The Legal Challenges of Big Data:

Putting Secondary Rules First in the Field of EU Data Protection

*Ugo Pagallo**

Considerable attention has been devoted in recent years to studying the legal challenges associated with Big Data. The main emphasis – including in, but not limited to, the field of data protection – has been on the role and content of the primary rules of the system. This stance makes perfect sense: it places the focus on the norms that should govern social and individual behaviour in terms that range from individual consent and data minimisation, accuracy and purpose limitation, integrity and confidentiality, to the principles of lawfulness, fairness, and transparency, as enshrined in Article 5 of the EU General Data Protection Regulation (GDPR). Still, I argue here that it is time to widen our perspective to include not only the hard laws of EU governance, but also to consider the role played by the secondary rules of the law. At the same time, we must evaluate the intent of the law in governing the process of technological innovation and the different ways in which human and social behaviours can be regulated. This article examines four types of secondary rules at work with(in) the GDPR and attempts to show how the mechanisms and procedures of legal flexibility provided by such rules may shed light on the kinds of primary rules needed within the field of Big Data.

I. Introduction

Big Data remains a fuzzy concept. The first popular definition of Big Data was provided in 2001 by Doug Laney's[1] model of the 'three Vs,' referring to volume (the size and scale of data), velocity (speed of data generation and processing) and variety (the different forms and range of the data analysed). More recently, a fourth 'V' has been proposed, namely the veracity of data, meaning that, say, user entry errors, redundancy, or corruption of the data should not affect their overall value.[2] Some argue that Big Data opens up a new perspective on reality, since patterns of data may suggest novel ways of grasping the world, to such an extent that we might even let those data speak for themselves.[3] Others concentrate on the computational and human challenges to sorting and analysing such data, and tackle issues inherent to the size and complexity of increasingly large datasets.[4] From this procedural point of view, the challenges of Big Data remind us of three different sets of problems. First, we have to take into account the technical and analytical barriers faced at the very time these data were generated and processed. Second, the focus should be on cases and issues that reveal unique ethical aspects and theoretical problems of Big Data associated with existing computing technologies. Among these issues, suffice it to mention matters of consent, anonymisation, privacy and data protection. Third, the complexity of Big Data and the algorithms used to analyse it prompt further epistemological questions of objectivity and loss of context. For example, we may lose certain aspects of the phenomena under scrutiny by reducing it to a given set of weights and variables. In addition, epistemic concerns may have to do with cases of inconclusive evidence leading to unjustified actions, or of inscrutable evidence

* Ugo Pagallo, Professor of Jurisprudence, Law School, University of Torino. For correspondence: <ugo.pagallo@unito.it>.
DOI: 10.21552/edpl/2017/1/7

1 Doug Laney, '3D Data Management: Controlling Data Volume, Velocity and Variety' (Metra Group Research Note, 2001) 6.

2 IBM, 'The Four V's of Big Data' (2014) <http://www.ibmbigdatahub.com/infographic/four-vs-big-data> accessed 15 January 2017.

3 Viktor Mayer Schönberger and Kenneth N Cukier, *Big Data: A Revolution Transforming How We Live, Work, and Think* (Houghton 2013).

4 See Brent D Mittelstadt and Luciano Floridi, 'The Ethics of Big Data: Current and Foreseeable Issues in Biomedical Contexts' (2016) 22(2) Science and Engineering Ethics 303-341.

leading to opacity, or of other kinds of misguided evidence leading to bias.[5] In normative terms, the challenges of Big Data may concern either unfair outcomes that lead to discrimination, or transformative effects that impact social and individual autonomy. Such normative challenges motivate a final overarching concern that can be summed up in terms of traceability, which goes hand in hand with issues of moral responsibility and the dilemmas of automation, ie, the acceptability of replacing or augmenting human decision-making with algorithms.[6]

Against this backdrop, this article focuses on the legal aspects of the normative challenges of Big Data, and especially on how Regulation (EU) 2016/679 on personal data protection (the GDPR) intends to govern this crucial facet of today's data-driven societies. Even when excluding the further legal challenges of Big Data from the analysis (such as matters of intellectual property and data ownership, for example),[7] this level of abstraction appears fruitful, since the GDPR aims to discipline the entire life-cycle of information regarding the production and processing of personal data through Big Data sets and techniques. Accordingly, three observables of the analysis with their variables have to be taken into account. Section I of the paper examines the regulatory claim of the law and the different ways in which legal systems intend to govern the process of technological research and social interaction by their own means. Here, we should distinguish between primary and secondary rules of the law: whereas the former aim to govern social and individual behaviour directly, the latter include rules of recognition, of adjudication, and of change, ie, the rules that allow the creation, modification, and suppression of the primary rules concerning individuals' conduct.[8] In both cases, it seems fair to affirm that the aim of the law to govern the process of technological innovation should neither hinder it, nor require over-frequent revision to manage such progress. In Section II, this sort of balance is further scrutinized in light of the primary rules of the GDPR. Such rules regard, among other things, matters of individual consent, the data minimization principle, pseudoanonymisation, and the exemption for statistical research that uses and reuses personal data. As to the instances of pseudoanonymisation, consider Apple's incorporation of differential privacy techniques in its data collection efforts for iOS and macOS, eg, the reuse of health data obtained through their apps for statistical purposes. As Apple's Senior Vice President of Software Engineering, Craig Federighi, declared at the Worldwide Developers Conference on 13 June 2016, Apple's efforts would mark the first wide-scale use of Aaron Roth's techniques. Roth is the mathematician who 'literally wrote the book' on how to learn as much as possible about a group while learning as little as possible about any individual in it.[9] The use of pseudoanonymisation techniques must be distinguished from the statistical approach to the protection of personal data. The meaning of the legal formula on the 'statistical purposes' of Big Data analysis concerns the difference between a model that, say, predicts which customers are likely to defect to competitors, so as to offer them better deals, and a model that predicts instead 'the likely overall percentage of customer churn.'[10]

Discussion of exemption for statistical purposes leads in Section III to the final observable to be analysed, namely the role and function of the secondary rules of the law and, more specifically, the rules of adjudication and change established by the GDPR. Four different types of secondary rules are at work with(in) the GDPR: (i) mechanisms of delegation of power; (ii) mechanisms of legal coordination; (iii) procedures for pre-emptive data protection; and (iv) procedures for effective judicial remedies. An instance of this type of legal mechanism and procedure can be seen in Article 36 on the powers of supervisory authorities. The provisions of this article clearly function as a set of instructions for individuals' con-

5 See Brent D Mittelstadt et al, 'The Ethics of Algorithms: Mapping the Debate' (July-December 2016) Big Data & Society 1-21.

6 See Ugo Pagallo and Massimo Durante, 'The Pros and Cons of Legal Automation and its Governance' (2016) 7(2) European Journal of Risk Regulation 323-334.

7 See for instance Josef Drexl, 'Designing Competitive Markets for Industrial Data between Propertisation and Access' (Max Planck Institute for Innovation & Competition Research, Paper No 16-13, 31 October 2016); Daniel L Rubinfeld and Michal S Gal, 'Access Barriers to Big Data' (26 August 2016) <https://ssrn.com/abstract=2830586> accessed 20 January 2017; and Gintaré Surblyté, 'Data-Driven Economy and Artificial Intelligence: Emerging Competition Law Issues' (Max Planck Institute for Innovation & Competition Research, Paper No 16-08, 5 August 2016).

8 The classical distinction in Herbert L A. Hart, *The Concept of Law* (Clarendon 1961).

9 On differential privacy 'the book' is Aaron Roth and Cynthia Work, 'The Algorithmic Foundations of Differential Privacy' (2014) 9(3-4) Foundation and Trends in Theoretical Computer Science 211-407.

10 Viktor Mayer Schönberger and Yann Padova, 'Regime Change? Enabling Big Data through Europe's New Data Protection Regulation' (2016) 17 Columbia Science and Technology Law Review 323.

duct: 'The [data] controller shall consult the supervisory authority prior to processing where a data protection impact assessment under Article 35 indicates that the processing would result in a high risk in the absence of measures taken by the controller to mitigate the risk.' Yet the supervisory authorities of Article 36 are those of each Member State where the controller has its main establishment: see the secondary rule of Article 55. This cross-reference implies efforts at coordination; otherwise, the system risks breaking down. On the one hand, we might envision beneficial competition among legal systems: a sort of EU version of Justice Brandeis's doctrine of experimental federalism, as espoused in *New State Ice Co. v Leibmann* (285 US 262 (1932)). On the other hand, some warn that national preferences, values, and fears will fatally determine the regulatory future in Europe.[11] Others critically note that – even after the European Parliament and the Council approved a number of new responsibilities for data controllers and a set of novel rights for data subjects relevant to decision-making algorithms and reuse of Big Data – several of the rules still seem vague and opaque. 'The GDPR can be a toothless or a powerful mechanism to protect data subjects dependent upon its eventual legal interpretation: the wording of the regulation allows either to be true.'[12]

The conclusion of the analysis brings us back to the regulatory claim of the law and how legal systems aim to govern technological innovation, much as individual and social conduct, by their own means (Section II), through either primary (Section III), or secondary rules (Section IV). Ultimately, the interplay between law and technology, and between GDPR and Big Data needs to be grasped as the interaction between competing regulatory systems that are contending against further regulatory systems, such as the forces of the market and of social norms. These regulatory claims may either clash or reinforce one another, and one regulatory system may even render the claim of another regulatory system superfluous. There are dozens of cases in which the legal intent to regulate the process of technological innovation has miserably failed. One example is EU Directive 46 from 2000, whose regulation of electronic money considered it as a mere surrogate of traditional currencies falling under the supervision of national financial authorities. New modalities of e-payment and transactions, such as PayPal, soon rendered the EU provisions inadequate, forcing lawmakers in Brussels to pass a new Directive (D-2009/110/EC). Another instance is that of Article 8 of the World Intellectual Property Organization's 1996 Copyright Treaty and Article 14 of the twin Performances and Phonograms Treaty. Twenty years after these international agreements were signed, it is clear that the legal rules have fallen short in coping with people's behaviour online and the dynamics of technological innovation. This paper shows that the GDPR's secondary rules play a crucial role in determining three kinds of balance: the balance between

(i) competitive regulatory systems;
(ii) efforts at coordination and risks of breaking down; and,
(iii) the protection of multiple legal rights, which nonetheless should not hinder responsible technological research of Big Data through manifold techniques, such as machine learning, ie, algorithms capable of autonomously defining or modifying decision-making rules;[13] or data analytics, namely the use of algorithms that make sense of huge streams of data.[14]

All in all, it may even be possible for GDPR's secondary rules to be interpreted in a way that makes the threefold balance feasible.

II. On Law and Technology

According to a glorious philosophical tradition extending from at least Kant to Kelsen, the law can conveniently be understood as a technique. As phrased in the *General Theory of the Law and the State*,[15] 'what distinguishes the legal order from all other social orders is the fact that it regulates human behaviour by means of a specific technique that hinges on the threat of physical coercion: 'if A, then B.' Now, if the law is a technique that regulates another technique,

11 ibid 331.

12 Mittelstadt et al, 'The Ethics of Algorithms' (n 5) 14.

13 See Martijn Van Otterlo, 'A Machine Learning View on Profiling' in Mireille Hildebrandt and Katja de Kries (eds), *Privacy, Due Process and the Computational Turn – Philosophers of Law Meet Philosophers of Technology* (Routledge 2013) 41-64.

14 See Luciano Floridi, 'Big Data and their Epistemological Challenge' (2012) 25(4) Philosophy & Technology 435-437; and Peter Grindrod, *Mathematical Underpinnings of Analytics: Theory and Applications* (Oxford University Press 2014).

15 Hans Kelsen, *General Theory of the Law and the State* (Anders Wedberg tr, Harvard University Press 1945/1949).

and if that other technique is the process of technological innovation, we may consider the law to be a meta-technology. From this point of view, it does not follow that we have to accept any of Kelsen's ontological commitments: the stance this article adopts on the law as meta-technology implies neither that the law is merely a means of social control, nor that no other meta-technological mechanisms exist. Rather, by insisting on the intent of the law to govern the process of technological innovation, the focus should be on the 'whys' and 'hows' of the regulation of human and social behaviour.[16]

Some suggest that we should distinguish four main legislative goals: (a) the achievement of particular effects; (b) functional equivalence between online and offline activities; (c) non-discrimination between technologies with equivalent effects; and, (d) future-proofing of the law that should neither hinder the advance of technology, nor require over-frequent revision to tackle such progress.[17] Others propose differentiating (a) technological indifference, ie, legal regulations that apply in identical ways, no matter what technology; (b) implementation neutrality, according to which regulations are by definition specific to that technology and yet do not favour one or more of its possible implementations; and, (c) potential neutrality of the law that sets up a particular attribute of a technology, although lawmakers can draft the legal requirement in such a way that even non-compliant implementations can be modified to become compliant.[18]

As to the ways in which the law can regulate both human and social behaviours, we should further distinguish between the traditional technique of rules that hinge on the menace of legal sanctions and techno-regulation, that is, legal regulation by design.[19] For example, the intent of the law to govern both human and social behaviours in the field of robotics can be divided into the following four categories: (a) the regulation of human producers and designers of robots and other artificial agents through law, eg, either through ISO standards or liability norms for users of robots; (b) the regulation of user behaviour through the design of Artificial Intelligence (AI) apps, that is, by designing them in such a way that unlawful actions of humans are not allowed; (c) the regulation of the legal effects of artificial behaviour through the norms set up by lawmakers, eg, the effects of robotic contracts and negotiations; and, (d) the regulation of artificial behaviour through design, that is, by embedding normative constraints into the design of the application.[20] This differentiation can be complemented with further work on how the environment of human-robot interaction can be regulated, and the legal challenges of 'ambient law.'[21] Accordingly, attention should be drawn to the set of values, principles, and norms that constitute the context in which the consequences of such regulations must be evaluated.[22] As stressed above in the introduction, most scholars would admit today that national preferences, values, and fears will play a crucial role in shaping the regulatory future of data protection in Europe.

Yet, by insisting on the different purposes and techniques of the law, we may reinterpret such regulatory claims in binary terms. They either concern the primary rules of the legal system, or have to do with the different kinds of secondary rules: recognition, adjudication, change. As previously mentioned in the introduction, the aim of the primary rules is to directly govern human and social behaviour either through techno-regulation, eg, some variants of the principle of privacy by design, or the manifold means of law as a meta-technology, such as achieving particular effects with hard laws (eg, the primary rules of GDPR); administrative regulation (eg, ISO standards); or soft law (eg, the powers of data protection authorities). The aim of the secondary rules of change

16 An overview in Ugo Pagallo, *The Laws of Robots: Crimes, Contracts, and Torts* (Springer 2013).

17 See Bert-Jaap Koops, 'Should ICT Regulation Be Technology-neutral? in Bert-Jaap Koops et al (eds), *Starting Points for ICT Regulation: Deconstructing Prevalent Policy One-liners* (TMC Asser 2006).

18 See Chris Reed, *Making Laws for Cyberspace* (Oxford University Press 2012).

19 See Ugo Pagallo, 'On the Principle of Privacy by Design and its Limits: Technology, Ethics, and the Rule of Law' in Serge Gutwirth et al (eds), *European Data Protection: In Good Health?* (Springer 2012) 331-346; Ugo Pagallo, 'Cracking down on Autonomy: Three Challenges to Design in IT Law' (2012) 14(4) Ethics and Information Technology 319-328; and Ugo Pagallo, 'Designing Data Protection Safeguards Ethically' (2011) 2(2) Information 247-265.

20 See Ronald Leenes and Federica Lucivero, 'Laws on Robots, Laws by Robots, Laws in Robots: Regulating Robot Behaviour by Design' (2016) 6(2) Law, Innovation and Technology 193-220.

21 Check, among others, the work of Mireille Hildebrandt and Bert-Jaap Koops, 'The Challenges of Ambient Law and Legal Protection in the Profiling Era' (2010) 73(3) Modern Law Review 428-460; and Mireille Hildebrandt, 'Legal Protection by Design: Objections and Refutations' (2011) 5(2) Legisprudence 223-248.

22 See Ugo Pagallo and Massimo Durante, 'The Philosophy of Law in an Information Society' in Luciano Floridi (ed), *The Routledge Handbook of Philosophy of Information* (Routledge 2016) 396-407 .

is to allow the creation, modification, and suppression of the primary rules. This aim can either concern the substitution of a given regulation, eg, the primary rules of the EU Data Protection Directive 46 from 1995, with the new set of primary rules of the GDPR, or they can concern mechanisms of legal flexibility. Consider the Federal Automated Vehicles Policy adopted by the US Department of Transportation in 2016. Here, we can appreciate the overall legislative goal of the policy, that is, the principle of 'implementation neutrality,' meaning that it does not intend to favour one or more of the possible applications in the field of self-driving cars. Another approach to the challenges of technological innovation has been worked out by the Japanese government through the creation of special zones for robotics empirical testing and development, namely, a form of living lab, or *Tokku*. After the Cabinet Office approved the world's first robotic special zone covering the prefecture of Fukuoka and the city of Kitakyushu in November 2003, further special zones have been established in Osaka and Gifu, Kanagawa and Tsukuba. The overall aim of these special zones is to set up a sort of interface for robots and society, in which scientists and laypeople alike can test whether manifold AI applications fulfil their task specifications in ways that are acceptable and comfortable to humans vis-à-vis the uncertainty of machine safety and legal liabilities that concern, eg, the protection for the processing of personal data.[23] Remarkably, a special zone for privacy and data protection was set up in the city of Kyoto in 2008.

The interplay between law and technology, eg, the rules of the GDPR and the challenges of Big Data, can thus be explored by distinguishing between the primary and secondary rules of the law. This differentiation sheds light on the different purposes and ways in which human and social behaviour has been regulated by the GDPR. The next section of the paper examines the primary rules of the GDPR in terms of the legal challenges of Big Data. Then, in Section IV, attention will be drawn to four different types of secondary rules at work with the GDPR.

III. The Primary Rules of the GDPR

Issues of data protection under EU law mainly have to do with the transparency of data collection, processing and use. Individuals have the right to know why their data is being processed, as well as the right to access that data and have it rectified. In the wording of Article 8(2) of the EU Charter of Fundamental Rights, 'such data must be processed fairly... and on the basis of the consent of the person concerned or some other legitimate basis laid down by law.' This type of protection through the principles of minimisation and quality of data, its controllability and confidentiality, aims to constrain the flow of information, and to maintain firm distinctions between individuals and society, so as to protect that which the German Constitutional Court has framed in terms of 'informational self-determination' since its *Volkszählungs-Urteil* ('census decision') of 15 December 1983. This general right to the *informationelle selbst-bestimmung* of individuals includes the right to determine whether personal data can be collected and potentially transmitted to others; the right to determine how that data may be used and processed; the right to access that data and, where necessary, to keep it up to date; and the right to delete that data and refuse at any time to have the data processed.

Within this general framework, the GDPR has substantially maintained the architecture of Directive 95/46/EC, which was largely based on the Organisation for Economic Co-operation and Development (OECD)'s 1980 'information-and-consent' Privacy Guidelines.[24] In short, the aim of the GDPR is to strengthen both individual's rights and the powers of the European authorities, while reinforcing the obligations and responsibilities of data controllers through a directly enforceable hard law-tool in the form of an EU Regulation. This threefold dimension of the GDPR is evident in some of its primary rules, including:

- Articles 21 and 22 on individual self-determination and automated decision-making, eg, profiling;

23 Further details in Ugo Pagallo, 'Robots in the Cloud with Privacy: A New Threat to Data Protection?' (2013) 29(5) Computer Law & Security Review 501-508. More specifically on the role of the secondary rules of the law in this context, see Ugo Pagallo, 'Even Angels Need the Rules: AI, Roboethics, and the Law' in Gal A Kaminka et al (eds), *ECAI 2016. Frontiers in Artificial Intelligence and Applications* (IOS Press 2016) 209-215; Ugo Pagallo, 'Three Lessons Learned for Intelligent Transport Systems that Abide by the Law' (2016) JusLetter IT <http://jusletter-it.weblaw.ch/issues/2016/24-November-2016/three-lessons-learne_9251e5d324.html> accessed 21 January 2017; and, Ugo Pagallo, 'When Morals Ain't Enough: Robots, Ethics, and the Rules of the Law' (2017) Minds and Machines, doi: 10.1007/s11023-017-9418-5.

24 See the OECD document at <https://www.oecd.org/sti/ieconomy/oecdguidelinesontheprotectionofprivacyandtransborderflowsofpersonaldata.htm> accessed 18 January 2017.

- Article 33 on notifications of personal data breaches to the supervisory authority 'competent in accordance with Article 55,' along with the powers of the latter to impose administrative fines pursuant to Article 83;
- Article 17 on the right to erasure, or the right to be forgotten, and Article 20 on data portability, as a new set of duties and obligations for data controllers.

On this basis, the question is then to ascertain whether and to what extent the real-time generation and processing of personal Big Data are compatible with the provisions, ie, the primary rules and principles of the new legal framework, such as the principle of purpose limitation and data minimisation.[25] After all, the value of data may become apparent after it has been used time and again for purposes other than that for which consent was originally requested. Since researchers and data controllers may not know at the time of data collection what the value of the Big Data is or how it might possibly be exploited in the future, the establishment of stricter legal guidelines for legitimate individual consent may constitute a formidable obstacle to an otherwise fruitful collection and use of Big Data.[26] Therefore, is there any way to keep the principles and rules of the GDPR from hindering this field of technological innovation and competitive business in Europe?

Actually, the new legal framework provides for two possible solutions. The first way to make the collection and use of Big Data compatible with the tenets of the GDPR concerns the use of pseudonymisation techniques. In the wording of Article 4(5), this means that personal data is processed

> in such a manner that the personal data can no longer be attributed to a specific data subject without the use of additional information, provided that such additional information is kept separately and is subject to technical and organisational measures to ensure that the personal data are not attributed to an identified or identifiable natural person.[27]

The second solution has to do with the exemption of data processing for statistical purposes, pursuant to Articles 5(1)(b) and (e), 14(5)(b), 17(3)(d), 21(6) and 89 of the GDPR. Whereas safeguards and derogations for the processing of personal data for statistical purposes are regulated together with similar protections and exemptions for scientific or historical research purposes, or for archiving purposes in the public interest, the two solutions provided by the new legal framework may of course overlap, eg, the employment of pseudonymisation techniques so as 'to ensure respect for the principle of data minimisation' in the processing of personal data for statistical purposes, in accordance with the wording of Article 89(1). Moreover, the sets of primary rules established in both cases propose an interesting interplay with the secondary rules at work with(in) the GDPR. Consider, for instance, Recital 162 of the Regulation, according to which

> where personal data are processed for statistical purposes... Union or Member State law should, within the limits of this Regulation, determine statistical content, control of access, specifications for the processing of personal data for statistical purposes and appropriate measures to safeguard the rights and freedoms of the data subject and for ensuring statistical confidentiality.[28]

Likewise, consider Articles 35 and 36 on a new generation of data protection impact assessments and prior consultation with the supervisory authorities that, again, point out the crucial role that Member

25 Check GDPR's art 5 on the 'principles relating to the processing of personal data,' which are: (a) the principles of lawfulness, fairness and transparency; (b) purpose limitation; (c) data minimisation; (d) accuracy; (e) storage limitation; and, (f) integrity and confidentiality.

26 As to the principle of individual consent, see the conditions set up by art 7 of the GDPR and compare with recital 42 on the mechanism of the burden of proofs. On the one hand, consent should be requested in an intelligible and easily accessible form: nowadays, suffice it to say that, for example, PayPal's terms of service are longer than Shakespeare's *Hamlet*, ie 36275 v 30066 words! On the other hand, 'consent should not provide a valid legal ground for the processing of personal data in a specific case where there is a clear imbalance between the data subject and the controller' (GDPR recital 43).

27 Accordingly, pseudonymisation techniques play a crucial role for the enforcement of the principle of data protection by design and by default [art 25), data security (art 30) and codes of conduct (art 40(2)(d)].

28 In truth, recital 162 offers an unsatisfactory definition of statistics, since 'statistical purposes mean any operation of collection and the processing of personal data necessary for statistical surveys or for the production of statistical results.' However, on the other hand, attention should be drawn to the fact that 'those statistical results may further be used for different purposes, including a scientific research purpose.' In addition, and more importantly, 'the statistical purpose implies that the result of processing for statistical purposes is not personal data, but aggregate data, and that this result or the personal data are not used in support of measures or decisions regarding any particular natural person.' See above in the introduction for the distinction between, say, profiling techniques and statistical purposes.

States and national legal systems will play here pursuant to, eg, Article 35(10).

To improve our understanding of the primary rules of the GDPR, let us now turn to how they interplay with the secondary rules. Only then will we be fully prepared to appreciate the political choices of EU lawmakers.

IV. The Secondary Rules of the GDPR

As mentioned above in the introduction, the secondary rules of the law comprise rules of recognition, of adjudication, and of change. The rules of recognition are the meta-rules by which all other rules of the system are identified and understood as valid, ie, that which counts as a valid law within that system. The rules of adjudication prescribe a remedy for all cases in which a rule has been violated, eg, the procedures that supervisory authorities should follow pursuant to Article 36, or 83, of the GDPR. Finally, the rules of change allow for creating, modifying, or suppressing the primary rules of the system. Among this set of rules, there are specific techniques, procedures, and even legal experiments. As to the techniques, recall the 'implementation neutrality'-principle endorsed by the Federal Automated Vehicles Policy of the US Department of Transportation in September 2016, which was mentioned above in Section II. As to the procedures, they include the meta-rules of 'procedural regularity,' so as to determine whether a decisional process is fair, adequate, or correct.[29] As to the forms of legal experimentation, recall the creation of special zones for robotics empirical testing and development set up by the Japanese government over the past twelve years, also mentioned in Section II above.

Of course, some of these secondary rules can interact and reinforce each other. Think, for instance, of the rules of adjudication as a viable way to strengthen some rules of change, eg, Justice Brandeis's doctrine of experimental federalism that was stressed in the introduction of this article. Yet, against the panoply of possible uses of secondary rules and their interplay with the primary rules of the system, what scenario will be reasonably favoured by the GDPR?

Let us assume in this context the tenets of the Dworkinian right answer-thesis, according to which a morally coherent narrative should grasp the law in such a way that, given the nature of the legal question and the story and background of the issue, scholars can attain the answer that best justifies or achieves the integrity of the law. By identifying the principles of the system that fit with the established law, jurists could apply such principles in such a way that presents every issue in the best possible light.[30] Accordingly, we could interpret all the cases in which the GDPR employs secondary rules to delegate powers back to national legal systems, eg, setting the safeguards to be in place for the processing of personal data for statistical purposes pursuant to Recital 162 and Article 89(1) of the Regulation, as a smart way to flesh out the content of the primary legal rules through a beneficial competition among legal systems at EU level. Admittedly, in light of the same mechanism, some warn instead of possible risks of fragmentation. The delegation of powers to Member States, in other words,

> will likely result in some two dozen different regulatory frameworks throughout the European Union. It will enable some nations to be more permissive of Big Data, and others more restrictive. It will certainly reduce the impact of the regulation as a harmonizing force of data protection regulation in Europe in the context of Big Data, and it will make life harder for companies and organizations operating not just in one but multiple Member States in Europe.[31]

As to risks of fragmentation, eg, multiple jurisdictions of national supervisory authorities in the field of EU data protection, however, such risks can be tackled either with technical standards, eg, the aforementioned meta-rules of 'procedural regularity,' or with efforts of coordination. GDPR Recitals 13, 36, 86, 135, etc – much like its Articles 60, 61, 75(4) and 97(2)(b) – convey this specific aim. Although this set of secondary rules do not guarantee per se a coher-

29 See for instance Joshua A Kroll et al, 'Accountable Algorithms' (forthcoming 2017) 165 University of Pennsylvania Law Review.

30 The reference text is of course Ronald Dworkin's *A Matter of Principle* (Oxford University Press 1985). Needless to say that, as much as in the introduction, when quoting Hart's distinction between the primary and secondary rules of the law, we may adopt this stance without buying any of the ontological commitments, eg law as literature, of Dworkin.

31 Schönberger and Padova (n 10) 327-328. Remarkably, these scholars nevertheless admit a possible positive side effect of this delegated setup: it 'may help established national players engaged in Big Data analysis.'

ent interaction between multiple national legal systems and their supervisory authorities, it seems fair to affirm that the GDPR provides a number of ways to cope with the centrifugal forces of the system, such as 'national preferences, values, and fears.'[32]

In addition, we should recall what was stressed in the introduction of this article concerning the regulatory claims of the law and, moreover, the competition between different regulatory systems, such as the forces of the market and of social norms. Such regulatory claims may not only clash, but also reinforce one another. Also, a regulatory system can render the claim of another regulatory system superfluous. Whatever scenario we consider, such competition does not take place in a normative vacuum, but is structured by the presence of values and principles. Therefore, policy makers and legislators should also keep in mind the degree of social acceptability and cohesion that affect their own decisions. Furthermore, technology can dramatically change these very expectations on what is socially acceptable and the extent to which social cohesion may be affected by technology. As Justice Alito emphasized in his concurring opinion in the Supreme Court's *United States v Jones* ruling from 23 January 2012 (565 US __), 'dramatic technological change may lead to periods in which popular expectations are in flux and may ultimately produce significant changes in popular attitudes.' As a result, on the one hand, the use of secondary rules may represent a mechanism of legal flexibility that allows us to address the interaction between regulatory systems wisely. On the other, specific types of secondary rules can improve the future-proofing of the law, assuring that it does not curtail technological innovation or require over-frequent revision to address such progress.

Interestingly, since the proposal for the new EU Data Protection Regulation was presented in January 2012, the European Commission has referred to the principle of implementation neutrality. In the wording of the 66th 'whereas' of the proposal, 'when establishing technical standards and organisational measures to ensure security of processing, the Commission should promote technological neutrality, interoperability and innovation.' Almost two years later, on 20 November 2013, the formula reappeared in the first round of amendments presented by the European Parliament vis-à-vis the initial phrasing of Article 86 on the powers of the Commission to adopt delegated acts under the conditions laid down by that article. According to the last amendment passed by the Parliament in 2013, ie, amendment number 196, 'the Commission shall promote technological neutrality while implementing the acts established by the current provision.' Two and a half years later, when the official text of the GDPR was finally published in the EU official journal on 4 May 2016, the reference to the notion of technological neutrality reappears in Recital 15: '

> In order to prevent creating a serious risk of circumvention, the protection of natural persons should be technologically neutral and should not depend on the techniques used. The protection of natural persons should apply to the processing of personal data by automated means, as well as to manual processing.

This meta-technological approach makes a lot of sense, of course, because it goes hand-in-hand with the previous secondary rules of experimental federalism, so as to set up a mechanism of legal flexibility. As already mentioned in the previous section, while the collection and use of Big Data can be compatible with the tenets of the GDPR thanks to the use of pseudonymisation techniques, it is noteworthy that the Regulation repeatedly refers to such techniques as one of the possible ways to provide 'appropriate safeguards' [Article 6(4)(e)]; 'appropriate technical and organisational measures' [Articles 25(1)) and 89(1)]; or, coupling the reference with further ways to protect personal data through encryption [Articles 6(4)(e) and 32(1)(a)].

In addition to forms of experimental federalism and the meta-technological approach of technological neutrality, attention should finally be drawn to harm, prejudice and risks or threats brought on by current Big Data trends, and how the GDPR intends to further address these issues through two different sets of secondary rules. On the one hand, the Regulation maintains the traditional equalisation of data subjects and natural persons [Article 4(1)]. And yet, rather than a unique data subject whose informational self-determination is specifically under attack, individuals will more often be targeted as a member of a group, or as a specimen falling within the set of ontological and epistemological predicates that cluster

32 ibid 331.

a group. New types of harm and threats should be expected as a result, since this trend is more about the new protection of 'sardines,' ie, individuals as members of a group, than 'Moby Dicks.' And while 'the individual sardine may believe that the encircling net is trying to catch it... it is not... it is trying to catch the whole shoal.'[33] Correspondingly, the traditional type of protection against individual harm in the field of data protection has to be supplemented with an analysis of the risks and threats to the processing and use of group data that may provoke new kinds of harm to most of us, namely the 'sardines.' This is the stance stressed by Article 29 Working Party (A29 WP)'s Opinion 3/2012 on developments in biometric technologies and, moreover, this is the procedural approach followed by the Court of Justice of the EU (CJEU) in its own readings of Articles 7 and 8 of the EU Charter of Fundamental Rights.[34] Whereas Big Data trends will increasingly give rise to cases that affect groups, rather than individuals, so that current rights of the personal data protection framework should be properly complemented with a new generation of collective rights,[35] the GDPR partially endorses this new kind of protection. Pursuant to the secondary rule of Article 80(1), 'the data subject shall have the right to mandate a not-for-profit body, organisation or association which has been properly constituted... to lodge the complaint on his or her behalf,' so as to exercise the right to an effective judicial remedy against a supervisory authority, or against a controller or processor, or the right to lodge a complaint with a supervisory authority and to receive compensation. Furthermore, in accordance with the mechanism of experimental federalism set up by Article 80(2), 'Member States may provide that any body, organisation or association referred to in paragraph 1 of this Article, independently of a data subject's mandate, has the right to lodge, in that Member State, a complaint,' which may concern either the right to an effective judicial remedy against a supervisory authority, or against a controller or processor, with the supervisory authority which is competent pursuant to Article 77 of the GDPR. Hence, the overall idea of this set of rules is not to replace today's personal data protection with a sort of US-like privacy group regime,[36] but rather to complement it with a new collective right to lodge complaints.[37] Since the data subject can be targeted and her privacy infringed due to her membership in a given (racial, ethnic, genetic, etc) data group, it makes sense to grant such a group, or 'any body, organisation or association which aims to protect data subjects' rights and interests,' a procedural right to a judicial remedy against the data controllers, processors or supervisory authorities.

On the other hand, scholars have insisted time and again on the shortcomings of the traditional 'information-and-consent' approach that the EU 1995 Directive first and the EU 2016 Regulation then inherited from the 1980 OECD Privacy Guidelines.[38]

> In its place, one could imagine a mechanism that focuses less on individual consent than on the regulation of permissible and prohibited uses of personal data, protecting individuals irrespective of whether they habitually click the consent button, while also enabling and facilitating accountable and ethical Big Data use... This requires quite a different approach by data processing entities, shifting away from rituals of consent to deliberate assessment procedures *ex ante*––not just of the benefits but also the potential risks and harms for individuals associated with a particular data

33 Luciano Floridi, 'Open Data, Data Protection, and Group Privacy' (2014) 27 Philosophy and Technology 3. This risk of individuals targeted as a member of a specific (racial, ethnic, genetic, etc) group was realized throughout the 2010-2011 Ivorian civil war. See Linnet Taylor, 'No Place to Hide? The Ethics and Analytics of Tracking Mobility Using African Mobile Phone Data' (2016) 34 Environment and Planning D – Society & Space 319-336.

34 In the Joined Cases C-293/12 and 594/12 *Digital Rights Ireland and Seitlinger and Others* (8 April 2014) ECLI:EU:C:2014:238, the CJEU accepted the claims by certain Austrian and Irish organizations of being victims of a violation of their rights under the provisions of the Data Retention Directive 24/2006. As to the Opinion of A29 WP, see WP 193: 'in this case, it is not important to identify or verify the individual but to assign him/her automatically to a certain category.'

35 I emphasized this critical evolution of the data protection framework years ago. See Ugo Pagallo, *Il diritto nell'età dell'informazione* (Giappichelli 2014). More recently, Ugo Pagallo, 'The Group, the Private, and the Individual: A New Level of Data Protection?' in Linnet Taylor, Luciano Floridi and Bart van der Sloot (eds), *Group Privacy: New Challenges of Data Technologies* (Springer 2017) 159-173.

36 Although crucial differences do exist and persist, we can say that US and EU regulations in the field of data protection are 'separated by common goals.' See David Vladeck, 'Separated by Common Goals: A U.S. Perspective on Narrowing the U.S.-EU Privacy Divide' in Artemi Rallo Lombarte and Rosario García Mahamut, *Hacia un nuevo derecho europeo de protección de datos* (Tirant lo blanch 2015) 207-243.

37 According to the US Supreme Court's ruling in *Boy Scouts of America v Dale* [530 US 640 (2000)], the privacy of a large civic membership organisation, as a single and unitary holder, such as the Boy Scouts, can be conceived analogously with an individual's privacy, that is as a corporate — rather than a collective — right. See Ugo Pagallo, 'The Group, the Private, and the Individual' (n 35).

38 An overview in Ugo Pagallo, *Il diritto nell'età dell'informazione* (n 35).

> use––and the necessity to devise and implement concrete mitigation strategies.[39]

Again, the GDPR partially endorses this new kind of protection by setting up a number of data protection impact assessments that should determine risks and threats for the processing and use of certain kinds of data. Pursuant to its Article 35(1), data controllers will thus have the responsibility of performing a data protection impact assessment 'where a type of processing in particular using new technologies, and taking into account the nature, scope, context and purposes of the processing, is likely to result in a high risk to the rights and freedoms of natural persons.' In particular, specific risks concern 'a systematic and extensive evaluation of personal aspects relating to natural persons which is based on automated processing' [Article 35(3)(a)]; 'processing on a large scale of special categories of data,' such as sensitive data or criminal records [Article 35(3)(b)]; and 'a systematic monitoring of a publicly accessible area on a large scale' [Article 35(3)(c)]. Whereas this set of secondary rules partly overlap with the aforementioned forms of experimental federalism and a new generation of collective and procedural rights in the field of data protection, their aim is quite clear: when dealing with the pace of technological innovation, data protection should be pre-emptive, rather than remedial, in order to ensure that privacy safeguards are at work even before a single bit of information has been collected.

Undoubtedly, whether this complex network of secondary rules set up by the GDPR will prove effective, or good enough, to tackle the challenges of Big Data trends, remains an open question. The difficulty does not only hinge on the type of harm, threat, or risks that the processing and use of Big Data may give rise to in terms of physical threat or injury, unlawful discrimination, loss of confidentiality, identity theft, financial loss, etc. Moreover, whether the GDPR will actually be a toothless or a powerful mechanism to protect data subjects depends on how its primary and especially, its secondary rules, at times 'rather vague and opaque,' are going to be interpreted.[40] This margin of uncertainty is not, in itself, inherently a fault. It may represent a wise mechanism of legal flexibility with which to tackle the challenges posed by the astonishing advancements in technology, competition between regulatory systems, future-proofing of the law and more.

V. Conclusions

Considerable attention has been devoted in recent years to studying the legal challenges associated with Big Data. The main emphasis – including in, but not limited to, the field of data protection – has been on the role and content of the primary rules of the system. This stance makes perfect sense, since it focuses on the norms that should govern social and individual behaviour. Accordingly, as illustrated in Section III, attention is drawn to the set of provisions and principles that should govern the collection and use of Big Data in Europe, ie, the norms that have to do with the primary rules on individual consent and data minimisation, accuracy and purpose limitation, integrity and confidentiality, as well as the principles of lawfulness, fairness, and transparency, as enshrined in Article 5 of the GDPR. In addition, we have examined two possible solutions to make the collection and use of Big Data compatible with the tenets of the GDPR: the use of pseudonymisation techniques and the exemption of data processing for statistical purposes, which, once again, concern the role and content of the primary rules of the system.

Still, in addition to the hard-tools of legal governance, it is time to widen our viewpoint. On the one hand, we need to take into account the role that the secondary rules of the law may play in this context. On the other, we should evaluate the intent of the law in governing the process of technological innovation and the 'whys' and 'hows' of the regulation of human and social behaviour. In light of the general framework on law and technology provided above in Section II, Section IV illustrated four types of secondary rules that can be identified with(in) the GDPR:

- Mechanisms of delegation of powers, eg, setting the safeguards to be in place for the processing of personal data for statistical purposes that, in the best possible scenario, could end up with a European form of experimental federalism;
- mechanisms of legal coordination that should prevent, or lessen, threats and risks of fragmentation that may be triggered by multiple jurisdictions of national supervisory authorities;

39 Schönberger and Padova (n 10) 332.

40 Mittelstadt et al, 'The Ethics of Algorithms' (n 5) 14.

- procedures for a pre-emptive, rather than remedial, protection of personal data, in order to guarantee that privacy safeguards are set up even before a single bit of information has been collected;
- procedures for an effective judicial remedy through a new collective right to lodge complaints that at least partially takes into account how Big Data treats types rather than tokens, and hence, groups rather than individuals.

Commentators frequently focus on the opaqueness and vagueness of certain provisions of the GDPR,[41] or on the centrifugal forces of the system, such as 'national preferences, values, and fears' that may prevail over the general design of the GDPR.[42] However, perhaps the secondary rules will be even more useful to us in dealing with those risks than an understanding of the margin of ambiguity, or of indecisiveness, in the text of the Regulation. By looking beyond the horizon of the GDPR's primary rules, in other words, and paying greater attention to its secondary rules and the meta-technological option on the principle of technological neutrality, we may gain a better grasp of how the GDPR can tackle the legal challenges of Big Data. These challenges concern:

- The breath-taking pace of technological innovation and, in more general terms, the competition between regulatory systems;
- the efforts of coordination that should prevail over the aforementioned risks of fragmentation;
- the protection of legal rights that should not hinder technological research and innovation in this field.

In the original design of the proposal for the new Data Protection Regulation presented in January 2012, the Commission self-attributed a number of powers – under the label of delegated acts, pursuant to Article 86 of the proposal – that, quite understandably, the EU Parliament fiercely opposed. At the end of the day, a reasonable compromise has been struck through the GDPR's secondary rules that provide a margin of flexibility to cope with the challenges of Big Data. In particular:

- The principles of technological neutrality and experimental federalism can make the GDPR flexible enough to govern the process of technological innovation;
- mechanisms of legal coordination can counterbalance mechanisms of delegation of power;
- procedures for pre-emptive protection of personal data and judicial remedies via new collective rights can contrast some of the specific legal challenges of Big Data, eg, the protection of group rights that complement the rights of individuals, without curtailing the vibrant research and development in this field.

As in other areas of technological innovation (eg, AI and robotics), Big Data advancements and techniques have put the secondary rules of the law in the spotlight.[43] Whether the principles, mechanisms and procedures set up by the GDPR's secondary rules will be successful in coping with the legal challenges of Big Data remains an open question whose answer depends on a number of complex factors (including the constraints of the GDPR's primary rules). It requires little power of the imagination to expect that the secondary rules of the law likely play an important role in this crucial domain of current legal systems. By providing mechanisms and procedures of legal flexibility, the secondary rules shed light on the kinds of primary rules needed within the field of Big Data.

41 ibid.

42 See Schönberger and Padova (n 10) 331.

43 See Ugo Pagallo, 'Even Angels Need the Rules' (n 23); Ugo Pagallo, 'Three Lessons Learned for Intelligent Transport Systems' (n 23); and Ugo Pagallo, 'When Morals Ain't Enough' (n 23).