ENABLING BIG DATA DESPITE GDPR SUBSTANTIVE UNCERTAINTY: COMPLIANCE PROGRAMS AND ARTICLE 25

Beginning May 25, 2018, all companies processing personal data of European data subjects must comply with the European Union's General Data Protection Regulation ("GDPR") or face stiff penalties. Much has been said about the GDPR's paradigm-shift in data protection rules, with special attention paid to how the Regulation will impact the burgeoning field of big data analytics. Some commentators assert that the GDPR will destroy such analysis; others argue that Big Data will flourish under the law. Regardless of the policy debate, companies must nevertheless follow the GDPR—and they are scrambling to do so.[1] Accordingly, these companies must devise systems and procedures to minimize their risk of infringement and liability under the GDPR.

Complicating compliance efforts is uncertainty about GDPR's application to big data analytics. The GDPR lays out many requirements for using personal data, ranging from strict principles to limitations on automated processing. Depending on their implementation, these principles could smother big data analytics with administrative limitations. Many commentators explore how the GDPR implementations might balance protecting individual privacy and benefiting from these new technologies.

Despite the importance of reconciling privacy rights and Big Data, another practical question remains: how do companies continue to employ big data analytics when faced with significant uncertainty regarding GDPR's substance? Companies found unlawfully processing data can be fined to the tune of 20,000,000 euros or 4% of annual global turnover—whichever is higher.[2] As explained below, the GDPR provides two mechanisms for companies to limit and potentially avoid liability for unlawful data processing: making good-faith efforts to adhere to the GDPR and adopting data protection by design and by default as Article 25 mandates. To demonstrate good-faith efforts and to comply with Article 25, companies must design, implement, and enforce strong procedural mechanisms—internal controls—to get to the right outcomes. But, what internal controls should companies implement, and how can regulators and companies know whether these controls are effective?

This Essay offers a framework for that last question. Part II highlights the tension between Big Data and the GDPR. Part III argues that an effective compliance program, including requirements under Article 25, limits potential liability under the GDPR. Because controllers must demonstrate good-faith efforts to comply with the Regulation to receive leniency, this section further argues regulators and controllers should look to the U.S. Federal Sentencing Guideline for Corporations to determine what is expected for good-faith efforts. Part IV assesses, in broad strokes, what the seven elements of the Guidelines might require of a controller striving to employ big data analytics under the GDPR. Part V concludes.

---

[1] According to some surveys, 93% of companies have GDPR compliance as their top legal demand. Wei Chieh Lim, *Will Data Protection Laws Kill Artificial Intelligence?*, CPO MAG. (Aug. 17, 2017), https://www.cpomagazine.com/2017/08/17/will-data-protection-laws-kill-artificial-intelligence/2/.

[2] *See* 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 "GDPR"), art. 83(5).

II. THE GDPR AND BIG DATA: FRIENDS, ENEMIES, OR FRENEMIES?

Does the GDPR permit effective and profitable big data practices? No consensus exists. While clear the Regulation applies to these technologies, questions remain about how enforcement of its provisions allow Big Data to achieve its purpose. The following discussion summarizes key provisions: Article 5(1)(b) (purpose limitation), Article 5(1)(c) (data minimization), and Article 22 (limiting automated decision-making).[3]

Article 5 sets out the Principles data controllers must follow when processing personal data.[4] One such principle is purpose limitation, codified as Article 5(1)(b). Purposes limitation obliges controllers to collect data only for "specified, explicit and legitimate purposes" and prohibits further processing of collected data "in a manner that is incompatible with those purposes."[5] Some criticize purpose limitations as undercutting the economic value and innovation from combining data sets to use in subsequent analyses.[6] Indeed, Big Data's promise comes from the "four Vs: the Volume of data collected, the Variety of sources, the Velocity [of] the analysis . . . and the Veracity of the data."[7] On the other hand, regulators have routinely disagreed with that assessment. Interpreting the same language under the GDPR's predecessor, the Article 29 Working Party observed "legislators intended to give some flexibility with regard to further use. Such further use may fit closely with the initial purpose or be different."[8] The United Kingdom's Information Commissioner's Office says that future processing is permissible so long as "it is fair."[9]

Article 5 further sets out a data minimization requirement. Specifically, Article 5(1)(c) demands personal data be "adequate, relevant and limited to what is necessary in relation to the purposes for which they are processed."[10] In claiming the GDPR smothers Big Data, critics point to the essence of such analyses: extracting unseen patterns from large data sets.[11] Indeed, if the "relevant" data points were already known, big data analytics would not offer such groundbreaking insights.[12] Proponents in turn argue that the Principle does not prevent companies from collecting lots of data; it only prevents

---

[3] Other rights and obligations bear on machine learning, such as the GDPR's elevation of special categories of data, the right to explanation, and the right to be forgotten.

[4] GDPR, art. 5.

[5] GDPR, art. 5(1)(b). Several exceptions exist, including one for statistical analysis. *See* GDPR, art. 5(1)(b) ("[F]urther processing for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes shall, in accordance with Article 89(1), not be considered to be incompatible with the initial purposes."). Article 89(1) states further processing "shall be subject to appropriate safeguards, in accordance with this Regulation, for the rights and freedoms of the data subject. Those safeguards shall ensure that technical and organisational measures are in place." However, Recital 162 states statistical analysis cannot be "used in support of measures or decisions regarding any particular natural person." Even though recitals are not binding, they call into question whether using machine learning algorithms fall within the "statistical purposes" exception.

[6] *See, e.g.*, Unlocking the Value of Personal Data: From Collection to Usage, WORLD ECON. FORUM 7 (Feb. 2013), http://www3.weforum.org/docs/WEF_IT_UnlockingValuePersonalData_CollectionUsage_Report_2013.pdf; *see also* Tal Z. Zarksy, *Incompatible: The GDPR in the Age of Big Data*, 47 SETON HALL L. R. 995, 1005 (2017).

[7] Zarsky, *GDPR in the Age of Big Data*, at 998–99.

[8] Opinion 03/2013 on Purpose Limitation, Article 29 Data Protection Working Party 21 (Apr. 2, 2013).

[9] Big Data, Artificial Intelligence, Machine Learning and Data Protection, U.K. INFO. COMM'NS OFFICE 38 (Sept. 04, 2017).

[10] GDPR, art. 5(1)(c).

[11] Zarsky, *GDPR in the Age of Big Data*, at 1010–11.

[12] For examples and an overview, see generally Viktor Mayer-Schönberger & Yann Padova, *Regime Change? Enabling Big Data Through Europe's Data Protection Regulation*, 17 COLUM. SCI. & TECH. L. R. 315 (2016).

them from collecting irrelevant *personal* data.[13] Moreover, companies adopting other technical approaches to anonymize data can escape these obligations.

Article 22 limits automated decision-making. It grants data subjects a right to avoid automated decision-making (including profiling) that produces legal effects or other significant effects.[14] Despite including several exceptions, Article 22 is viewed as a "rejection of the Big Data revolution" because its exceptions require explicit consent and an understanding of the underlying decision, the latter of which is difficult, if not impossible, to create at the moment.[15] According to Antoinette Rouvroy, a member of the European Data Protection Supervisor's Ethics Advisory Group, the goal Article 22 embodies is "both unrealistic and deeply paradoxical."[16] On the other hand, researchers view Article 22 as an opportunity to address challenging problems that might otherwise go unaddressed.[17]

As this Part explored, several provisions of the GDPR have the potential to unnecessarily burden big data analyses. The burden will turn, in large part, on how regulators enforce the provisions and what technologies and techniques are created in response. In this Essay, these ambiguities are termed substantive uncertainties. Despite substantive uncertainties, companies should not rest on their laurels; doing so risks memorializing the company for all the wrong reasons.[18] Instead, companies should strive—in good faith—to comply. Not only will doing so reduce the chances of processing information unlawfully, it will also mitigate potential sanctions for any errors.

III. GDPR LENIENCY: EFFECTIVE COMPLIANCE PROGRAMS AND ARTICLE 25

Despite authorizing astronomic penalties, the GDPR acknowledges that not all unlawful processing should be prosecuted to the fullest extent. Indeed, the GDPR demands regulators consider several factors in deciding an appropriate fine.[19] These factors and their associated guidance suggest that good-faith efforts to comply and implementing data protection by design and by default might inoculate a company. One mechanism to demonstrate good-faith efforts and to implement Article 25 is by designing and enforcing an effective compliance program. Simply put, to handle substantive uncertainty, the GDPR relies on procedural mechanisms to identify conduct worthy of sanctions.

Good-faith efforts to comply bear on intentionality and negligence and demonstrate serious contemplation of big data analytics' risks. Common sense suggests intentionally- and negligently-unlawful activity should be punished more harshly than unintentionally-unlawful activity; Article 83(2)(b) incorporates that common sense into the GPDR.[20] According to the Article 29 Working

---

[13] *Big Data, Artificial Intelligence, Machine Learning and Data Protection*, at 40–41; *see also* Ann Cavoukian, David Stewart, and Beth Dewitt, *Using* Privacy by Design *to Achieve Big Data Innovation Without Compromising Privacy*, INFO. AND PRIV. COMM'N OF ONTARIO 16 (June 10, 2014).

[14] GDPR, art. 22(1).

[15] Zarsky, *GDPR in the Age of Big Data*, at 1017.

[16] 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. 11 (Jan. 11, 2016).

[17] *See* Bryce Goodman and Seth Flaxman, *European Union Regulations on Algorithmic Decision-Making and a "Right to Explanation,"* 2016 ICML Workshop on Human Interpretability in Machine Learning 5 (Aug. 31, 2016), https://arxiv.org/abs/1606.08813.

[18] For example, LabMD was the defendant in the case that established inadequate security practices can be "unfair" under Section 5 of the Federal Trade Commission Act. Thus, "LabMD" will forever conjure thoughts of poor security.

[19] *See* GDPR, art. 83(2).

[20] *See* GDPR, art. 83(2)(b) ("When deciding whether to impose an administrative fine and deciding on the amount of the administrative fine . . . due regard shall be given to . . . the intentional or negligent character of the infringement."); *see also*

Party, regulators shall look for "objective elements of conduct" when deciding whether intentional misconduct or negligence occurred.[21] These objective elements might include "unlawful processing authori[z]ed explicitly by the top management hierarchy . . . in disregard for existing policies[,] . . . failure to read and abide by existing policies, human error, failure to check for personal data in information published, failure to apply technical updates in a timely manner, [and] failure to adopt policies."[22] Each of these elements can be addressed by an effective compliance program. Moreover, companies who demonstrably take their obligations under the GPDR seriously will likely receive leniency. Failures that "demonstrat[e] contempt for the provisions of the law" are more likely to be fined.[23] Regulators must also give "due regard [to] . . . the degree of responsibility of the controller" while accounting for obligations imposed under Article 25.[24] In other words, has a company "d[one] what it could be expected to do give the nature, the purposes or the size of the processing"?[25] Designing and enforcing an effective compliance program signal genuine respect for the purposes of the law, and ensure companies have done what they can.

The GDPR further demands controllers implement data protection by design and by default. Notably, data protection by design insists controllers consider data protection through the whole product lifecycle—from ideation to decommission—and implement "appropriate technical and organi[z]ational measures [that are] designed to implement data-protection principles . . . in an effective manner.[26] In determining appropriate measures, controllers must consider the "state of the art, the cost of implementation and the nature, scope, context and purposes of processing as well as the risks of varying likelihood and severity for rights and freedoms of natural persons posed by the processing."[27] Data protection by design simply prescribes a process: controllers must approach data protection from the very beginning.[28] The adequacy of a data protection by design program is another factor regulators consider when determining penalties.[29]

As covered above, the GDPR implicitly—as is the case with good-faith compliance—and explicitly—as is the case with Article 25—endorses compliance programs as a way to limit or avoid penalties for unlawful activity entirely. A parallel of this system is seen in the U.S. Federal Sentencing Guidelines for Corporations ("Sentencing Guidelines").[30] Like the GDPR, the Sentencing Guidelines

---

Guidelines on the Application and Setting of Administrative Fines for the Purposes of the Regulation 2016/679, ART. 29 DATA PROTECTION WORKING PARTY 12 (Oct. 3, 2017) ("It is generally admitted that intentional breaches, demonstrating contempt for the provisions of the law, are more severe than unintentional ones and therefore may be more likely to warrant the application of an administrative fine.") (henceforth "Penalty Guidance").

[21] Penalty Guidance, at 12.

[22] *Id.*

[23] *See* Penalty Guidance, at 12.

[24] GDPR., art. 83(2)(d).

[25] Penalty Guidance, at 13.

[26] GDPR, art. 25(1).

[27] *Id.*

[28] *See* Penalty Guidance, at 13. ("Rather than being an obligation of goal, these provisions introduce *obligations of means*, that is, the controller must make the necessary assessments and reach the appropriate conclusions."(emphasis added)).

[29] *See id.*; *see also* GDPR, art. 83(2)(d).

[30] *See* Chapter 8 — Sentencing of Organizations, Guidelines Manual, U.S. SENTENCING COMMISSION (2016), https://www.ussc.gov/guidelines/2016-guidelines-manual/2016-chapter-8 (henceforth "Sentencing Guidelines"). Because companies can be convicted under U.S. law, the Sentencing Guidelines summarize mitigating factors in determining appropriate fines after conviction.

offer reduced fines, and perhaps amnesty, for companies who adopt effective compliance programs.[31] Unlike the GDPR, however, the Sentencing Guidelines also offer seven elements that comprise an effective compliance program. To ensure they benefit from the GDPR's limited liability mechanism, controllers dealing with substantive uncertainty should look to the Sentencing Guideline as a framework for developing an effective compliance system.[32] The next Part explores what such a compliance system might entail.

IV. SENTENCING GUIDELINE'S SEVEN ELEMENTS OF AN EFFECTIVE COMPLIANCE PROGRAM

The primary goal of a compliance program is to prevent and detect unlawful conduct; as mentioned above, a GDPR compliance program also hopes to prevent (and detect) objective elements of misconduct. However, a compliance program means more than having processes in place; it requires promoting an organizational culture of lawful and ethical compliance.[33] The following discussion explores the seven elements in the Sentencing Guidelines and pontificates on what internal controls satisfying each element might look like.

The first element of an effective compliance program requires establishing standards and procedures to prevent *and detect* unlawful activity.[34] With the GDPR, this might involve establishing and enforcing procedures to ensure appropriate balancing tests are conducted and adequate measures are implemented. For example, one of IBM's preventative controls is conducting a data protection impact assessment ("DPIA") for every product, offering, and service.[35] Moreover, effective compliance systems require detective controls. One such detective control might be reviewing DPIAs and any measures implemented to protect fundamental rights. If the balancing was conducted inaccurately or measures were inadequate, unlawful processing might occur; reviewing detects these dangers. Another preventative control might be writing Article 25 requirements into vendor contracts. To ensure all vendors are bound by such provisions, OCR-enabled contract analysis, offered by companies such as Evisort, can validate the contracts. Finally, these procedures should create a feedback loop so the company responds to new caselaw and innovations. Policies for continuing education and employee participation in industry groups will keep controllers abreast of recent developments.

---

[31] *See generally*, *id.*; *see also* Memorandum from Paul J. McNulty, Deputy Attorney General, U.S. DEP. OF JUSTICE, at 4. ("In . . . determining whether to bring charges, . . . prosecutors must consider . . . the existence and adequacy of the corporation's pre-existing compliance program." (emphasis in original)) (henceforth "McNulty Memorandum").

[32] It is worth noting that other frameworks for effective compliance systems exist, such as ISO 19600:2014, https://www.iso.org/obp/ui/#iso:std:iso:19600:ed-1:v1:en.

[33] Sentencing Guidelines, § 8B2.1(a) ("To have an effective compliance and ethics program, . . . an organization shall . . . promote an organizational culture that encourages ethical conduct and a commitment to compliance with the law."); *see also* McNulty Memorandum, at 14 ("[T]he critical factors in evaluating any program are . . . whether corporate management is enforcing the program or is tacitly encouraging or pressuring employees to engage in misconduct to achieve business objectives."); Speech of Elizabeth Denham, Commissioner of the U.K. Information Commissioner Office, given at the Data Protection Practitioner's Conference 2017 (Mar. 06, 2017), https://ico.org.uk/about-the-ico/news-and-events/news-and-blogs/2017/03/data-protection-practitioners-conference-2017/ ("It's about a framework that should be used to *build a culture of privacy that pervades an entire organisation*. It goes back to that idea of doing more than being a technician, and seeing the broader responsibility and impact of your work in your organisation on society." (emphasis added)).

[34] Sentencing Guidelines, § 8B2(b)(1).

[35] Tom Macaulay, *How IBM is Preparing for GDPR*, COMPUTERWORLDUK (Nov. 28, 2017), https://www.computerworlduk.com/data/how-ibm-is-preparing-for-general-data-protection-regulation-gdpr-3668249/.

To satisfy the second element of an effective compliance program, the program must ensure information about and details of the compliance program flow between employees implementing it day-to-day and governing authorities, such as boards.[36] To ensure adequate information dissemination, DPOs should design a formal reporting structure, whereby front-line employees report on compliance successes and challenges. These reports must percolate to the governing authority through regular updates. To ensure governing authorities have sufficient understanding of their compliance program, they should be updated by DPOs, can be briefed by outside counsel, or can partake in director education programs.[37]

The third element requires companies use reasonable effort to avoid giving authority over the program to individuals who have a history of unlawful conduct or "other conduct inconsistent with an effective compliance . . . program."[38] Simply put, foxes should not guard hen houses. Because data practices vary widely across the world, and because Big Data analytics might use technology from many different parties, this element is best understood as requiring controllers to be diligent in selecting vendors. It also applies to people. For example, controllers should avoid giving risky individuals, such as those who intentionally violated the GDPR at another corporation, responsibility over their compliance program. Alternatively, a compliance program may need to screen vendors (and vendors' vendors) for past GDPR violations or questionable data protection histories.

The fourth element requires periodic and practical communication of the organization's standards and procedures through "effective training programs" and "otherwise disseminating information" appropriate to the specific role.[39] At a minimum, controllers should adequately train employees on company policies to comply with the GDPR.

The fifth element requires taking reasonable steps to ensure the program is followed, its efficacy is regularly evaluated, and it includes a channel for resolving uncertainties and reporting violations anonymously.[40] Standard compliance techniques are helpful here, such as hiring outside counsel or consultants to review and certify the program, auditing databases and paper records to monitory policy compliance, and implementing anonymous hotlines.

The sixth element requires the program be promoted and enforced through rewards and punishments.[41] Basically, employees must be rewarded for good compliance and punished for unlawful activity. When employees are in a position to prevent unlawful activity, they must be punished for failing to take reasonable steps to prevent or detect it.

The seventh and final element demands a compliance program act like an algorithm; that is, change in response to feedback. Specifically, it must adapt to detected unlawful activities with the goal of preventing similar conduct in the future.[42] As part of their policies and standards, controllers should

---

[36] Sentencing Guidelines, § 8B2.1(b)(2)(A)–(C).
[37] Guidelines from the National Association of Corporate Director might prove useful in developing basic understanding of the relevant information and Director responsibility. *See e.g.*, Corey E. Thomas, *The Corporate Director's Guide to GDPR*, NAT. ASS'N. OF CORPORATE DIRECTORS (Aug. 15, 2017), https://blog.nacdonline.org/2017/08/directors-guide-to-gdpr/.
[38] Sentencing Guidelines, § 8B2.1(b)(3).
[39] *Id.,* § 8B2.1(b)(4)(A).
[40] *Id.*, § 8B2.1(b)(5)(A)–(C).
[41] *Id.*, § 8B2.1(b)(6).
[42] *Id.*, § 8B2.1(7).

have a committee responsible for investigating potential unlawful activity and recommending changes to the compliance program as necessary.

V. CONCLUSION

Big Data technologies promise to unlock untold benefits. Yet these same innovations risk surreptitiously exposing data subjects to discrimination, unfairness, and other detrimental consequences. To reignite respect for individual privacy, the European Union passed the GDPR. In doing so, however, the European Union implemented legislation that is substantively uncertain as applied to Big Data. For companies seeking to avoid mortal penalties, they face a choice: continue operating and risk substantial fines, or shut down.

This Essay argues there is a third option: adopt an effective compliance system. Doing so demonstrates good-faith efforts to comply with the Regulation, which will likely lead to leniency from regulators. Because the GDPR is unclear on what constitutes an effective compliance program, this Essay recommends companies look to the Sentencing Guidelines. While controllers must nevertheless operationalize the seven elements, they can rest assured that doing so offers a buffer from the harshest regulatory sanctions. And with that buffer, they can continue improving the world.