Comparative Law Review

VOLUME 11/2

ISSN: 2038-8993

COMPARATIVE LAW REVIEW
The Comparative Law Review is a biannual journal published by the
I. A. C. L. under the auspices and the hosting of the University of Perugia Department of Law.

Office address and contact details:
Department of Law - University of Perugia
Via Pascoli, 33 - 06123 Perugia (PG) - Telephone 075.5852437
Email: complawreview@gmail.com

Comparative Law Review is registered at the Courthouse of Monza (Italy) - Nr. 1988 - May, 10th 2010.

COMPARATIVE LAW REVIEW

VOL. 11 /2

Remembering Stefano Rodotà

"LEGAL BIG DATA": FROM PREDICTIVE JUSTICE TO PERSONALISED LAW?

Andrea Sstazi

TABLE OF CONTENTS

The phenomenon of Big Data intersects with comparative law and justice in several noteworthy profiles.
First, the comparative approach leads to identifying the peculiar characteristics of the data through a conceptual framework of the same in the perspective of other disciplines, in particular economics and information technology.
Then, in view of the different legal issues posed by Big Data, comparative law can help develop and provide data management and analysis services across national borders.
Finally, the application of data analysis methods to legal issues can give rise to "Legal Big Data" through which it might be possible to observe evolutionary patterns and paths of law, foresee or adopt jurisprudential decisions, develop and apply laws or regulations based on solid argumentative and comparative elements.

I. INTRODUCTION: BIG DATA AND COMPARATIVE LAW

Nowadays, computer science, internet networks and the connections that are established between these and material things allow us to collect, manage and analyze large quantities of data, so-called Big Data, which are collected and analyzed through advanced systems and technologies, and which allow to develop innovations, goods and services at unprecedented speed and socio-economic impact: this is the so-called Data Driven Innovation[1].

Big Data differs from "small data" in what are their characteristics, summarized in the so-called. 5V, i.e. volume, variety, speed, value and veracity[2]. In fact, what allows to extract value

[1] In this regard, see, *ex multis*: OECD, *Big Data: Bringing Competition Policy to the Digital Era*, October 2016. https://one.oecd.org/document/DAF/COMP(2016)14/en/pdf, 5, according to which: "Big Data is the information asset characterized by such a high volume, velocity and variety to require specific technology and analytical methods for its transformation into value"; Id., *Data-Driven Innovation. Big Data for Growth and Well-Being*, October 2015. http://www.oecd.org/sti/data-driven-innovation-9789264229358-en.htm; Id., *Data-driven Innovation for Growth and Well-being*, Interim Synthesis Report, October 2014. https://www.oecd.org/sti/inno/data-driven-innovation-interim-synthesis.pdf; Schönberger V.M., Ramge T., *Reinventing Capitalism in the Age of Big Data*, New York: Basic Books, 2018, 1 ff.; Schönberger V.M., Cukier K., *Big Data. A Revolution That Will Transform How We Live, Work, and Think*, London: John Murray, 3 ff.

[2] See: Stucke M.E., Grunes A.P., *Big Data and Competition Policy*, Oxford: Oxford University Press, 2016, 16; De Mauro A., A formal definition of Big Data based on its essential features. *Library Rev.*; 2016, 122; Gandomi A., Haider M., Beyond the hype: Big data concepts, methods and analytics. *Intern. Journ. Inform. Manag.* 2015, 137; Lukoianova T., Rubin V.L., Veracity Roadmap: Is Big Data Objective, Truthful and Credible? *Advan. Classif. Research Online* 2014, 4. In particular, according to the Authors, the 5Vs are made

from data is their "smart analysis", that is, through advanced methods such as predictive analytics, data mining and data science, which make use of technologies such as cloud computing, sensors of the internet of things, machine learning, artificial intelligence, etc.[3].

The phenomenon of Big Data intersects with comparative law under several important profiles. First of all, the comparative approach leads to identify the peculiar characteristics of the data through a conceptual framework of the same in the perspective of other disciplines, in particular information technology and economics[4].

Then, comparative law is of particular importance in order to ensure that data management and analysis services can be developed and provided beyond national borders, while highlighting requirements and limits for this to take place in compliance with the various current regulations.

Furthermore, the collection and analysis of "Legal Big Data" can allow observing the evolutionary patterns and paths of the law, foreseeing or adopting jurisprudential decisions, drafting and applying laws or regulations, based on solid argumentative and comparative elements.

I. DATA DRIVEN INNOVATION AND BIG DATA: CONCEPTUAL FRAMEWORK IN MULTIDISCIPLINARY PERSPECTIVE

In computer science, systems are regularly modeled using data, information and knowledge[5]. This is a model that has long been used. The first version of the model is static, that is, a

up of: 1) volume of available data which is enormous and coincides with the overall size of the phenomenon; 2) variety of data and unstructured data sets or heterogeneity of sources and formats; 3) the speed with which the databases are fed and the high frequency with which the data circulate from a point of origin to a collection point; 4) value of the data which depends on the economic potential and the social value that can be attributed to the data as new production factors; 5) veracity of the data, or their authenticity and reliability.

[3] See: Gellert R., *Data Protection and Notions of Information: A Conceptual Exploration*, 2018. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3284493; Ackoff R., From Data to Wisdom. *Journ. Appl. Syst. Anal.* 1989, 3; Gandomi A., Haider M., Beyond the hype: Big data concepts, methods and analytics, cit., 140, who state that: "Big data is worthless in a vacuum. Its potential value is unlocked only when leveraged to drive decision making. To enable such evidence-based decision making, organizations need efficient processes to turn high volumes of fast-moving and diverse data into meaningful insights".

[4] On the importance of the other disciplines for the purpose of comparative analysis, see among others: Michaels R., Transnationalizing Comparative Law. *Maastr. Journ. Eur. Comp. Law* 2016, 352; Spamann H., Empirical Comparative Law. *Ann. Rev. Law Soc. Sc.* 2015, 131; Reitz J.C., How To Do Comparative Law. *Amer. Journ. Comp. Law* 1998, 617, who highlighted how one of the benefits of using the comparative method is "the tendency to force the researcher to expand the analysis to include the whole legal system and its relationship with the rest of human culture and its material and spiritual context in order to understand the differences and similarities observed"; one may see also: A. Stazi, *Biotechnological Inventions and Patentability of Life. The US and European Experience*, Cheltenham: Edward Elgar, 2015, 258 ff.

[5] See: European Commission, *The economics of ownership, access and trade in digital data*, JRC Digital Economy Working Paper 2017-01. https://ec.europa.eu/jrc/sites/jrcsh/files/jrc104756.pdf, 6 ff.; Ronquillo C., Currie L.M., Rodney P., The Evolution of Data-Information-Knowledge-Wisdom in Nursing Informatics. *ANS Adv Nurs Sci.* 2016, E1.

hierarchy: knowledge is made up of information that is composed of data. But this model is not adequately explanatory.

In practice, data is essentially acquired through three main channels. First of all, some data are offered voluntarily, that is, intentionally provided by an individual-user of services or products. Secondly, there are the observed data, or behavioral data acquired automatically by the activities of users or machines. Finally, some data are deduced, transforming in a non-trivial way data provided voluntarily and/or observed while they are still in relationship with a specific individual or machine[6].

With respect to the use of the data collected in this way, four modes are currently configured, namely: non-anonymous use of data at individual level, anonymous use of data at individual level, aggregated data and contextual data. Non-anonymous use of data at the individual level is typically aimed at providing services to the individual. The anonymous use of data on an individual level does not take place directly for this purpose, but for example to train machine learning algorithms and/or for purposes unrelated to those for which the data were originally collected. The aggregate data refer to more standardized data that has been irreversibly aggregated, such as for example national statistical information, sales data, etc. Finally, contextual data refer to data that does not derive from data at an individual level, such as information on the road network, satellite data etc.[7]

Hence, data is structured and organized through computation - human or automatic - in information, which in turn is interpreted and used as knowledge. Here emerges a value chain, in which data are organized into information through the use of tools and processing methods, and this information is then interpreted and used as knowledge through the use of attention[8].

In the debate on the subject, metaphors are frequently used on the value of data, including first of all the one according to which "data is the new oil"[9]. Economic analysis, on the other hand, leads to believe that this statement cannot be shared for three reasons, namely:

---

[6] In this regard, see: European Commission, *Competition policy for the digital era - A report by Jacques Crémer, Yves-Alexandre de Montjoye, Heike Schweitze*r, April 2019. http://ec.europa.eu/competition/publications /reports/kd0419345enn.pdf, 24-25.

[7] See again: *European Commission, Competition policy for the digital era - A report by Jacques Crémer, Yves-Alexandre de Montjoye, Heike Schweitzer,* cit., 25 ff.

[8] Thus: Gandomi A., Haider M., *Beyond the hype: Big data concepts, methods, and analytics*, cit., 137-144; Drexl J., *Designing Competitive Markets for Industrial Data - Between Propertisation and Access*, Max Planck Institute for Innovation & Competition, Research Paper No. 16-13, October 2016. https://ssrn.com/abstract=2862975, 17; Varian H.R., Beyond Big Data. *Busin. Econ.* 2014, 27.

[9] See: Newman N., Search, Antitrust and the Economics of the Control of User Data. *Yale Journ. Reg.*, 2014, 436; Lerner A.V., *The role of Big Data in online platform competition*, 2014. https://papers.ssrn.com/abstract=2482780, 3.

i) There is no scarcity of data, on the other hand, they are not only abundant but increasingly abundant. An asset that is increasingly abundant behaves differently than one of which there is scarcity when it comes to value. The data is produced every second. In fact, with the passage of time, the amount of data in circulation increases exponentially. It is a mechanism profoundly different from oil[10].

ii) Data is not "consumed". When a person uses gasoline for her car, another person cannot use that same gasoline for her own. She cannot use even a small part of it. Oil and petrol cannot be shared. On the contrary, as regards the data, the user can transfer them to different service providers.

iii) Data is a non-rival asset: it can also be used by several people at the same time[11].

As a result, the data often has a very particular utility curve, with decreasing returns to scale. If you don't have any data, even a minimal amount is useful; then, as the amount of data increases, their usefulness begins to decrease. One of the reasons why the marginal utility of data decreases is that the number of variables in a set grows linearly, while correlations grow exponentially. It is the so-called Big Data tragedy, where more data will involve spurious or even insignificant correlations and significantly higher costs for identifying significant correlations[12].

While the value of the data in itself is small, it acquires value when it is organized in such a way as to obtain information from them or where knowledge can be extracted from that. If we look only at the raw data, we do not find the value to which it can instead give rise if properly analyzed. The key to extracting value from data is to have innovative ideas on how to use it and computer scientists and data scientists able to develop effective software and artificial intelligence systems to obtain useful information[13].

In this scenario, what is of central importance, beyond the amount of data available, the programming and processing skills, are the rhythm of learning and the ability to combine data analysis, machine learning and human learning in the best possible way[14]. Human

[10] In this sense, for example, the amount of data produced by the machines increases exponentially. Regarding the differences between data and oil, see widely: Frank M., Roehrig P., Pring B., *What To Do When Machines Do Everything: How to Get Ahead in a World of AI, Algorithms, Bots, and Big Data*, Hoboken: Wiley, 2017, 65 ff

[11] See: Gandomi A., Haider M., *Beyond the hype: Big data concepts, methods, and analytics*, cit., 137-44; J. Drexl, Drexl J., *Designing Competitive Markets for Industrial Data - Between Propertisation and Access*, cit. p. 28; Varian H.R., *Beyond Big Data*, cit.

[12] In this sense, see: Taleb N.N., *Antifragile: Things That Gain from Disorder*. New York: Random House, 2012.

[13] In this regard, see: European Commission, *Competition policy for the digital era - A report by Jacques Crémer, Yves-Alexandre de Montjoye, Heike Schweitzer*, cit., 27 ff.; Pitruzzella G., Big Data and Antitrust enforcement. *Riv. ital. Antit.* 2017, 80.

[14] See: Stiglitz J.E., Greenwald, B.C. Creating a Learning Society: A New Approach to Growth, Development, and Social Progress, New York: Columbia University Press, 2015.

knowledge specific to a particular field, therefore, still retains a high value in the data economy, not least because it is necessary to ask the right questions.

II. THE QUALIFICATION OF CROSS-BORDER TRAFFIC REQUIREMENTS

In the Big Data scenario, it is also of particular importance to ensure that data management and analysis services can be developed and provided beyond national borders, in compliance with the different regulations in force in the various legal systems[15].

In this perspective, first of all, with regard to data ownership, the European continental model is being questioned, because whatever the nature of the rights that individuals can claim on their data, their applicability seems, in most cases, illusory. The European model must also compete with other models including, first and foremost, the US model which, in practice, attributes the possibility for companies to make the most of data by drawing its value[16].

In this context, also in consideration of what has been noted above, the most relevant issues pertain not so much to the personal data that belong to individuals but rather to the management of the vast databases that are among the main components of the so-called datasphere.

A similar approach seems to be accepted also in those Asian systems where the fundamental question is not to protect individual rights, but rather to guarantee the general interest of the community. The model that is based on the protection of personal data, therefore, seems to be decreasing in terms of size, numerically and geographically[17].

In Japan, for example, non-personal and anonymized data do not have specific legal protection. Apart from contract law, individual data are not protected, while structured datasets have the possibility of being protected by the regulation on trade secrets and

---

[15] In this regard, see among others: Drexl J., Legal Challenges of the Changing Role of Personal and Non-Personal Data in the Data Economy, in De Franceschi A., Schulze R. (ed.), *Digital Revolution - New Challenges for Law*, Cambridge: Intersentia, 2019, 5 ff.; Falce V., Copyrights on data and competition policy in the Digital Single Market Strategy. *Riv. ital. Antit.* 2018, 33.

[16] See: Sylvestre Bergé J., Grumbach S., Zeno-Zencovich V., The 'Datasphere', Data Flows beyond Control, and the Challenges for Law and Governance. *Eur. Journ. Comp. Law Gov.* 2018, 159; A. Coos A., *EU vs US: How Do Their Data Privacy Regulations Square Off?*, 17 January 2018. https://www.endpointprotector.com/blog/eu-vs-us-how-do-their-data -protection-regulations-square-off; see also the analysis conducted by William Fry, *Europe for Big Data*. https://www.williamfry.com/docs/default-source/reports/william-fry-europe-for-big-data-report.pdf?sfvrsn=2, 2 ff.

[17] In this sense, see: Sylvestre Bergé J., Grumbach S., Zeno-Zencovich V., *The 'Datasphere', Data Flows beyond Control, and the Challenges for Law and Governance*, cit., 163; Burk D.L., *Privacy and Property in the Global Datasphere*, Minnesota Legal Studies Research Paper No. 05–17. https://papers.ssrn.com/sol3/papers.cfm? Abstract_id = 716862.

copyright. This is due to the fact that structured datasets have characteristics that differentiate them from individual data and are substantiated in the fact that corporate entities apply their own vision and intuitions to that data.

Although the protection regime granted in Japan on trade secrets is not significantly different from that in force in the European Union and in the United States, the regulation on the protection of copyright presents some differences with respect to the sui generis rights granted in the EU and at the level of requirements in the United States. In Japan, moreover, protection against illegal acts is not effective since it does not extend to individual data and the Supreme Court attributes this possibility as well as the related protection only for data sets protected by copyright and by the regulation on trade secrets[18]. On the other hand, contracts offer greater flexibility and therefore the possibility of protecting individual data, unless a contract violates the rules for the protection of public order.

Since non-personal and anonymized data do not have specific legal protection, the Japanese government is considering the implementation of new policies, both legislative and non-legislative, similarly to what the European Commission is doing[19]. Recently, in fact, EU has successfully concluded the talks with Japan aimed at adopting a mutual adequacy decision, by which it is agreed to recognize the respective data protection systems as "equivalent"[20].

With respect to the issue of data circulation, in terms of defining the related legal conditions, the general system set up by the European Union for the protection of personal data is very ambitious, since it aims to regulate the flow of personal and non-personal data, while at national level the legal bases related to the circulation of data vary considerably from one area to another, depending on the nature of the data and the justification for their confidentiality or control[21].

---

[18] Thus: Kaburaki Y., *Legal Protection for Non-Personal Data in Japan - Comparative Perspective with the EU and the US*, 2017. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3208068, 51.

[19] See: Kaburaki Y., *Legal Protection for Non-Personal Data in Japan - Comparative Perspective with the EU and the US*, cit., 47-50.

[20] See the press release on the conclusion of the talks between the European Union and Japan on mutual adequacy, available on: http://europa.eu/rapid/press-release_IP-18-4501_it.htm, as well as the decision available on: https://ec.europa.eu/info/sites/info/files/draft_adequacy_decision.pdf. In particular, the agreement on mutual adequacy is aimed at creating a secure data transmission space, based on a high level of protection of personal data. European citizens whose personal data will be transferred to Japan will benefit from strong data protection, in line with EU rules. The agreement will also complement the EU-Japan Economic Partnership Agreement, and European businesses will benefit from the unhindered flow of data with Japan, a key trading partner, and privileged access to 127 million Japanese consumers. Under the EU General Data Protection Regulation, an adequacy decision is the most direct way to ensure the security and stability of data flows.

[21] In this regard, see: Lazaro C., Le Métayer D., The Control over Personal Data: True Remedy or Fairy Tale?, *SCRIPTed*, 2015, 3; Dechesne F. et al., A comparison of data protection legislation and policies across the EU, *Comput. Law Sec. Rev.* 2018, 234.

The cross-border circulation of data poses specific problems of international law, whether it is a matter of regulating such flows through the adoption of international agreements or of taking unilateral decisions of international scope[22].

For example, with regard to the transatlantic data transfer, the so-called Safe Harbor and the following Privacy Shield have provided for a mechanism to support transatlantic trade through which companies operating on both sides of the Atlantic must comply with certain data protection requirements when transferring personal data from the European Union and from the Switzerland to the United States[23].

In particular, in order to join the Privacy Shield program, organizations must guarantee by self-certification that they respect the parameters of the agreement, also publicly committing to respect this constraint. The new regulation provides for stringent protection obligations for companies that transfer data and specific tools for the protection of people. Thanks to this agreement, moreover, for the first time, the US Administration has formally guaranteed that public authorities' access to personal data will be subject to compliance with a series of specific and defined limits, guarantees and control mechanisms.

Although membership of the Privacy Shield is voluntary, once an organization publicly undertakes to comply with the requirements therein, that commitment becomes enforceable under United States law.

Regulation (EU) 2016/679, so-called GDPR[24], on the other hand, is already demanding a higher level of legislative convergence. An incomplete study of over 30 countries outside Europe, in Africa, Asia and elsewhere, shows how six new "GDPR principles" have already

[22] For an overview of the issues and regulatory options, see: Casalini F., López González J., *Trade and Cross-Border Data Flows*, OECD Trade Policy Papers No. 220. Paris: OECD Publishing, 2019. https://www.oecd-ilibrary.org/trade/trade-and-cross-border-data-flows_b2023a47-en; Zeno-Zencovich V., Free flow of data. Is international law the appropriate answer?. Forthcoming in: Fabbrini F., Celeste E., Quinn J. (eds.), *Data Protection Imperialism and Digital Sovereignty*. Oxford: Hart Publishing, 2020.

[23] See: Commission Decision of 26 July 2000 pursuant to Directive 95/46/EC of the European Parliament and of the Council on the adequacy of the protection provided by the safe harbour privacy principles and related frequently asked questions issued by the US Department of Commerce, 2000/520/EC, OJEU L 215/7; Commission Implementing Decision (EU) 2016/1250 of 12 July 2016 pursuant to Directive 95/46/EC of the European Parliament and of the Council on the adequacy of the protection provided by the EU-U.S. Privacy Shield, C/2016/4176, GUUE L 207/1; *Maximillian Schrems v Data Protection Commissioner* [2015] CJEU C-362/14. In doctrine, see: Sylvestre Bergé J., Grumbach S., Zeno-Zencovich V., *The 'Datasphere', Data Flows beyond Control, and the Challenges for Law and Governance*, cit., p. 167-168; Miller R.A. (ed.), *Privacy and Power. Transatlantic Dialogue in the Shadow of the NSA-Affair*, Cambridge: Cambridge University Press, 2017; Svantesson D., Kloza D. (eds.), *Trans-Atlantic Data Privacy Relations as a Challenge for Democracy*, Antwerp: Intersentia, 2017.

[24] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), OJ L 119, 4.5.2016. For a systematic comment on the provisions of the Regulations, see among others: Kuner C. et al. (eds.), *The EU General Data Protection Regulation (GDPR). A Commentary*. Oxford: Oxford University Press, 2020.

been adopted by at least ten countries, and all the new principles enshrined in the GDPR only by one country.

The GDPR is also spreading through a new global phenomenon which is represented by the voluntary convergence of companies for which there is no legal obligation to the aforementioned regulation[25].

Moreover, there is a phenomenon of regulatory convergence also at a global level, as evidenced by the Data Protection Convention no. 108, which was drawn up by the Council of Europe and to which non-European countries have joined since 2011[26]. Recently, it has been "modernized" with new standards, including many, but not all those provided by the GDPR.

On the other hand, there are currently a number of potential obstacles to the adoption of high standards globally. First, countries could make commitments through regional agreements that require lower standards, including to allow for the export of data, and therefore could legislate to implement it. Secondly, free trade agreements can impose stricter bans than the global GATS agreement and restrictions on the export of personal data, creating conflicting standards[27].

In any case, it is appropriate to highlight how the legal regime of a dataset may be different from the regime of data intended as single entities. Therefore, the processing of a bulk of Big Data could coexist with the processing of individual data on the basis of a different regime[28].

The legal profiles of this coexistence, on the other hand, also following the European Union regulatory interventions on personal and non-personal data, remain widely debated. Lastly, in particular, with reference to the Regulation on the free flow of non-personal data in Europe, two main concerns were raised, namely on the one hand the indeterminacy and dynamism of the notion of non-personal data as a reference point of the legislation; on the

[25] In this regard, see: Greenleaf G., *Global Convergence of Data Privacy Standards and Laws: Speaking Notes for the European Commission Events on the Launch of the General Data Protection Regulation (GDPR) in Brussels & New Delhi*, UNSW Law Research Paper No. 18-56, May 2018. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3184548, 3.

[26] See the 1981 Strasbourg Convention, or Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data, available at the following link: https://www.coe.int/it/web/conventions/full-list/-/conventions/treaty/108. On 18 May 2018, the Council of Europe adopted a protocol amending the text of the Convention aimed at modernizing it to provide a legal framework more suited to a time when violations of the right to data protection have become a major concern. . In particular, the protocol provides a robust and flexible legal framework to facilitate the flow of data across borders and provide effective guarantees.

[27] See: Greenleaf G., *Global Convergence of Data Privacy Standards and Laws: Speaking Notes for the European Commission Events on the Launch of the General Data Protection Regulation (GDPR) in Brussels & New Delhi*, cit., 4.

[28] In this sense, see again: Sylvestre Bergé J., Grumbach S., Zeno-Zencovich V., *The 'Datasphere', Data Flows beyond Control, and the Challenges for Law and Governance*, cit., 168-69.

other, the possible occurrence of conflicts with the GDPR and with the fundamental right to the protection of personal data[29].

On the other hand, the EU Regulation 2018/1807 on non-personal data[30] appears to be based on a cautious approach aimed at reconciling the objective of promoting the free movement of the same with the fundamental rights at stake. It is intended, in fact, to ensure the free movement of data within the Union by focusing in fact on the elimination of data localization obligations, on the regulation of making data available to the competent authorities, and on the impulse to develop codes of conduct for the implementation of data portability.

III. "LEGAL BIG DATA" AS A NEW INSTRUMENT OF COMPARATIVE LAW

Finally, from a different point of view, the application of today's data collection and analysis technologies in the legal field can give rise to "Legal Big Data", collected and analyzed for descriptive, predictive or prescriptive purposes, that is, to observe the evolutionary patterns and paths of the law, to foresee or adopt jurisprudential decisions, to elaborate and apply laws or regulations, based on solid argumentative and comparative elements[31].

As is known, legal discipline has been at the center of a long debate regarding the possibility of considering law as a science or not[32]. This debate may recall the fact that the scientific method, based on an interactive process - formulating hypotheses, developing testable predictions, testing predictions, developing general theories - does not fit the law, which is essentially deontic and non-descriptive in nature, and in turn has developed its own *sui generis* methods[33].

---

[29] See: Graef I. et al., *Feedback to the Commission's Proposal on a Framework for the Free Flow of Non-Personal Data*, 2018. https://ssrn.com/ abstract = 3106791.

[30] Regulation (EU) 2018/1807 of the European Parliament and of the Council of 14 November 2018 on a framework for the free flow of non-personal data in the European Union, OJ L 303, 28.11.2018.

[31] Similarly to what is happening in other disciplines. See, among others: Custers B., Methods of data research for law. In: Mak V. et al. (eds.), *Research Handbook on Data Science and Law*. Cheltenham: Edward Elgar, 2018, 355 ff .; Van Ettekoven B.J., Prins C., Data analysis, artificial intelligence and the judiciary system, ivi, 425 ff.; Leeuw F.L., Schmeets H., *Empirical Legal Research, A Guidance Book for Lawyers, Legislators and Regulators*. Cheltenham: Edward Elgar, 2016, 8 ff.

[32] On the theme of the relationship between science and law (and legal comparison), see: Fagan F., Big Data Legal Scholarship: Toward a Research Program and Practitioner's Guide. *Virg. Journ. Law Tech.*, 2016, 7; as well as: Langford P., Bryan I., McGarry J. (eds.), *Kelsenian Legal Science and the Nature of Law*. Springer: Berlin, 2017; Coyle S., Pavlakos G. (eds.), *Jurisprudence or Legal Science? A Debate about the Nature of Legal Theory*. Oxford: Hart Publishing, 2005; Mertz E., *The Role of Social Science in Law*. Farnham: Ashgate, 2008; Faignman D.L., To Have and Have Not: Assessing the Value of Social Science to the Law as Science and Policy, *Emory Law Journ.*, 1989, 1007.

[33] See, among others: Taekema S., Van Klink B., De Been W. (eds.), *Facts and Norms in Law: Interdisciplinary Reflections on Legal Method*. Cheltenham: Edward Elgar, 2016; Bruncken E. et al. (eds.), *Science of Legal Method*. Farmington Hills: Gale, 2013; Moenssens A., Scientific Method Compared to Legal Method. In: Jamieson A., Moenssens A., *Wiley Encyclopedia of Forensic Science*, Vol. 5, Chichester: Wiley,

Although it is commonly said that law falls within the social sciences, in fact, there is a fundamental difference between what is commonly included in social sciences, such as economics or sociology, and law. The first are of a descriptive nature, in the sense that in order to achieve a result it is necessary to study, investigate, measure social phenomena and extract from such researches a general rule that describes what has been observed. Law, on the other hand, is essentially prescriptive, or deontic. It doesn't say how things are, but how they should be.

Therefore, while datification is an essential, though not the only, characteristic of the social sciences, this is not required for law, which expresses immeasurable values. The construction of a legal system, the logic through which the rules are applied and the relationships that must be established between the rules cannot be dated, in the sense that they cannot be described in numerical terms and even if this were possible it would be of little significance[34]. Likewise, this view can also recall the fact that much of legal research is generally not reproducible. Indeed, legal research conducted at the doctrinal level consists in expressing interpretative opinions developed through legal arguments and reasoning.

As is known, there are many factors that contribute to the increase in the volume of legal acts, but two stand out in particular: multi-level governance and socio-economic developments, driven by the influence of additional forces such as technological innovation, which bring existing legal frameworks under discussion and increase the need to develop new and more adequate regulatory responses[35].

Classical comparative law is of a doctrinal nature[36], as is the legal analysis provided in the comparative reports provided, for example, for the European Commission, even if the latter often uses questionnaires that can be integrated by empirical methods such as the use of surveys on a limited number of interested parties[37].

On the other hand, all in all, there is more law and more legal research in today's world than at any other historical moment, and with the help of technology this trend is on the rise.

2009, 2296. On the relations between law and logic, see: Navarro P.E., Rodríguez J.L., *Deontic Logic and Legal Systems*, Cambridge: Cambridge University Press 2014.

[34] See: Zeno-Zencovich V., Through a Lawyer's Eyes: Data Visualization and Legal Epistemology, in Degravem E. et al. (eds.), *Law, Norms and Freedoms in Cyberspace - Droit, Normes et Libertés*. Liber Amicorum Yves Poullet, Brussels: Larcier, 2018, 462 ff.

[35] In this regard, see: Fagan F., Successor Liability from the Prospective of Big Data. *Virg. Law & Bus. Rev.* 2015, 391; Macey J., Mitts J., Finding Order in the Morass: The Three Real Justifications for Piercing the Corporate Veil. *Corn. Law Rev.* 2014, 99.

[36] See: Hahn T., *From Big data to Smart data*, Siemens Future Forum 2014. https://w3.siemens.com/topics/global/en/events/hannover-messe/program/Documents/pdf/ Smart-Data-to-Business-Michal-Skubacz.pdf, 13.

[37] On such instruments, see in general: Keman H., Woldendorp J.J. (eds.), *Handbook of Research Methods and Applications in Political Science*. Cheltenham: Edward Elgar, 2016, 262 ff.; Dunn W.N., *Public Policy Analysis*, New York: Routledge, 2016, 65 ff.

Technology has certainly affected legal research and related methodologies, and this finding has been highlighted by scholars in one way or another for years now[38]. The benchmarking or indexing exercises, for example, show how the law can be transformed into data and further researched in order to integrate the most classic legal investigations[39].

In this perspective, while the elaboration of a legal norm or of a judicial practice can be based on non-legal texts and other types of data such as those of the large statistical repertories, the function of the latter when used in the context of the legislation is generally focused on the justification of the new rules[40].

On the contrary, data analysis methods can also be used to develop factual bases for the application of pre-existing rights established previously by law. Thus conceived, data analysis methods could be used to develop the factual basis aimed at strengthening the assertion of pre-existing rights. The object of this type of analysis are the repositories of the fact itself, for example a historical database of decisions that span various areas of protection of rights[41].

Judicial decision-making models can also be a useful tool for researchers to establish what is the positive law that exists in practice, or law in action, and how it should be interpreted in specific cases[42].

In general, therefore, data analysis methods can be used to describe theory and legal rules. What has emerged scientifically so far, however, at least in terms of Big Data's ability to describe legal rules, is essentially taxonomic. On the other hand, in terms of the ability to describe legal theory, Big Data does not appear to be as useful, indeed it has been claimed that they would rather provide another empirical method for falsification[43].

However, taxonomic studies are important, especially where there is a doctrinal debate concerning the series of events that leads to the judicial application of a particular doctrine,

[38] See: Smith T., The Web of Law. *San Diego Law Rev.*, 2007, 309; Katz D.M., Bommarito M.J., Measuring the Complexity of the Law: The United States Code. *Artif. Intell. Law* 2014, 337.

[39] In this regard, see: Siems M., Numerical Comparative Law - Do We Need Statistical Evidence in Law in Order to Reduce Complexity?. *Card. Journ. Intern. Comp. Law* 2005, 521.

[40] See again: Smith T., *The Web of Law*, cit., 309 ff.

[41] See: Micklitz H.W., An Expanded and Systematized Community Consumer Law as Alternative or Complement ?. *Eur. Bus. Law Rev.* 2002, 583.

[42] See Custers B., *Methods of data research for law*, cit., 361 and 364 ff., who mentions among others the example of the Watson data analysis system, developed by IBM, which through artificial intelligence interprets the questions in natural language and answers these questions after having consulted a collection of digital sources such as encyclopedias, books, journals, scientific publications and websites; on law in action and law in books, v. *ex multis*: Halperin J.L., Law in Books and Law in Action: The Problem of Legal Change. *Maine Law Rev.* 2011, 46.

[43] Kitchin R., Big Data, new epistemologies and paradigm shifts. *Big Data & Soc.* April-June 2014, 4; the Author claims that the way to analyze data, even "(w)hilst this empiricist epistemology is attractive, it is based on fallacious thinking with respect to the four ideas that underpin its formulation". In particular, the author cites other scholars and notes that Big Data and the related analysis techniques mark the transition to a new era of knowledge production which coincides with "the end of theory".

or where there is a continuous change in the legal logic underlying it. For example, Macey and Mitts have been able to find a reading key and establish an authoritative taxonomy in the *mare magnum* of over 9,000 decisions on the so-called corporate veil, in which corporate law judges have put aside the limited liability of the company and held the shareholders or directors personally responsible for the company's actions or debts[44].

Thus, the data analysis methods offer classification techniques that severely limit the analyst's subjective prejudices and therefore can promote the resolution of legal issues in a shared way. Through a method known as "topic modeling", the analyst can use an algorithm to simultaneously examine a practically unlimited number of judicial decisions, without having to specify the reasons, opinions, motivations or characteristics underlying the case[45].

The algorithm itself does a job without knowing its purpose. The analyst specifies the number of topics that must be part of the model and the algorithm produces a list with a certain number of keywords. The analyst must then create categories for classification based on the contents of the list. The creation of a category requires a certain dose of subjectivity, which however is strictly limited when compared with other methods and understood in a scientific sense[46].

Doctrinal taxonomies created on the basis of data analysis appear more advanced than their counterparts built with traditional methods. Therefore, through large datasets and algorithmic models, the analyst can severely limit subjectivity through an automated analysis of the words contained in the judicial texts. The results are scientifically reproducible. The large data taxonomies, therefore, seem to represent progress in the search for objectivity and may be able to generate greater consensus in legal doctrine as a result[47].

In fact, the essential requirement for all attempts to find correlations is to collect the necessary data. Any researcher who tries to study legal developments by gathering all sorts of relevant statistics soon discovers that these are very difficult to obtain, due to two different factors: on the one hand, relevant statistics from a substantially legal perspective are scarce; on the other hand, the research databases themselves and the statistics are rare and difficult to verify[48].

---

[44] See Macey J., Mitts J., *Finding Order in the Morass: The Three Real Justifications for Piercing the Corporate Veil*, cit., 113 ff.

[45] See: Fagan F., *Big Data Legal Scholarship: Toward a Research Program and Practitioner's Guide*, cit. 15; Young D.T., How Do You Measure a Constitutional Moment? Using Algorithmic Topic Modeling to Evaluate Bruce Akerman's Theory of Constitutional Change. *Yale Law Journ.* 2013, 1990.

[46] In this regard, see: Sedgewick R., Flajolet P., *An Introduction to the Analysis of Algorithms*. Boston: Addison-Wesley, 2013, 465.

[47] So again: Fagan F., *Big Data Legal Scholarship: Toward a Research Program and Practitioner's Guide*, cit., 13 ff.

[48] See: Goanta C., Big Law, Big Data. *Law Meth., Special Issue - Comp. Law*, 2017. http://www.lawandmethod.nl/tijdschrift/lawandmethod/2017/10/lawandmethod-D-17-00007, 9.

In the scenario so far highlighted, therefore, an interesting option seems to be that of considering the collection and analysis of Legal Big Data as a new instrument of comparative law, for the purpose of investigating and outlining the models for processing, interpretation and application of the law in the various legal systems.

In other words, the most complex and voluminous legislation and legal decisions or interpretations can be codified in the form of quantitative observations on the basis of the data available in comparative law, or they can be analyzed with the help of new research technologies[49].

In this context, the rules or jurisprudential decisions of the different countries can be considered as data that must be analyzed and processed according to a specific methodology. Observing the law and using research to extrapolate data for further analysis can revolutionize the way legal research is understood. Viewing the underlying data can lead to the observation of new models and the development of new conclusions regarding legal developments. The potential for this has already been explored through the use of innovative methods such as the use of legal research software for so-called network analysis, a method used especially in information technology and sociology that maps and measures the relationships between people, groups, computers or information[50].

In this sense, for example, a new methodology has been proposed to measure the convergence effect of the European harmonization policies of the discipline on sales contracts to consumers: the "Convergence Index", which was developed in order to fill a gap in the consumer law literature of the European Union and showed the usefulness of adopting measurement indices[51].

The starting point was that the methods used to understand European consumer contract law were not sufficient to address the essential weaknesses of the current regulatory framework, since the classic methods of comparing the legal provisions and understanding

[49] See again: Goanta C., *Big Law, Big Data*, cit., 13 ff.

[50] In this regard, see: Netherlands eScience Center, *How can network analysis lead to a new way of studying court decisions?*, 2017 https://medium.com/escience-center/how-can-network- analysis-lead-to-a-new-way-of-studying-court-decisions-686ccf4d46aa; Lupu Y., Voeten E., Precedent in International Courts: A Network Analysis of Case Citations by the European Court of Human Rights. *Brit. Journ. Polit. Sc.*, 2012, 413; Fowler J.H., Network Analysis and the Law: Measuring the Legal Importance of Supreme Court Precedents, *Polit. Anal.*, 2007, 324.

[51] The methodology is based on Siems' work on "comparative numerical law", considered as a translation of the law into numbers (Siems M., *Comparative law*. Cambridge: Cambridge University Press, 2014, 146-187; and Id., *Numerical Comparative Law - Do We Need Statistical Evidence in Law in Order to Reduce Complexity?*, cit.), and on the Consumer Law Compendium database established by Schulte-Nölke (Schulte-Nölke H., Twigg-Flesner C., Ebers M., *EC Consumer Law Compendium: The Consumer Acquis and its Transposition in the Member States*. Berlin: De Gruyter, 2009), also inspired by the OECD Manual of Composite Indicators (*Handbook on constructing composite indicators: methodology and user guide*, 2008. http://www.oecd.org/std/42495745.pdf).

the related similarities and differences were not suitable for the analysis which must be carried out on a vast quantity of observations.

Therefore, the Convergence Index is an aggregator that focuses on measuring the level of legal convergence that occurs following the transposition of European directives into national legal systems[52]. The outcome of the analysis carried out by using this index allowed the performance of the selected Member States to be displayed on certain European directives on consumer contracts, with results of considerable interest[53].

In the context of comparative law, therefore, the possibility of treating legal information such as legislation, jurisprudence and doctrine as data through the tools applied to Big Data could represent an innovative and effective solution. Comparative law has so far been based on a small-scale comparison, and the greater the number of jurisdictions chosen, the greater the research effort in terms of people, time and other resources.

From an accessibility point of view, the technical solutions that can be used in legal research already exist and can be easily adapted to the specific needs of researchers[54]. In this sense, consider for example the use of the aforementioned network analysis: while the traditional method of analyzing case law has been to summarize the number of judicial decisions consulted for a given legal problem, the analysis of the set of cases provides the opportunity to view an entire area with the help of technical support and to derive information based on a large and essentially exhaustive amount of data[55].

---

[52] The Index comprises thirteen separate indicators, seven of which reflect the level of European governance, while the remaining six reflect national legislation. The indicators were chosen on the basis of an in-depth comparative study that examined the selected directives, that is, 85/577/EEC on contracts negotiated away from business premises, 93/13/EEC on unfair terms in consumer contracts, 97/7/EC on the protection of consumers in distance contracts, 1999/44/EC on the sale of consumer goods, and 2005/29/EC on unfair commercial practices, and the selected Member States - Belgium, France, Germany, Ireland, the Netherlands, Romania, and the United Kingdom - to understand what influences the convergence of legal systems in the field of European consumer law. As regards the object of the comparison, the Index reflects only the transposition process within the regulatory framework and does not extend to implementation, i.e. the application of European standards by courts or national professionals (because on this the sources of information are inconsistent).

[53] See: Goanta C., *Big Law, Big Data*, cit., 9 ff. In a critical perspective, see: Siems M., The End of Comparative Law. *Journ. Comp. Law*, 2007, 133; Mattei U., Some Realism about Comparativism: Comparative Law Teaching in the Hegemonic Jurisdiction. *Amer. Journ. Comp. Law*, 2002, 87; Reimann M., The End of Comparative Law as an Autonomous Subject. *Tul. Eur. Civ. Law Forum*, 1996, 49. Recently, the matter was then subject to a new regulatory intervention with the Directive on strengthening the application and modernization of EU consumer protection rules, aimed at ensuring effective sanctions and clear rules to contrast the quality differences in the goods of consumption, and greater transparency for consumers who shop online either through a cash payment or by providing personal data in exchange for digital content or services. Regarding data issues in the Directive approval process, see: Drexl J., *Legal Challenges of the Changing Role of Personal and Non-Personal Data in the Data Economy*, cit., 25 ff.

[54] In this sense, see: Michaels R., *Transnationalizing Comparative Law*, cit., 352 ff.

[55] See, for example, as regards the use of *network analysis* in the study of judicial decisions: Van Kuppevelt D., Van Dijck, Answering Legal Research Questions About Dutch Case Law with Network Analysis and Visualization. In: Wyner A., Casini G. (eds.), *Legal Knowledge and Information Systems*, Amsterdam: IOS Press, 2017, 95-100, and the project of the Netherlands eScience Center, *Case Law Analytics: Discovering new patterns in Dutch court decisions*, available on: https://www.esciencecenter.nl/projects/case-law-

IV. LEGAL BIG DATA AND JUSTICE: FROM PREDICTIVE JUSTICE TO PERSONALISED LAW?

Considering the law as a set of data to be collected and analyzed using today's technologies, therefore, can facilitate the understanding of legal issues and modify the legal practice allowing the development of new frameworks or forecasts[56], up to hypotheses leading to the possibility of a true and proper personalization of the law[57].

Through the analysis of Legal Big Data, it seems possible the so-called predictive justice, that is, to predict the results of court cases. In the United States, for example, when major Supreme Court rulings are awaited, there is much speculation in the media and by experts to predict whether the behavior of the judges - appointed by the President of the United States - will be in line with their political views or their decisions will be surprising.

In recent years, several experiments based on Big Data have provided remarkably precise predictions on the outcome of decisions, with correctness rates of forecasts between 70 and 75 percent. Even greater was the result achieved by a model developed by British and US researchers for the decisions of the European Court of Human Rights, with forecasts almost 80% correct[58].

Anticipating the outcome of court cases can be very useful for legal professionals, as it can help assess whether to present the case before a court. When the probability of success is

analytics. In practice, law firms seeking to keep up with others are already employing Big Data-based tools; see, for example, Juristat (www.juristat.com), which provides for the success of the patent process, or Ravel's Judges Analytics (https://www.ravellaw.com/judges), which allows users to map each decision taken by a particular judge.

[56] In this regard, see: Alarie B., Niblett A., Yoon A., How Artificial Intelligence Will Affect the Practice of Law. *Univ. Tor. Law Journ.*, 2018, 106; Brescia R.H. et al., Embracing Disruption: How Technological Change in the Delivery of Legal Services Can Improve Access to Justice. *Alb. Law Rev.*, 2014, 553.

[57] In this regard, see: Busch C., Implementing Personalized Law: Personalized Disclosures in Consumer Law and Data Privacy Law. *Univ. Chi. Law Rev.*, 2019, 309; Casey A.J., Niblett A., A Framework for the New Personalization of Law. ivi, 333; Busch C., De Franceschi A., Granular Legal Norms: Big Data and the Personalization of Private Law. In: Mak V., Tjong Tjin Tai E., Berlee A. (eds.), *Research Handbook on Data Science and Law*, cit., 408-24; Casey A.J., Niblett A., The Death of Rules and Standards. *Ind. Law Journ*., 2017, 1401; Hacker P., Personalizing EU Private Law: From Disclosures to Nudges and Mandates. *Eur. Rev. Priv. Law*, 2017, 651; Ben-Shahar O., Porat A., Personalizing Negligence Law. *New York Univ. Law Rev.*, 2016, 627; Busch C., The Future of Pre-contractual Information Duties: From Behavioral Insights to Big Data. In: Twigg-Flesner C. (ed.), *Research Handbook on EU Consumer and Contract Law*, Cheltenham: Edward Elgar, 2016, 221-40; Porat A., Strahilevitz L.J., Personalizing Default Rules and Disclosure with Big Data. *Mich. Law Rev.*, 2014, 1417.

[58] See Custers B., *Methods of data research for law*, cit., 360 ff.; Katz D.M., Bommarito M., Blackman J., A General Approach for Predicting the Behavior of the Supreme Court of the United States. *PLoS One*, 2017. https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0174698; Aletras N. et al., Predicting Judicial Decisions of the European Court of Human Rights: a Natural Language Processing Perspective. *PeerJ. Comp. Sc.*, 2016. https://peerj.com/articles/cs-93.

low, the legal advisor might perhaps rather recommend an agreement to her client, with the relative benefits also in terms of deflation of the litigation[59].

Other types of forecasts can also be useful in legal practice. For example, in combating crime it could be very useful to analyze criminal data to make predictions about who will commit crimes, where the crime will occur and which people, buildings and objects will be at risk as potential targets. This approach is usually called predictive police[60].

Another legal area in which Big Data-based predictions can be useful is probation or conditional liberty. In different countries, criminal courts base their decisions on whether or not there is a repeat offender and on assessing the risk of how likely the repeat offender is.

In many risk assessment models, prior convictions play an important role, resulting in the paradigm that "if you offend once, you are likely to commit offenses again; if you offend twice, you will surely commit offenses again and again"[61]. Although these reports can be statistically correct, at group level they can prevent any other conclusion for those individuals who are actually willing and able to improve their behavior. This type of use of Legal Big Data can therefore aggravate the difficulties that such profiled people already have in obtaining a job, an education and a betterment of life[62].

Another area in which the collection and analysis of Legal Big Data can contribute to the development of law is that of improving the law and regulations. The use of Big Data for the development and improvement of laws and regulations can contribute to socio-legal research work, developing theories and regulations based on solid evidence.

Legal Big Data can be useful to improve the contents of laws and regulations, but also to evaluate the acceptance of them by the community. By combining legal data with behavioral data, for example on social media etc., it becomes possible to evaluate which rules or which types of rules are better respected and/or can be easier to apply[63].

---

[59] See: Osbeck M.K., *Lawyer as Soothsayer: Exploring the Important Role of Outcome Prediction in the Practice of Law*, University of Michigan Public Law Research Paper No. 604, 2018. https://papers.ssrn.com/sol3 /papers.cfm?abstract_id=3138211.

[60] In this regard, see: Porcedda M.G., Wall D.S., Data science, data crime and the law. In: Mak V., Tjong Tjin Tai E., Berlee A. (eds), *Research Handbook on Data Science and Law*, cit., 214 ff.; Perry W.L. et al., *Predictive Policing: The Role of Crime Forecasting in Law Enforcement Operations*. Santa Monica: RAND Corporation, 2013.

[61] See: O'Neill K., *Weapons of Math Destruction*, New York: Crown, 2016; Harcourt B.E., *Against Prediction; Profiling, Policing and Punishing in an Actuarial Age*. Chicago: University of Chicago Press, 2006.

[62] In this regard, see: Custers B., *Methods of data research for law*, cit., 363-64.

[63] When trying to create rules that are better respected, the concept of "nudging" is highlighted, that is the offer of incentives, such as positive reinforcements or indirect suggestions, to try to make desirable behavior attractive, without forcing people to adopt that behavior or limit their freedoms. See, among others: Yeung K., 'Hypernudge': Big Data as a Mode of Regulation by Design. *Inf. Comm. Soc.*, 2016, 1; Thaler R.H., Sunstein C.R., *Nudge: Improving Decisions About Health, Wealth, and Happiness*. New York: Penguin Books, 2009.

In the near future, then, Big Data, super-human information processing skills and artificial intelligence could redefine the optimal complexity of legal rules and refine their content to a level of granularity that was previously unattainable. In such a scenario, granular or personalised legal rules could take into account the heterogeneity of the actor in a degree that impersonal laws are unable to do, allowing to repair the relationship between legal certainty and individual equity[64].

To make such a personalization of the law operational, of course, the legislator would have to establish at an abstract level the criteria on which the personalization will be based and define the consequences connected with the granularization of different personality profiles[65].

On the other hand, the collection and analysis of Legal Big Data gives rise to problematic issues and presents limits for the purposes of use in the legal field.

The most significant disadvantage is evidently that the adoption of data search methods limits the use of human intuition, since the volumes of Legal Big Data are usually too large to allow useful overviews and insights. The amount of data allows to increase the reliability of the results, but at the same time it can give rise to reliability problems since the results are statistical reports that describe probabilities, which could be of limited use for decision making, in particular in a legal context[66].

Another drawback is that in many situations data analysis methods can produce an abundance of patterns and relationships, most of which may not be new or useful. In many situations, especially in the legal field, it can be useful to know the underlying causal mechanisms. Finding out or even proving the underlying causality can be much more difficult and often requires further research.

A further question may arise because although large amounts of Legal Big Data are available, this data may have been collected in the past for other purposes. Aside from the legal restrictions on data reuse, this re-proposition could lead to problems, since when the data is used for new purposes it may no longer correspond exactly to these purposes. As a result,

---

[64] On the dynamic relations between legal certainty and individual equity, see: Fenwick M., Siems M., Wrbka S. (eds.), *The Shifting Meaning of Legal Certainty in Comparative and Transnational Law*, Oxford: Hart Publishing, 2017; Ávila H., *Certainty in Law*. Berlin: Springer, 2016; Neuhaus P.H., Legal Certainty versus Equity in the Conflict of Laws. *Law Cont. Prob.*, 1963, 795.

[65] In this sense: Busch C., De Franceschi A., *Granular Legal Norms: Big Data and the Personalization of Private Law*, cit., 6 ff., who refer to some application options of this approach, in the field of contract law as personalised disclosure, in the field of tort law as a standard of diligence, and in the field of family law and succession as predefined custom rules.

[66] See: Custers B., *Methods of data research for law*, cit., 374, and Id., Effects of Unreliable Group Profiling by Means of Data Mining. In: Grieser G., Tanaka Y., Yamamoto A. (eds.), *Lecture Notes in Artificial Intelligence*. Heidelberg: Springer, 2003, 290 ff.

many of the discovered models can be based on approximate indicators, rather than on the actual factors that determine the results. Furthermore, the reuse of data can give rise to more complex privacy problems than those inherent to their use[67].

Linked to this is the problem of so-called self-confirmation. Since the data search methods are based on available historical data, the results of the analysis will mainly concern the past rather than the future. Therefore, gradual changes can be discovered and used to make predictions about the future, but more disruptive changes can be much more difficult to consider[68].

A further example of a problem of prejudice is that of "self-fulfilling prophecies". A typical example of this is when law enforcement surveillance focuses on neighborhoods with ethnic minorities, with the result that databases fill up with people from those minorities. When the databases are later used to find patterns of which people are more prone to fall into criminal behavior, since the data was partial, people from these ethnic minorities will be able to be profiled as a more likely criminal behavior[69].

In these cases, the problematic analysis of the Legal Big Data emerges with respect to the transparency of the data selection, the analysis methods and processes, the chosen criteria and the algorithms used. All these aspects are relevant with regard to the due process principle, whose protection is conditional on the existence of guarantees relating to the transparency of the methods and processes adopted, the contestability of the results of the analysis, and the responsibility for the decisions taken[70].

Finally, granular or personalised legal rules could give rise to relevant issues in terms of interference with fundamental rights to the protection of personal data, freedom and equality. The privacy risks arise from the fact that such a regulatory approach is based on the

---

[67] On the subject of data reuse, see: Custers B., Ursic H., Big Data and Data Reuse: A Taxonomy of Data Reuse for Balancing Big Data Benefits and Personal Data Protection. *Intern. Data Priv. Law*, 2016, 4; Ursic H., Custers B., Legal Barriers and Enablers to Big Data Reuse - A Critical Assessment of the Challenges for the EU Law. *Eur. Data Prot. Law Rev.*, 2016, 209.

[68] See Hoffman S., Big Data Analytics: What Can Go Wrong. *Ind. Health Law Rev.*, 2018, 227.

[69] This is what emerged, for example, from recent US research which showed a clear prejudice between different population groups in the practice of assigning a score relating to the risk of recidivism. In this regard, see: Angwin J. et al., Machine Bias. *Pro Publ.*, May 2016. https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing; Larson J. et al., *How We Analyzed the COMPAS Recidivism Algorithm*, ivi. https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm; Petit N., *Artificial Intelligence and Automated Law Enforcement: A Review Paper*, 2018. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3145133, 5-6; Van Ettekoven B.J., Prins C., *Data analysis, artificial intelligence and the judiciary system*, cit., 442-43; Custers B., *Methods of data research for law*, cit., 373-74.

[70] On the subject, see: Petit N., *Artificial Intelligence and Automated Law Enforcement: A Review Paper*, cit., 10-11; Crawford K., Schultz J., Big Data and Due Process: Toward a Framework to Redress Predictive Privacy Harms. *Bost. Coll. Law Rev.*, 2014, 93; Moses L.B., Chan J., Using Big Data for Legal and Law Enforcement Decisions: Testing the New Tools. *Univ. New South Wales Law Journ.*, 2014, 643; Keats Citron D., Technological Due Process. *Wash. Univ. Law Rev.*, 2008, 1249.

collection of personal data and on the profiling of individuals. On this point, the European GDPR provides the right not to be subject to decisions of significant impact based on exclusively automated processes, including profiling, with exceptions among which in particular the consent of the interested party[71].

Personalised law, then, would abandon the equal application of general standards to all individuals, so the question arises as to whether this approach would be compatible with the fundamental principles of equality and freedom. In this sense, it is useful to remember that the process of identifying the relevant differences from a legal point of view is a regulatory process, and not merely an empirical one[72].

V. CONCLUSION

In conclusion, beyond the problematic issues mentioned above, the possible developments related to the Legal Big Data that have emerged recently are starting to transform, so far mainly in the United States but increasingly also in other countries, on one hand legal research, legislation and jurisprudence - to have them based on solid argumentative and comparative elements - but also, on the other hand, the dynamics and markets of the legal professions[73].

Although the analysis of Legal Big Data does not appear to be able to lead to results capable of equaling human reasoning, as we have seen, it can still have a significant impact on law and legal practice[74]. However, considering the issues and limitations mentioned above, it is unlikely that such developments will completely replace legal work, legal research or legislative drafting and jurisprudential application processes.

Law is a deontic-normative discipline, for which the added value provided by the intellectual and emotional abilities of the human being appears irreplaceable through data, machines and methods of processing and analysis.

---

[71] Regulation (EU) 2016/679, art. 22. In this regard, see: Busch C., De Franceschi A., *Granular Legal Norms: Big Data and the Personalization of Private Law*, cit., 14-15; Busch C., *The Future of Pre-contractual Information Duties: From Behavioral Insights to Big Data*, cit., 237-238; on the other hand, see: Wachter S., Mittelstadt B., Floridi L., Why a right to explanation of automated decision-making does not exist in the General Data Protection Regulation. *Intern. Data Priv. Law*, 2017, 76.

[72] See: Busch C., De Franceschi A., *Granular Legal Norms: Big Data and the Personalization of Private Law*, cit., 15-16; on the usefulness of Big Data for this purpose, see also: Hacker P., *Personalizing EU Private Law: From Disclosures to Nudges and Mandates*, cit., 659.

[73] See, among others: Custers B., *Methods of data research for law*, cit., 374-75; Pistone M.R., Horn M.B., *Disrupting Law School: How Disruptive Innovation will Revolutionize the Legal World*. San Francisco: Christensen Institute, 2016; Susskind R., *Tomorrow's Lawyers*, Oxford: Oxford University Press, 2013.

[74] In this regard, see: Surden H., Machine Learning and Law. *Wash. Law Rev.*, 2014, 87.

Moreover, in this regard it is worth remembering that the analysis of the data is strictly connected to the data entered and the questions posed. Therefore, to evaluate the result of the data analysis, it is necessary to verify the quality of the data entered, the ways in which it was collected, and whether the human or human-programmed analyst has asked the right questions: in this sense, notwithstanding the emerging legal relevance of the professional categories of computer scientist and data scientist, the persistent centrality of the role of the jurist is evident[75].

However, the problematic considerations referred to do not lead to the conclusion that law, and especially comparative law, cannot benefit from the methods of analysis of Legal Big Data. Conversely, these methods can facilitate the collection of information and provide previously difficult to find or unexpected knowledge, and thus increase the efficiency, accuracy and reliability of legal research, regulatory development, jurisprudential application and legal practice[76].

[75] See: Remus D., Levy F.S., *Can Robots Be Lawyers? Computers, Lawyers and the Practice of Law*, 2016. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2701092.

[76] In this sense, see again: B. Custers, *Methods of data research for law*, cit., 377.