recital 71 goes further to note that an individual has the right to receive an "explanation" after the fact as to how the decision was reached. Yet some scholars have questioned whether this statement is sufficiently backed by specific language in the actual articles of the Regulation. *See* Wachter et al., *supra* note 9, at 11.

112 Note that one of the versions considered for article 22 called for applying it in situations which were "predominantly" automated—a version eventually abandoned during the legislative process. *See* Wachter et al., *supra* note 9, at 32.

113 DPD, *supra* note 4, arts. 12(a), 15; Wachter et al., *supra* note 9, at 11 (explaining that the rules regarding this matter set out in the GDPR are very similar to those of the DPD).

114 Zarsky, *Transparent Predictions*, *supra* note 20, at 1551; Douwe Korff, *Data Protection Laws in the EU: The Difficulties in Meeting the Challenges Posed by Global Social and Technical Developments* 86 (European Comm'n Directorate-Gen. Justice, Freedom & Sec., Working Paper No. 2, 2010), http://ec.europa.eu/justice/policies/privacy/docs/studies/new_privacy_challenges/final_report_working_paper_2_en.pdf.

115 *See* Wachter et al., *supra* note 9, at 23.

dignity of having a human decision-maker address his or her personal matter.[116] Second, the fact that these automated processes unfold without providing sufficient insights to those affected by it undermines the right to "due process."[117] Similarly, carrying out automated processes without sufficient supervision generates concerns that such processes are error-ridden, discriminatory, tainted, and flawed.[118] On a deeper level, this rule might reflect the distrust humans have towards computerized systems and machines generally—admittedly, a notion that might be no longer valid.

Article 22 (and the "access" rights which pertain to the process it addresses) directly impacts Big Data practices. Furthermore, in a recent report, Rouvroy admits that achieving the article's objective in the Big Data age is "both unrealistic and deeply paradoxical."[119] The tension between Big Data and article 22 unfolds on several levels. First, prohibiting automated analysis obviously undermines many of the Big Data practices discussed throughout this Article. Second, even if one of the many exceptions to the prohibition on automation (such as limited human intervention) is met, the specific disclosures noted above which call for enabling a human response to the machines' decisions are still required. To meet these disclosure obligations, Big Data processes must be conducted in a manner that would assure they are interpretable—i.e., they can be explained to the inquiring individual.[120] Constantly meeting an "interpretability" requirement might call upon those designing the automated processes to compromise some of the system's precision to enable the delivery of this form of detailed disclosure.[121] Third, allowing human interjection would further encumber the automated process and slow down the innovative technologies they bring about.

Article 22 is perhaps the most salient example of the GDPR's rejection of the Big Data revolution. It signals a deep distrust towards these automated processes, yet does not specify why this attitude was adopted. In response to this rule, it is possible that firms will indeed be required to substantially change their technological architectures and even business models, opting for less efficient practices which comply with this rule.

However, it is very likely that of all the GDPR's provisions discussed

---

116 Zarsky, *Transparent Predictions*, *supra* note 20, at 1551–52.

117 HILDEBRANDT, *supra* note 54, at 198.

118 These justifications could be derived from GDPR, *supra* note 1, rec. 71. For a recent discussion of the flaws of automated and automatic processes, see generally Andrea Roth, *Trial by Machine*, 104 GEO. L.J. 1245 (2016).

119 Rouvroy, *supra* note 10, at 34.

120 Zarsky, *Transparent Predictions*, *supra* note 20, at 1519.

121 Future research must establish the extent of the trade-off between interpretability and data mining efficiency. If the costs of meeting interpretability are low, this critique is somewhat curbed.

thus far, article 22 will have the least effect in practice. As noted, the provision could be sidestepped by inserting very minimal human interaction. In addition, the explanations set forth to meet legal requirements might be of very limited scope, while providing very general language. Still, the powerful anti-Big Data gesture article 22 conveys will remain, and might be projected in regulatory decisions made in other contexts.

IV. CONCLUSION: WHAT'S NEXT FOR EUROPE?

Part III illustrated a bleak outlook regarding the ability to engage in effective Big Data analysis in an EU governed by the GDPR. It further addressed other negative effects these technologies will have on the GDPR's coherence. In this Part, the Article will briefly examine what the impact of the noted conclusions might be and what the existence of the GDPR's current format might actually lead to. In what follows, this Part will address three possible outcomes for Europe: one optimistic, one pessimistic, and in between them, a realistic forecast and point of view.

Let us begin with some European optimism. Here one might argue that the GDPR will march the citizens of Europe and even those of the entire world to a new era in which new forms of data analytics will dominate the markets—those that provide sufficient safeguards to protect fundamental rights and adhere to the other rules addressed in the Regulation. When addressing this matter, Mireille Hildebrandt explains that EU data protection law might allow citizens to have their cake and eat it too; they will benefit from enhanced data protection, while enjoying the innovations enhanced data analytics bring about.[122] Even if this preferable outcome would not be achievable, a balance between the interests will be struck to assure that only a proportionate breach of rights would occur.

These optimistic dynamics need not be limited to the EU. Given the GDPR's international jurisdiction, this balanced outcome will migrate to other jurisdictions beyond the EU as well, in a process Anu Bradford refers to as the "Brussels Effect." Thus, the change the GDPR brings about will even affect global firms operating domestically in the United States, and thus American consumers and citizens.[123] Even though the GDPR does not apply in the latter instance, multi-national firms will nonetheless comply with these norms in their domestic settings. Given the high costs of complying with multiple regulatory regimes in the digital age, global firms might merely opt for complying with one regulatory model everywhere—and this would prove to be the stricter European model. They might also choose to apply GDPR-like rules to U.S. citizens due to fear of the public and political backlash that

122 HILDEBRANDT, *supra* note 54, at 211.

123 Anu Bradford, *The Brussels Effect*, 107 NW. U. L. REV. 1, 64 (2012).

would follow disclosures that U.S. firms provide foreign customers with greater protection than American customers.

This rosy prediction should be met with skepticism and pessimism. At first, the noted global-political dynamic must be questioned. Recent political trends of national separatism will most likely allow firms to justify different treatment in different countries. Thus, the actions of U.S. businesses pushing back EU regulators will not be frowned upon by either the government or substantial segments of the public. The same could most likely be said of the U.K. Furthermore, even within the EU, there is no guarantee the public would be able to have their cake and eat it, too (as Mireille Hildebrandt argued). Given the difficulties in carrying out Big Data analytics, local entrepreneurs will focus their innovative spirits elsewhere, or move to other countries where they could pursue their objectives and Big Data-related business models without disturbance. Such outcomes will be harmful to the EU, as it will lead to the loss of income and jobs. Furthermore, and at the end of the day, privacy interests might not be protected after all, as local EU residents might nonetheless turn to firms overseas to use their data analytic products and services. The firms operating in these latter instances might indeed still be subjected to the GDPR's jurisdiction, but enforcement of these situations will surely prove challenging.[124] It is obviously difficult to predict which of these opposing scenarios will unfold, yet it is very likely both will occur in varying contexts.

Between these two somewhat extreme predictions, a third, pragmatic forecast has been set forth. It states that the GDPR provides sufficient forms of exceptions and loopholes to allow rich Big Data dynamics to nonetheless unfold. For instance,[125] Mayer-Schönberger and Padova argue that the GDPR can facilitate Big Data analysis, and that its introduction constitutes progress from the rules set out by the DPD. They reach this conclusion after analyzing the same provisions discussed above, while highlighting the various (noted) instances in which member states are permitted to introduce exceptions which might enable some form of analyses. They are also optimistic that at least some such states will indeed move ahead and introduce such exceptions.

I certainly hope this noted scenario will indeed reflect reality in the near future, but will conclude by adding substantial skepticism towards this view as well. While exceptions exist, the noted analysis outlined in the previous sections demonstrates that the GDPR's overall narrative is quite clear—it views Big Data analysis as highly problematic. It reintroduces article 22 which directly clashes with the very foundations of the practices of Big Data

[124] For a similar analysis, see Zarsky, *The Privacy–Innovation Conundrum*, *supra* note 7, at 161.

[125] Mayer-Schönberger & Padova, *supra* note 48, at 315.

analysis; it continues to apply a layered approach to data protection (with "special" categories) which Big Data clearly renders unworkable. In view of these two clear messages, local regulators will be reluctant to expand exceptions pertaining to other rules—mainly data minimization and purpose specification—which encumber the Big Data processes. It is therefore fair to assume that the analysis set forth in Part III will hold, and that the pessimistic prediction noted above will eventually materialize.

To conclude, it is very difficult to predict what the GDPR's actual impact on Big Data analytics will prove to be, and vice versa. Many factors—such as the effectiveness of the EU enforcement and fine mechanisms as well as the benefits the Big Data technologies will bring about—are still unclear. Yet, the scenario that the GDPR's incompatibility will lead to an impact that would be both negative and substantial must be taken under serious consideration. While the EU's strong position towards the protection of privacy rights is admirable, it is possible that the full implications the GDPR will have for the important Big Data practices, and their benefits, have not been fully and properly considered. Therefore, the opinions here noted must be kept in mind as this new Regulation moves towards enactment and implementation.