COMPUTER LAW & SECURITY REVIEW 32 (2016) 238–255

Available online at www.sciencedirect.com

ScienceDirect

www.compseconline.com/publications/prodclaw.htm

Computer Law & Security Review

Personal data for decisional purposes in the age of analytics: From an individual to a collective dimension of data protection

Alessandro Mantelero [a,b,*]

[a] *Nanjing University of Information Science & Technology (NUIST), School of Public Administration, Jiangsu, China*

[b] *Polytechnic University of Turin, Department of Management and Production Engineering, and Nexa Center for Internet & Society, Turin, Italy*

Keywords:
Big data
Right to privacy
Data protection
Group privacy
Collective interests
Data protection authorities
Risk assessment

A B S T R A C T

In the big data era, new technologies and powerful analytics make it possible to collect and analyse large amounts of data in order to identify patterns in the behaviour of groups, communities and even entire countries.

Existing case law and regulations are inadequate to address the potential risks and issues related to this change of paradigm in social investigation. This is due to the fact that both the right to privacy and the more recent right to data protection are protected as individual rights. The social dimension of these rights has been taken into account by courts and policymakers in various countries. Nevertheless, the rights holder has always been the data subject and the rights related to informational privacy have mainly been exercised by individuals.

This atomistic approach shows its limits in the existing context of mass predictive analysis, where the larger scale of data processing and the deeper analysis of information make it necessary to consider another layer, which is different from individual rights. This new layer is represented by the collective dimension of data protection, which protects groups of persons from the potential harms of discriminatory and invasive forms of data processing.

On the basis of the distinction between individual, group and collective dimensions of privacy and data protection, the author outlines the main elements that characterise the collective dimension of these rights and the representation of the underlying interests.

* Department of Management and Production Engineering, Politecnico di Torino, C.so Duca degli Abruzzi, 24, 10120 Torino, Italy.
E-mail address: alessandro.mantelero@polito.it.

http://dx.doi.org/10.1016/j.clsr.2016.01.014

1. Introduction and scope of the analysis

Big data analytics make it possible to infer predictive information from large amounts of data in order to acquire further knowledge about individuals and groups, which may not necessarily be related to the initial purposes of data collection.[1] Moreover, analytics group people together by their qualitative attributes and habits (e.g. low-income people, "working-class mom", "metro parents"[2]) and predict the future behaviour of these clusters[3] of individuals.[4]

This approach is adopted, for instance, by some health insurance companies, which extract predictive information about the risks associated with segments of clients on the basis of their primetime television viewing, propensity to buy general merchandise, ethnicity, geography or use of mail order buying.[5]

In these cases, predictions based on correlations[6] do not only affect individuals, which may act differently from the rest of the group to which have been assigned,[7] but also affect the whole group and set it apart from the rest of society. An example in this sense is provided by the "neighbourhood's general credit score" adopted by credit companies,[8] which induces companies to provide opportunities for people living in a given neighbourhood in a way that bears no relationship to their individual conditions, but is based on the aggregate score of the area.[9]

These issues are not new and may be considered the effect of the evolution of profiling technologies, in a context characterised by an increased volume of information available and powerful software analytics.[10] Nevertheless, previous forms of categorisation and profiling were based on a few standard variables (e.g. sex, age, family income, marital status, place of residence); therefore, their predictive ability was limited. Today, big data analytics use hundreds of different variables to infer predictive information about groups of people and, in many cases, these variables concern aspects that are not clearly related to the final profiles created by analytics.

Moreover, users are often unaware of these forms of data analysis and of the impact that some information may have on their membership of one or another group created by analytics. Finally, decision makers use the outcomes generated by big data analytics to take decisions that affect individuals and groups, without allowing them any participation in the process, which remains primarily based on obscure data management and frequently takes place in situations of imbalance between data gatherers and data subjects.

In the light of the above, the use of big data analytics creates "a new truth regime",[11] in which general strategies are adopted

[1] See David Bollier, 'The Promise and Perils of Big Data' (Aspen Institute, Communications and Society Program 2010) <http://www.aspeninstitute.org/sites/default/files/content/docs/pubs/The_Promise_and_Peril_of_Big_Data.pdf> accessed 27 February 2014. See also Pertti Ahonen, 'Institutionalizing Big Data methods in social and political research' (2015) Big Data & Society 1–12 <http://bds.sagepub.com/content/2/2/2053951715591224> accessed 21 July 2015.

[2] This is one of the categories used by US data brokers to define specific segments of population based on models of predictive behaviour. In this sense, the category "metro parents" includes consumers "primarily in high school or vocationally educated [. . .] handling single parenthood and the stresses of urban life on a small budget", see Federal Trade Commission, 'Data Brokers: A Call for Transparency and Accountability' (2014), 20 and Appendix B <https://www.ftc.gov/system/files/documents/reports/data-brokers-call-transparency-accountability-report-federal-trade-commission-may-2014/140527databrokerreport.pdf> accessed 27 February 2014.

[3] In this article, the notion of cluster is used to identify a set of individuals that are directly or indirectly grouped on the basis of common qualitative elements (class of age, habits, geographic distribution, etc.).

[4] See Recital nn. 51, 58 and 58a of the Proposal for a Regulation of the European Parliament and of the Council on the protection of individuals with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation) text adopted by the Council of the European Union, Brussels, 19 December 2014 (hereinafter abbreviated as EU Proposal).

[5] See Satish Garla, Albert Hopping, Rick Monaco and Sarah Rittman, 'What Do Your Consumer Habits Say About Your Health? Using Third-Party Data to Predict Individual Health Risk and Costs. Proceedings' (SAS Global Forum 2013) <http://support.sas.com/resources/papers/proceedings13/170-2013.pdf> accessed 28 February 2015; see also Federal Trade Commission (n 2) 20 and Appendix B.

[6] See Bollier (n 1). See also Mireille Hildebrandt, 'Profiling: From Data to Knowledge. The challenges of a crucial technology' (2006) 30(9) Datenschutz und Datensicherheit 548.

[7] See also Barbara D. Underwood, 'Law and the Crystal Ball: Predicting Behavior with Statistical Inference and Individualized Judgment' (1979) 88 Yale Law Journal 1408.

[8] This score predicts credit risks of individuals that live in a small geographic area and it is defined on the basis of aggregate credit scores.

[9] See Pam Dixon and Robert Gellman, 'The Scoring of America: How Secret Consumer Scores Threaten Your Privacy and Your Future' (2014), 21, 44, <http://www.worldprivacyforum.org/wp-content/uploads/2014/04/WPF_Scoring_of_America_April2014_fs.pdf> accessed 10 March 2015. See also Frank Pasquale, *The Black Box Society. The Secret Algorithms That Control Money and Information* (Harvard University Press 2015) 22–26; Danielle Keats Citron and Frank Pasquale, 'The Scored Society: Due Process For Automated Predictions' (2014) 89 Wash. L. Rev. 1; Meike Kamp, Barbara Körffer and Martin Meints, 'Profiling of Customers and Consumers – Customer Loyalty Programmes and Scoring Practices' in Mireille Hildebrandt and Serge Gutwirth (eds.), *Profiling the European Citizen. Cross-Disciplinary Perspective* (Springer 2010) 205–211; Anton H. Vedder, 'Privatization, Information Technology and Privacy: Reconsidering the Social Responsibilities of Private Organizations' in Geoff Moore (ed), *Business Ethics: Principles and Practice* (Business Education Publishers 1997) 215–226.

[10] See also Serge Gutwirth and Mireille Hildebrandt, 'Some Caveats on Profiling' in Serge Gutwirth, Yves Poullet and Paul de Hert (eds.) *Data protection in a profiled world* (Dordrecht, London 2010) 32–33.

[11] See Antoinette Rouvroy, 'Des données sans personne: le fétichisme de la donnée à caractère personnel à l'épreuve de l'idéologie des Big Data' (2014) 9 <http://works.bepress.com/antoinette_rouvroy/55> accessed 8 March 2015; Antoinette Rouvroy, 'Algorithmic Governmentality and the End(s) of Critique' (2013) <https://vimeo.com/79880601> accessed 10 March 2015.

on a large scale on the basis of representations of society generated by algorithms, which predict future collective behaviour.[12] These strategies are then applied to specific individuals, given the fact that they are part of one or more groups generated by analytics.[13]

The use of analytics and the adoption of decisions based on group behaviour rather than on individuals are not limited to commercial and market contexts. They also affect other important fields, such as security and social policies, where a different balancing test should be applied, given the importance of public interest issues.

One example of this is provided by predictive policing solutions like "PredPol",[14] a software used by US local police forces to anticipate, prevent and respond more effectively to crime, on the basis of cross check data, places and techniques of recent crimes. PredPol and similar software are able to predict future crimes and their location, but they also induce a "self-fulfilling cycles of bias". This is due to the fact that police departments allocate more resources to the areas suggested by analytics and this increases crime detection at local level, with the result of reinforcing the original prediction. At the same time, a reduced police presence in other areas reduces crime detection and produces an adverse prediction for these areas.[15] The consequence of these software solutions is a potential geographical discrimination, which might not directly affect individuals, but has an impact on local communities in terms of social stigma or inadequate provision of police services. In this sense, there is a collective interest in a correct and accurate use of data.[16]

These scenarios show the collective dimension of decisions adopted using data analytics and their potential bias.[17] Against this background, Korzybski's statement "a map is not the territory"[18] sums up the focus of this article. The logic of the author of the map, the way in which the territory is

[12] See Pasquale (n 9); Viktor Mayer-Schönberger and Kenneth Cukier, *Big Data. A Revolution That Will Transform How We Live, Work and Think* (John Murray 2013); Bollier (n 1); McKinsey Global Institute *Big data: The next frontier for innovation, competition, and productivity* (2011) <http://www.mckinsey.com> accessed 16 April 2012. See also Bellagio Big Data Workshop Participants, 'Big data and positive social change in the developing world: A white paper for practitioners and researchers' (Oxford Internet Institute 2014) <http://www.rockefellerfoundation.org/uploads/files/c220f1f3-2e9a-4fc6-be6c-45d42849b897-big-data-and.pdf> accessed 28 June 2015; Ira S. Rubinstein, 'Big Data: The End of Privacy or a New Beginning?' (2013) 3 (2) International Data Privacy Law 74–87; danah boyd and Kate Crawford, 'Six Provocations for Big Data' (paper presented at Oxford Internet Institute's "A Decade in Internet Time: Symposium on the Dynamics of the Internet and Society", Oxford, September 21, 2011) <http://ssrn.com/abstract=1926431> accessed 16 April 2012; danah boyd and Kate Crawford, 'Critical Questions for Big Data: Provocations for a Cultural, Technological, and Scholarly Phenomenon' (2012) 15(5) Information, Communication, & Society 662–679; Omer Tene and Jules Polonetsky, 'Privacy in the Age of Big Data. A Time for Big Decisions' (2012) 64 *Stan. L. Rev. Online* 63–69 <http://www.stanfordlawreview.org/sites/default/files/online/topics/64-SLRO-63_1.pdf> accessed 14 March 2013.

[13] See Federal Trade Commission (n 2) IV–V ("Potentially sensitive categories include those that primarily focus on ethnicity and income levels, such as 'Urban Scramble' and 'Mobile Mixers,' both of which include a high concentration of Latinos and African Americans with low incomes. Other potentially sensitive categories highlight a consumer's age such as 'Rural Everlasting,' which includes single men and women over the age of 66 with 'low educational attainment and low net worths,' while 'Married Sophisticates' includes thirty-something couples in the 'upper-middle class . . . with no children' "). See also Bollier (n 1); Hildebrandt, 'Profiling: From Data to Knowledge. The challenges of a crucial technology' (n 6) 549–550.

[14] See Walter L. Perry, Brian McInnis, Carter C. Price, Susan C. Smith and John S. Hollywood, 'Predictive Policing. The Role of Crime Forecasting in Law Enforcement Operations' (The RAND Corporation 2013), <http://www.rand.org/content/dam/rand/pubs/research_reports/RR200/RR233/RAND_RR233.pdf> accessed 10 March 2015. See also Pasquale (n 9) 41–46, 48–51; Robinson + Yu, 'Civil Rights, Big Data, and Our Algorithmic Future. A September 2014 report on social justice and technology' (2014), 18–19 <http://bigdata.fairness.io/wp-content/uploads/2014/09/Civil_Rights_Big_Data_and_Our_Algorithmic-Future_2014-09-12.pdf> accessed 10 March 2015; Alessandro Mantelero and Giuseppe Vaciago, 'Social media and big data' in Babak Akhgar, Andrew Staniforth and Francesca Bosco (eds.) *Cyber Crime & Cyber Terrorism. Investigator's Handbook* (Elsevier 2014), 175–196; Andrew Guthrie Ferguson, 'Predictive Policing: The Future of Reasonable Suspicion' (2012) 62 Emory L. J. 259 <http://www.law.emory.edu/fileadmin/journals/elj/62/62.2/Ferguson.pdf> accessed 29 January 2014; Rosamunde van Brakel and Paul De Hert, 'Policing, surveillance and law in a pre-crime society: Understanding the consequences of technology based strategies' (2011) 20(3) Journal of Police Studies 20(3): 163 – 192 http://www.vub.ac.be/LSTS/pub/Dehert/378.pdf accessed 27 July 2015.

[15] See Kelly K. Koss, 'Leveraging Predictive Policing Algorithms to Restore Fourth Amendment Protections in High-Crime Areas in a Post-Wardlow World' (2015) 90 Chi.-Kent. L. Rev. 301, 311–312.

[16] See also Perry et. al. (n 14), 118–125; Oscar H. Gandy Jr., 'Exploring Identity and Identification in Cyberspace' (2000) 14 Notre Dame J.L. Ethics & Pub. Pol'y 1085, 1100 <http://scholarship.law.nd.edu/ndjlepp/vol14/iss2/10> accessed 10 July 2015.

[17] See also Tal Z. Zarsky, 'Transparent Predictions' (2013) 4 U. Ill. L. Rev. 1503, 1510–1513.

[18] Alfred Korzybski, 'A Non-Aristotelian System and its Necessity for Rigour in Mathematics and Physics' in Alfred Korzybski (ed) *Science and sanity: An Introduction to Non-Aristotelian Systems and General Semantics* (Institute of General Semantics 1933) 747, 750 <http://lipn.univ-paris13.fr/~duchamp/Books&more/Neurosciences/Korzybski/%5BAlfred_Korzybski%5D_Science_and_Sanity_An_Introduc%28BookFi.org%29.pdf> accessed 15 March 2015; see also Kate Crawford, 'Algorithmic Illusions: Hidden Biases of Big Data', presentation at Strata 2013, <https://www.youtube.com/watch?v=irP5RCdpilc> accessed 15 March 2015.

represented and the potential errors of representation can produce different maps of the same territory. Maps are not neutral. In the same way, in social investigations, the strategies used to group data, the logic of big data analytics and their potential bias have an influence on the final representation of groups and society.[19]

This "categorical" approach characterising the use of analytics leads policymakers to adopt common solutions for individuals belonging to the same cluster generated by analytics. These decisional processes do not consider individuals *per se*, but as a part of a group of people characterised by some common qualitative factors.

This leads to a reflection on privacy and data protection.[20] The use of personal information and big data analytics to support decisions exceeds the boundaries of the individual dimension and assumes a collective dimension, with potential harmful consequences for some groups.[21] In this sense, prejudice can result not only from the well-known privacy-related risks (e.g. illegitimate use of personal information, data security), but also from discriminatory and invasive forms of data processing.[22]

The dichotomy between individuals and groups is not new and it has already been analysed with regard to the legal aspects of personal information. Nonetheless, the right to privacy and the (more recent) right to the protection of personal data have been largely safeguarded as individual rights, despite the social dimension of their rationale.[23]

The focus on the model of individual rights is probably the main reason for the few contributions by privacy scholars on the collective dimension of privacy and data protection. Hitherto, only few authors have investigated the notion of group privacy. They have represented this form of privacy as the privacy of the facts and ideas expressed by the members of a group in the group environment or in terms of protection of information about a group.

Against this background, this article is not an attempt to provide a new interpretation of group privacy or to investigate the relationships between the legal and sociological notions of the group, which is only briefly touched on in the following paragraphs. Rather, it focuses on the new kind of groups that results from the use of big data analytics to represent the "territory" of our society. In this light, the article investigates the consequences of this algorithmic representation, in terms of protection of collective rights.

From this perspective, the first part of this article deals with the traditional notions of individual privacy and group privacy; it points out the novelty of the groups generated by algorithms,[24] which are ontologically different from the groups referred to by the original notion of "group privacy".[25] In this sense, big data analytics generate new groups, which did not previously exist in society, variable aggregations of individuals whose personal information is mined in order to extract predictive inferences.

The different origin and morphology of these groups make it necessary to investigate the collective dimension of privacy and data protection, which is different from the manifestation of the individual right to be let alone in the group context or the protection of information regarding the group. For this reason, the second part of the article focuses on the main elements that characterise this collective dimension in the context of big data analytics[26] and examines the nature of the collective interests at issue in this regard, their representation and the balance with other conflicting interests.

2. The notion of group in privacy literature: Group privacy and individual rights

Privacy scholars have devoted few contributions to group privacy and collective interests in data processing. A first approach considers group privacy as the right to privacy concerning information shared within a group by its members.[27] In this sense, there is no autonomous notion of group privacy, but only a peculiar attitude of individual privacy in the context of groups. Individual privacy describes the conditions under which a "right

[19] It should be noted that different architectures of algorithms may produce different results, although on the basis of the same factors. See Dixon and Gellman (n 9), 2 ("new consumer scores use thousands of pieces of information about consumers' pasts to predict how they will behave in the future. Issues of secrecy, fairness of underlying factors, use of consumer information such as race and ethnicity in predictive scores, accuracy, and the uptake in both use and ubiquity of these scores are key areas of focus").

[20] The origin of data, the nature of information and its legal protection (i.e. right to privacy or data protection) are not particularly relevant in the context described in the previous paragraphs. In this article, the analysis of privacy and data protection is more focused on the use of information and on the relationship between individual and collective dimensions, rather than on the traditional aspects of secrecy and data quality. See also Fred H. Cate and Viktor Mayer-Schönberger, 'Data Use and Impact. Global Workshop' (The Center for Information Policy Research and The Center for Applied Cybersecurity Research, Indiana University 2013) iii <http://cacr.iu.edu/sites/cacr.iu.edu/files/Use_Workshop_Report.pdf> accessed 27 February 2014; Alessandro Mantelero, 'The future of consumer data protection in the E.U. Rethinking the "notice and consent" paradigm in the new era of predictive analytics' in this Review (2014), vol 30, issue 6, 643–660.

[21] See also Kate Crawford, Gustavo Faleiros, Amy Luers, Patrick Meier, Claudia Perlich and Jer Thorp, 'Big Data, Communities and Ethical Resilience: A Framework for Action' (2013) 6–7 <http://www.rockefellerfoundation.org/app/uploads/71b4c457-cdb7-47ec-81a9-a617c956e6af.pdf> accessed 5 April 2015; danah boyd, Karen Levy, Alice Marwick, 'The Networked Nature of Algorithmic Discrimination', in Seeta Peña Gangadharan, Virginia Eubanks and Solon Barocas Data and Discrimination: Collective Essays (Open Technology Institute and New America 2014) 56 <http://www.newamerica.org/downloads/OTI-Data-an-Discrimination-FINAL-small.pdf> accessed 14 April 2015.

[22] See also The White House, Executive Office of the President, 'Big Data: Seizing Opportunities, Preserving Values' (2014) <http://www.whitehouse.gov/sites/default/files/docs/big_data_privacy_report_may_1_2014.pdf> accessed 26 December 2014. See also Zarsky (n 17) 1560–1563; Vedder (n 9).

[23] See below fn. 61.

[24] Section 2.

[25] See Edward J. Bloustein, 'Group Privacy: The Right to Huddle' (1977) 8 Rutgers-Cam. L.J. 219 and Edward J. Bloustein, *Individual and Group Privacy* (Transaction Books 1978) 123–186; Alan F. Westin, *Privacy and Freedom* (Atheneum 1970) 42–51.

[26] Section 3.

[27] See Bloustein, *Individual and Group Privacy* (n 25).

to be let alone" should be recognised, while group privacy determines the type of personal information sharing that goes within a group.[28] Group privacy therefore refers to the privacy of the facts or ideas expressed by members of a group in the group environment and provides a guarantee that this information will not be revealed outside the group.[29]

This notion of group privacy focuses on secrecy and intimacy and, for this reason, is mainly based on the level of trust existing among the members of a group. The consequence is a duty of confidentiality.[30] Group privacy concerns the breach of this duty.[31] Nevertheless, this does not represent a change in the traditional perspective, which remains based on the individual's right to privacy.[32]

A slightly different notion of group privacy is represented by the idea of "organizational privacy",[33] which focuses on control over information concerning collective entities and on the self-determination of these entities.[34] In this sense, group privacy still relies on confidentiality and regards the interests of the group itself in the protection of facts, acts or decisions that concern its internal affairs and its organisational autonomy.[35] Thus questions regarding "organizational privacy" do not only concern whether legal persons might have a legitimate claim to privacy,[36] but also revolve around the indirect protection of individuals who constitute the collective entities and their group interests.[37]

These two theories on group privacy concern the peculiar nature of the sharing of personal information within a group. They provide a sort of context-related notion of individual privacy. A different approach focuses on information regarding groups *per se* and does not focus on groups as a sum total of the individuals that make them up, with the related internal dynamics.[38] In this perspective, a group is an autonomous entity (an organised or a non-organised collective entity) and

[28] See Bloustein, *Individual and Group Privacy* (n 25) 127–130.

[29] See Bloustein, *Individual and Group Privacy* (n 25) 129–134. In the description of the various contexts in which the right to privacy is relevant in the light of the group dimension, the author considers marital, priest–penitent, lawyer–client and physician–patient relationships. In all these cases, the right to privacy is mainly related to intimacy and secrecy.

[30] It should be noted that terms like confidentiality or "relational privacy" have been also used to describe the aspects concerning the notion of group privacy that has been investigated by Bloustein. See, e. g., Lawrence O. Gostin, *Public health law: power, duty, restraint* (University of California Press 2008) 316; Christine M. Emery, 'Relational privacy. A Right To Grieve in The Information Age: Halting The Digital Dissemination of Death-Scene Images' (2011) 42 Rutgers L. J. 765. Authors used the notion of relational privacy to point out the contextual nature of this right, its intimate nature and its protection, but without focusing on the group dimension *per se*. See James Rachels, 'Why Privacy is Important' (1975) 4(4) Philosophy & Public Affairs 323–333; Charles Fried, 'Privacy [A moral analysis]' (1968) 77(3) Yale L. J. 475–493. See also Kendall Thomas, 'Beyond the Privacy Principle' (1992) 92(6) Columbia Law Review 1431, 1445–1446. On the contrary, the study conducted by Bloustein focuses on the group environment and provides a more detailed analysis of the right to privacy in this context. Moreover, this author puts the notion of group privacy in relationship with the dynamics of groups and the sociological theories on groups. Finally, the mentioned notion of "relational privacy" is very vague, since it is used by legal scholars to describe different kinds of social-related aspects concerning privacy, from privacy of the relatives concerning the death of members of their family to intimate sexual aspects, up to the more recent dimension of social network interaction. See also Lorraine G. Kisselburgh, 'Reconceptualizing privacy in technological realms: Theoretical frameworks for communication' (2008) Annual meeting of the International Communication Association, TBA, Montreal, Quebec, Canada <http://citation.allacademic.com/meta/p233000_index.html> accessed 20 February 2015; Beate Rössler, *The value of privacy* (Polity 2005) 130–133.

[31] See Bloustein, *Individual and Group Privacy* (n 25) 137–140, 180–181.

[32] See Bloustein, *Individual and Group Privacy* (n 25) 125 ("Group privacy is an extension of individual privacy. The interest protected by group privacy is the desire and need of people to come together, to exchange information, share feelings, make plans and act in concert to attain their objectives").

[33] See Westin (n 25).

[34] Collective entities may be autonomous and independent of the sum of their members, but, at the same time, they are the sum of the individual persons who make them up. For this reason, although organisational privacy can be considered as an autonomous right of legal persons, in many cases it also represents an indirect protection of the individual rights of their members and of the secrecy of members' interaction in the context of the organisation. See Westin (n 25) 42. See also Lee A. Bygrave, *Data Protection Law. Approaching Its Rationale, Logic and Limits* (Kluwer Law International 2002) 175–176, 186.

[35] See also Westin (n 25). For an analysis of the theoretical approach adopted by Westin, see also Bygrave, *Data Protection Law. Approaching Its Rationale, Logic and Limits* (n 34) 247–252.

[36] This issue is part of the more general debate on personality rights of collective entities and on the dualism characterising legal persons, from von Savigny and van Gierke to Kelsen, Hohfeld and Hart. See Friedrich Karl von Savigny, *Jural relations : or, The Roman law of persons as subjects of jural relations : being a translation of the second book of Savigny's System of modern Roman law*, translated by W. H. Rattigan (Wildy & Sons 1884); Otto von Gierke, *Die Genossenschaftstheorie und die deutsche Rechtsprechung* (Weidmann 1887); Wesley N. Hohfeld and Walter W. Cook, *Fundamental legal conceptions as applied in judicial reasoning : and other legal essays* (Yale University Press 1923); Herbert L. A. Hart, Definition and theory in jurisprudence (Stevens 1954); Hans Kelsen, *Reine Rechtslehre. Mit einem Anhang: Das Problem der Gerechtigkeit* (F. Deuticke 1960). See also John Dewey, 'The Historic Background of Corporate Legal Personality' (1926) 35(6) Yale L. J. 655–673; Katsuhito Iwai, 'Persons, Things and Corporations: The Corporate Personality Controversy and Comparative Corporate Governance' (1999) 47(4) The American Journal of Comparative Law 583–632.

[37] On this dual dimension, which characterises collective entities and the legal protection of related interests, see Bygrave, *Data Protection Law. Approaching Its Rationale, Logic and Limits* (n 34) 175–176, 250–253.

[38] See Bygrave, *Data Protection Law. Approaching Its Rationale, Logic and Limits* (n 34) 173–298.

group privacy refers to information that identifies and describes the group.[39]

Under this third interpretation, group privacy protects information referring to collective entities – both legal persons and organisations or groups without a formal and independent identity – and acts as an extension of individual data protection to these entities.[40] Although this notion of group privacy is different from the other definitions briefly described above, it seems not to challenge the traditional perspective that characterises privacy and data protection. The group dimension affects the manifestation of these rights in a specific context (the group), but they still revolve around the model of individual rights, although referring to a collective entity or its members.[41] Nevertheless, this approach is the closest to the collective dimension of data protection, mainly when it focuses on non-organised collective entities.[42]

Despite these differences between the theories about group privacy, this brief overview shows that the existing studies are for the most part based on the individual rights model where they consider the group dimension of privacy and data protection. These two rights are related to given individuals who are members of a group, or to the group itself as an autonomous collective body. In both cases, the architecture of these rights is not inspired by the idea of the group's collective and non-aggregative interests.

This approach to the issues related to groups is consistent with the traditional protection of the rights in question. The right to privacy and the right to data protection have been treated as individual rights in both the U.S. and European experiences, though based on differing origins and evolutions.

In the U.S., at the end of the 19th century, Warren and Brandeis shaped the modern idea of privacy,[43] which was different from the previous notion of protection of private life based on property.[44] In spite of this, the right to privacy, although redefined as a personality right, remained largely based on the individual dimension.[45] Neither the notion of decisional privacy nor its constitutional dimension, originating in the ground-breaking opinion given by Brandeis in his role as Supreme Court judge,[46] abandoned the individualistic nature of the right.

On the other side of the Atlantic, individual privacy protection stemmed from the same social factors (the invasive attitude of the "penny press" and new media) that justified the response of the U.S. legal system to privacy invasion and the protection of the right to be let alone.[47] However, the European legal notion of privacy did not draw its origins from the U.S. experience, but was independently shaped by legal scholars and the courts.[48] From the theoretical point of view, the right to privacy was placed in the sphere of individual rights, as in the U.S., but in European case law and literature there is a closer connection with the general theory of personality rights.[49]

Moreover, unlike in the U.S., the European notion of privacy has not acquired the wider dimension of the U.S. decisional privacy, but has remained more focused on informational privacy. This does not mean the right to self-determination with regard to government and public bodies has not been recognised in Europe, but that it rests on the different fundamental

[39] See Bygrave, *Data Protection Law. Approaching Its Rationale, Logic and Limits* (n 34) part III. The author extensively describes the various issues concerning data protection rights of collective entities. See also Jean-Pierre Chamoux, 'Data Protection in Europe: The Problem of the Physical Person and their Legal Person' (1981) 2 J. Media Law & Practice 70–83. See also Article 3(2)(b) of Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data (Council of Europe, Strasbourg, 28.01.1981) and Recital 24 of the Directive 95/46/EC.

[40] See Bygrave, *Data Protection Law. Approaching Its Rationale, Logic and Limits* (n 34) 186, 241–282, 288–289. See also Bart van der Sloot, 'Do privacy and data protection rules apply to legal persons and should they? A proposal for a two-tiered system' in this Review (2015), vol. 31, issue 1, 26, 33–45.

[41] But see, more recently, Lee A. Bygrave and Dag Wiese Schartum, 'Consent, Proportionality and Collective Power' in Serge Gutwirth, Yves Poullet, Paul De Hert, Cécile de Terwangne and Sjaak Nouwt (eds.), *Reinventing Data Protection?* (Springer 2009), 157–173.

[42] See Bygrave, *Data Protection Law. Approaching Its Rationale, Logic and Limits* (n 34) 283–295. Bygrave points out the peculiar nature of non-organised collective entities, which are created by persons or organisations outside the group. Moreover, the author suggests some remedies to protect the interests of these entities and their members according to the existing data protection framework.

[43] Samuel D. Warren and Luis D. Brandeis, 'The Right to Privacy' (1890) 4(5) Harv. L. Rev. 193–220.

[44] See Warren and Brandeis (n 43) 205, 213; Westin (n 25) 337–345; David W. Leebron, 'The Right to Privacy's Place in the Intellectual History of Tort Law' (1991) 41 Case W. Res. L. Rev. 769, 775–778, 781; Robert C. Post, 'Rereading Warren and Brandeis: Privacy, Property and Appropriation' (1991) 41 Case W. Res. L. Rev. 647, 663–670; Amitai Etzioni, *The Limits of Privacy* (Basic Books 1999) 189.

[45] See Warren and Brandeis (n 43) 219; Westin (n 25) 330–364; Etzioni (n 44) 189, 196; Daniel J. Solove, *Understanding Privacy* (Harvard University Press 2008) 12–37, 78–98.

[46] See Brandeis' opinions in *Olmstead v. United States* 277 US 438, 471 (1928). See also *Sweezy v. New Hampshire* 354 US 234 (1957); *NAACP v. Alabama* 357 US 449 (1958); *Massiah v. US* 377 US 201 (1964); *Griswold v. Connecticut*, 381 US 479 (1965); *Roe v. Wade* 410 US 113 (1973).

[47] See Michael Schudson, *Discovering the News. A Social History of American Newspaper* (Basic Books 1978) 12–60.

[48] See, e.g., Trib. civ. Seine, 16 June 1858, D.P., 1858.3.62. See also Lee A. Bygrave, 'Privacy Protection in a Global Context. A Comparative Overview' (2004) 7 Scandinavian Studies in Law 319, 326–331; James Q. Whitman, 'The Two Western Cultures of Privacy: Dignity versus Liberty' (2004) 113 Yale L.J. 1151–1221; Frits W. Hondius, *Emerging Data Protection in Europe* (North Holland Publishing Company 1975) 6.

[49] See Stig Stromhölm, *Right of Privacy and Rights of Personality. A comparative Survey* (Norstedt & Soners 1967) 28–31. See also Hans Giesker, *Das Recht der Privaten an der eigenen Geheimsphäre. Ein Beitrag zu der Lehre von den Individualrechten* (Müller 1905); Josef Kohler, *Urheberrecht an Schriftwerken und Verlagsrecht* (F. Enke 1907) 441.

freedoms recognised by European charters and conventions, not solely on the right to privacy.[50]

Despite these differences, the nature of the right to privacy depends primarily on the individual rights model on both sides of the Atlantic.[51] The collective dimension of this right has been recognised in the U.S. and Europe, but protected mainly indirectly, as an aggregation of individual privacy issues and not as an autonomous dimension.[52]

The same considerations can be applied to the legal regimes of personal information, which is regulated under data protection statutes. Although data protection laws have drawn their origins from citizens' concerns about government social control,[53] regarding the collective dimension of data protection, the regulation focuses on data subjects and their rights.[54] Collective interests have been actively protected as the sum total of various individual needs. Hence, lawmakers, the courts and data protection authorities have addressed these interests with remedies that are mainly focused on individual rights and their enforcement.

In light of the above, the approach based on the individual rights model, adopted by legal scholars with regard to group privacy, is in line with the general legal notions of privacy and data protection. It is also consistent with the theoretical studies on group theory conducted in the field of sociology.

The various approaches of legal scholars seem to reflect the more general controversy between individualistic and organic sociological theories about the nature of groups.[55] On the one hand, attention to the individual dimension of privacy and the interactions between different individuals[56] is consistent with the notion of group as the sum of the relationships existing among its members (individualistic theory). On the other hand, when the analysis takes into consideration information concerning the group itself as a whole,[57] the group is seen as an autonomous unit that assumes the form of an organised collective entity (organic theory).

In this context, the legal approach that considers group privacy as relating to the individual's privacy issues within a group[58] is in line with the individualistic theory, which sees groups as entities in which individuals interact with each other in a continuous and relatively stable manner. Moreover, from a sociological perspective, the members of a group are aware of being part of the group and the group is usually recognised as an autonomous social structure. According to this position, a group is the product of concurrent decisions of various persons who are striving to reach a common goal or share common experiences, values or interests.

[50] See the influential decision adopted by the Federal German Constitutional Court (Bundesverfassungsgericht), 15 December 1983, Neue Juristische Wochenschrift, 1984. <https://www.zensus2011.de/SharedDocs/Downloads/DE/Gesetze/Volkszaehlungsurteil_1983.pdf?__blob=publicationFile&v=9> accessed 25 June 2014. For an English translation, see Human Rights Law Journal 1984, 5: 94. See also the article 8 of the European Convention on Human Rights and the related case law of the European Court of Human Rights, Council of Europe, 'Case Law of the European Court of Human Rights Concerning the Protection of Personal Data' (2013) <http://www.coe.int/t/dghl/standardsetting/dataprotection/TPD_documents/DP%202013%20Case%20Law_Eng%20%28final%29.pdf> accessed 20 March 2015; European Court of Human Rights, 'Protection of personal data' (2014) <http://www.echr.coe.int/Documents/FS_Data_ENG.pdf> accessed 21 March 2015. See also Stefano Rodotà, 'Data Protection as a Fundamental Right' in Gutwirth, Poullet, De Hert, de Terwangne and Nouwt (n 41), 77–82; Joseph A. Cannataci, 'Lex Personalitatis & Technology-driven Law' (2008) 5(1) SCRIPTed 1–6 <DOI: 10.2966/scrip.050108.1>. On the notion of personality right, see Giorgio Resta, 'Personnalité, persönlichkeit, personality' (2014) 1(3) Comparative Perspectives on the Protection of Identity in Private Law 215–243.

[51] See Bygrave, 'Privacy Protection in a Global Context. A Comparative Overview' (n 48) 324–325; Solove, *Understanding Privacy* (n 45) ch. 2. See also Colin J. Bennett and Charles D. Raab, *The Governance of Privacy. Policy instruments in global perspective* (Ashgate 2003) ch. 1; Robert C. Post, 'The Social Foundations of Privacy: Community and Self in the Common Law Tort' (1989) 77 Cal. L. Rev. 957; Julie E. Cohen, 'Examined Lives: Informational Privacy and the Subject' (2000) 52 Stan. L. Rev. 1373, 1426–1428.

[52] See *inter alia* Article 29 Data Protection Working Party, 'Letter to Mr. Larry Page, Chief Executive Officer' (2013) <http://ec.europa.eu/justice/data-protection/article-29/documentation/other-document/files/2013/20130618_letter_to_google_glass_en.pdf> accessed 27 February 2014; Irish Data Protection Commissioner, 'Facebook Ireland Ltd. Report of Re-Audit' (2012) <http://dataprotection.ie/documents/press/Facebook_Ireland_Audit_Review_Report_21_Sept_2012.pdf> accessed 27 February 2014; Italian Data Protection Authority, 'Injunction and Order Issued Against Google Inc.' (2013) <http://www.garanteprivacy.it/web/guest/home/docweb/-/docweb-display/docweb/3133945> accessed 27 February 2014. Only in a few hypotheses, collective privacy is recognised as an autonomous dimension, which is different from individual privacy. This happens in labour law, where the representatives of employees concur on the adoption of the decisions concerning surveillance in the workplace on behalf of the workers. See, European Commission, 'Second stage consultation of social partners on the protection of workers' personal data' (undated) 7, 10, 16–17 <http://ec.europa.eu/social/main.jsp?catId=708> accessed 10 January 2015. See also specific references to the provisions of European national labour laws in Mark Freedland, 'Data Protection and Employment in the European Union. An Analytical Study of the Law and Practice of Data Protection and the Employment Relationship in the EU and Its Member' (1999) 40–43 <http://ec.europa.eu/social/main.jsp?catId=708> accessed 25 January 2015; Frank Hendrickx, 'Protection of Workers' Personal Data in the European Union' (undated) 33–35, 98–101, <http://ec.europa.eu/social/main.jsp?catId=708> accessed 18 January 2015. See also Article 4 of the Italian labour statute (L. 300/1970). See below para 3.2.

[53] See Westin (n 25) 158–168, 298–326; Adam C. Breckenridge, *The Right to Privacy* (University of Nebraska Press 1970) 1–3; Secretary's Advisory Committee on Automated Personal Data Systems, 'Records, Computers and the Rights of Citizens' (1973) <http://epic.org/privacy/hew1973report/> accessed 27 February 2014. See also Solove, *Understanding Privacy* (n 45) 4–5. See also Myron Brenton, *The Privacy Invaders* (Coward-McCann 1964); Vance Packard, *The Naked Society* (David McKay 1964); Arthur R. Miller, *The Assault on Privacy Computers, Data Banks, Dossiers* (University of Michigan Press 1971) 54–67; Viktor Mayer-Schönberger, 'Generational development of data protection in Europe?' in Philip E. Agre and Marc Rotenberg (eds), *Technology and privacy: The new landscape* (MIT Press 1997) 221–227.

[54] See Colin J. Bennett, *Regulating Privacy: Data Protection and Public Policy in Europe and the United States* (Cornell University Press 1992) 29–33, 47; Mayer-Schönberger, 'Generational development of data protection in Europe?' (n 53) 219, 221–222. See also Bygrave and Wiese Schartum (n 41) 169.

[55] See Bloustein, *Individual and Group Privacy* (n 25) 124.

[56] Ibid.

[57] See Westin (n 25) 42; Bygrave, *Data Protection Law. Approaching Its Rationale, Logic and Limits* (n 34) 173–282.

[58] See Bloustein, *Individual and Group Privacy* (n 25) 124.

On the other hand, the organic theory of groups is consistent with the notion of group privacy in terms of "organizational privacy",[59] which focuses on the informational self-determination of the group itself. In this sense, group privacy is more closely connected to the secrecy of the group's activities than to the secrecy of personal information shared within the group by its members. The organic theory is also consistent with the more recent approach focused on data protection,[60] which is not necessarily related to the secrecy of information, but also regards publicly available data on collective entities.

Finally, as we have seen, the central role of the individual rights model in protecting privacy and personal information does not mean that legal systems disregard the social dimensions of these rights. Both privacy and data protection play an important role in safeguarding not only individual interests, but also the quality of society in general. Freedom of association, limits to disproportionate surveillance practices, and prevention of discrimination based on sensitive personal data are just few examples of the social effects of safeguarding the right to privacy and personal information. Values such as democracy and pluralism are strictly related to the protection of these rights.[61]

However, the courts may address issues related to general interests only when they receive complaints from rights holders, but the right holders may have no interest in these issues, be unaware of the general interest,[62] or be in no position to react to potential threats to their interests, owing to situations of power imbalance.[63]

Independent authorities may better address these issues of general interest,[64] but we should remember that these authorities often act on a discretionary basis and this may have negative effects, in terms of under-deterrence. The licensing model, which has been adopted in some cases by national regulators, represents a possible solution in assessing the risks associated with specific technologies or business models, and to prevent under-deterrence.[65]

For these reasons, independent authorities may play an important role in safeguarding interests related to the collective dimension of privacy and data protection in the big data environment. Even so, adequate solutions are required to enlarge their role and move from a discretionary approach to a general and mandatory assessment of the impact of technologies and business models on data protection.[66]

3. A new dimension of protection

In the big data era, new technologies and powerful analytics make it possible to collect and analyse huge amounts of data to try and identify patterns in the behaviour of groups of individuals[67] and to take decisions that affect the internal dynamics of groups, with consequences for the collective issues of the people involved.

Nevertheless, these groups are different from those considered in the literature on group privacy, in fact that they are created by data gatherers selecting specific clusters of information. Data gatherers shape the population they set out to investigate. They collect information about various people, who do not know the other members of the group and, in many cases, are not aware of the consequences of their belonging to a group.[68] This is the case of consumer group profiling,[69] scoring solutions[70] and predictive policing applications,[71] mentioned above.

The issues relating to privacy that arise from this new situation are different from the issues of individual privacy and group privacy. We are neither in the presence of forms of analysis that involve only individuals, nor in the presence of groups in the traditional sociological meaning of the term, given group members' lack of awareness of themselves as part of a group and the lack of interactions among people grouped into various clusters by data gatherers.

[59] See above fn. 33. See also Bygrave, *Data Protection Law. Approaching Its Rationale, Logic and Limits* (n 34) 250.

[60] See above fn. 38. On the debate regarding the application of privacy notion to collective entities, see Bygrave, *Data Protection Law. Approaching Its Rationale, Logic and Limits* (n 34) 241–256.

[61] See Simitis, 'Auf dem Weg zu einem neuen Datenschutzrecht' (1984) 3 Informatica e diritto 111; Schwartz, 'Privacy and Participation: Personal Information and Public Sector Regulation in the United States (1995) 80 Iowa L. Rev. 553, 560–561. For a general overview, see Bygrave, *Data Protection Law. Approaching Its Rationale, Logic and Limits* (n 32) 133–143, 150–157; Anita L. Allen, *Uneasy Access: privacy for women in a free society* (Rowman & Littlefield 1988) ch. 2.

[62] Extensive video surveillance programs, which have been adopted by municipalities or police departments, offer an example in this sense: citizens are aware of being monitored, but, in many cases, do not care about surveillance and are not interested in the social impact of these control solutions.

[63] See para 3.2.

[64] See fn. 126.

[65] See below para 3.3.

[66] See below para 3.3.

[67] Moreover, this is also possible without directly identifying data subjects; see Andrej Zwitter, 'Big Data ethics' (2014) Big Data & Society 1, 4–5. See also Paul Ohm, 'Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization' (2010) 57 UCLA L. Rev. 1701–1777; Philippe Golle, 'Revisiting the uniqueness of simple demographics in the US population' in Ari Juels (ed), *Proc. 5th ACM workshop on Privacy in electronic society* (ACM 2006) 77–80; Latanya Sweeney, 'Simple Demographics Often Identify People Uniquely' (Carnegie Mellon University 2000) <http://dataprivacylab.org/projects/identifiability/paper1.pdf> accessed 24 January 2015; Latanya Sweeney, 'Foundations of Privacy Protection from a Computer Science Perspective' in *Proc. Joint Statistical Meeting, AAAS, Indianapolis* (2000) http://dataprivacylab.org/projects/disclosurecontrol/paper1.pdf, accessed 24 January 2015.

[68] See Mireille Hildebrandt, 'Defining Profiling: A New Type of Knowledge?' in Hildebrandt and Gutwirth (n 9) 19–20. See also Executive Office of the President of the United States-Council of Economic Advisers, 'Big Data Differential Pricing' (2015) 18 <https://www.whitehouse.gov/sites/default/files/docs/Big_Data_Report_Nonembargo_v2.pdf> accessed 25 March 2015; Gandy (n 16) 1085, 1088, 1095; Hildebrandt, 'Profiling: From Data to Knowledge. The challenges of a crucial technology' (n 6) 549–550.

[69] See also Ryan Calo, 'Digital Market Manipulation' (2014) 82 George Washington Law Review 995.

[70] See Federal Trade Commission (n 2). But see Articles 18 and 20 of the Directive 2014/17/EU on credit agreements for consumers relating to residential immovable property and amending Directives 2008/48/EC and 2013/36/EU and Regulation (EU) No 1093/2010.

[71] See above fn. 14.

We must therefore extend the field of investigation to the collective interests of the persons whose personal data are being collected, analysed and grouped. The differing nature of these groups of individuals requires a different approach that cannot be exclusively based on individual rights.

The new scale entails the recognition of a new layer, represented by the rights of groups of individuals to the protection of their collective privacy and data protection. Moreover, since the predictive nature of big data analytics is designed to assist decisions that affect a plurality of individuals in various fields, we must also consider the social and ethical effects associated with this type of analysis.[72]

This kind of approach differs from the theoretical framework proposed by legal scholars in shaping the notion of group privacy,[73] but it can give a specific answer to the issues arising from the present and future scenarios of the data-driven society.

3.1. *Collective data protection and its rationale*

The collective dimension of data protection has its roots in the individual's right to privacy and shares some similarities with group privacy, but differs from both these previous notions. On the one hand, notions of individual privacy and data protection do influence the definition of the boundaries of this collective dimension, but the greater scale affects the morphology of the interests involved and their enforcement. At the same time, group privacy – as hitherto described by legal scholars – represents the notion that is closest to the idea of collective data protection.

On the other hand, collective data protection does not necessarily concern facts or information referring to a specific person,[74] as with individual privacy and data protection. Nor does it concern clusters of individuals that can be considered groups in the sociological sense of the term. In addition, collective rights are not necessarily a large-scale representation of individual rights and related issues.[75] Finally, collective data protection concerns non-aggregative collective interests,[76] which are not the mere sum of many individual interests.[77]

The importance of this collective dimension depends on the fact that the approach to classification by modern algorithms does not merely focus on individuals, but on groups or clusters of people with common characteristics (e.g. customer habits, lifestyle, online and offline behaviour, etc.).[78] Data gatherers are mainly interested in studying groups' behaviour and predicting this behaviour, rather than in profiling single users. Data-driven decisions concern clusters of individuals and only indirectly affect the members of these clusters. One example of this is price discrimination based on age, habits or wealth.

The most important concern in this context is the protection of groups from potential harm due to invasive and discriminatory data processing. The collective dimension of data processing is mainly focused on the use of information,[79] rather than on secrecy[80] and data quality.

We need to adopt a broader notion of discrimination here, one that encompasses two different meanings. In a negative sense, discrimination is "the unjust or prejudicial treatment of different categories of people". In a more neutral and potentially positive sense, though, discrimination may be the "recognition and understanding of the difference between one thing and another".[81] Both these dimensions assume relevance in the context of big data analytics.

We will focus below on the first meaning, since the unfair practices characterised by discriminatory purposes are generally

[72] See Article 29 Data Protection Working Party, 'Statement on the role of a risk-based approach in data protection legal frameworks' (2014) 4 <http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2014/wp218_en.pdf> accessed 27 February 2014; Paul M. Schwartz, 'Data Protection Law and the Ethical Use of Analytics' 22–26 <http://www.huntonfiles.com/files/webupload/CIPL_Ethical_Undperinnings_of_Analytics_Paper.pdf> accessed 27 February 2014; David Wright, 'A framework for the ethical impact assessment of information technology' (2011) 13 Ethics Inf. Technol. 199–226. See also Luciano Floridi, *The 4TH Revolution. How the Infosphere is Reshaping Human Reality* (Oxford University Press 2014) 189–190; Helen Nissenbaum, *Privacy in Context. Technology, Policy, and the Integrity of Social Life* (Stanford University Press 2010) 231; Rayan M. Calo, 'Consumer Subject Review Boards: A Thought Experiment' (2013) 66 Stan. L. Rev. Online 97, 101–102; Cynthia Dwork and Deirdre K. Mulligan, 'It's not Privacy and It's not Fair' (2013) 66 Stan. L. Rev. Online 35, 38; Bygrave, *Data Protection Law. Approaching Its Rationale, Logic and Limits* (n 34) 61–62, 339; Julie E. Cohen, 'What Privacy is For' (2013) 126 Harv. L. Rev. 1904, 1925–1926; Crawford et al. (n 21) 4.

[73] See above para 2.

[74] In many cases, private companies and governments have no interests in profiling single customers or citizens, but wish to discover the attitudes of clusters of individuals. Their main purpose is to predict future behaviours of segments of the population to achieve economic or political goals. See Bollier (n 1).

[75] See Francesco Capotorti, 'Are Minorities Entitled to Collective International Rights?' in Yoram Dinstein and Mala Tabory (eds), *The Protection of Minorities and Human Rights* (Martinus Nijhoff Publishers 1992) 507, 511.

[76] See Dwight G. Newman, 'Collective Interests and Collective Rights' (2004) 49(1) American Journal of Jurisprudence 127, 131. See also below in the present section. On the contrary, an aggregative approach seems to be consistent with the notion of group privacy described by Bloustein, *Individual and Group Privacy* (n 25) 123–186.

[77] Contra Vedder (n 9), who claims that the notion of collective privacy "reminds of collective rights", but subjects of collective rights are groups or communities. Conversely, the groups generated by group profiling are not communities of individuals sharing similar characteristics and structured or organised in some way. For this reason, Vedder uses the different definition of "categorial privacy", see below fn. 101.

[78] See above para 1 and below in the text.

[79] See Cate and Mayer-Schönberger (n 20) iii; Mantelero, 'The future of consumer data protection in the E.U. Rethinking the "notice and consent" paradigm in the new era of predictive analytics' (n 20).

[80] See Bloustein. *Individual and Group Privacy* (n 25) 182.

[81] See <http://www.oxforddictionaries.com/it/definizione/inglese/discrimination> accessed 29 January 2015.

forbidden and sanctioned by law.[82] This article concerns involuntary forms of discrimination in cases where big data analytics provide biased representations of society.[83]

For example, in 2013 a study examined the advertising provided by Google AdSense and found statistically significant racial discrimination in advertisement delivery.[84] Similarly, Kate Crawford has pointed out certain "algorithmic illusions"[85] and described the case of the City of Boston and its StreetBump smartphone app to passively detect potholes. The application had a signal problem, due to the bias generated by the low penetration of smartphones among lower income and older residents. While the Boston administration took this bias into account and solved the problem, less enlightened public officials might underestimate such considerations and make potentially discriminatory decisions.[86]

Another example is the Progressive case, in which an insurance company obliged drivers to install a small monitoring device in their cars in order to receive the company's best rates. The system considered as a negative factor driving late at night, but did not take into account the potential bias against low-income individuals, who are more likely to work night shifts, compared with late-night party-goers, "forcing them [low-income individuals] to carry more of the cost of intoxicated and other irresponsible driving that happens disproportionately at night".[87]

These cases represent situations in which a biased representation of groups and society results from flawed data processing[88] or a lack of accuracy in the representation.[89] This produces potentially discriminatory effects as a consequence of the decisions taken on the basis of analytics.

On the other hand, the other sense of discrimination involving different treatment of different situations may represent an intentional goal for policy makers, which is in line with the rule of law. This is the case of law and enforcement bodies and intelligence agencies, which adopt solutions to discriminate between different individuals and identify targeted persons. Here there is a deliberate intention to treat given individuals differently, but this is not unfair or illegal providing it is within existing legal provisions. Nonetheless, as in the previous case, potential flaws or a lack of accuracy may cause harm to citizens.

For instance, criticisms have been raised with regard to the aforementioned predictive software adopted in recent years by various police departments in the US.[90] Leaving aside the constitutional profiles associated with these applications and the peculiar balance of interests of this use of data, there have been cases where people were named as potential offenders due to merely remote connections with authors of serious crimes.[91] Criticisms also concern the use of risk assessment procedures based on analytics coupled with a categorical approach (based on typology of crimes and offenders) in U.S. criminal sentencing.[92]

[82] See *inter alia* European Commission, 'Developing Anti-Discrimination Law in Europe. The 28 EU Member States, the Former Yugoslav Republic of Macedonia, Iceland, Liechtenstein, Norway and Turkey compared' (2013) <http://www.non-discrimination.net/content/media/Developing%20Anti-Discrimination%20Law%20in%20Europe%20EN%2029042014%20WEB.pdf> accessed 28 March 2015; Evelyn Ellis and Philippa Watson, *EU Anti-Discrimination Law* (2nd edn Oxford University Press 2015). See also Article 14 of the Convention for the Protection of Human Rights and Fundamental Freedoms; Article 21 of the Charter of Fundamental Rights of the European Union; Article 19 of the Treaty on the Functioning of the European Union; Directive 2000/43/EC; Directive 2000/78/EC. See also Wim Schreurs, Mireille Hildebrandt, Els Kindt and Michaël Vanfleteren, 'Cogitas, Ergo Sum. The Role of Data Protection Law and Non-discrimination Law in Group Profiling in the Private Sector' in Hildebrandt and Gutwirth (n 9) 258–264.

[83] See Citron and Pasquale (n 9), 14; Birny Burnbaum (2013). Insurers' Use of Credit Scoring for Homeowners in Ohio: A Report to the Ohio Civil Rights Commission.

[84] See Latanya Sweeney, 'Discrimination in Online Ad Delivery' (2013) 56(5) Communications of the ACM 44–54. See also Bianca Bosker, 'Google's Online Ad Results Guilty Of Racial Profiling, According To New Study' *The Huffington Post* (2 May 2013) <http://www.huffingtonpost.com/2013/02/05/online-racial-profiling_n_2622556.html> accessed 27 March 2015.

[85] Crawford (n 18).

[86] See Kate Crawford, 'The Hidden Biases in Big Data' (2013) Harv. Bus. Rev. April 1, 2013, <https://hbr.org/2013/04/the-hidden-biases-in-big-data> accessed 29 January 2015. Similar considerations can be made in the case of the predictive policing systems mentioned above in the text; see also fn. 14. See also Jonas Lerman, 'Big Data and Its Exclusions' (2013) 66 Stan. L. Rev. Online 55.

[87] See Robinson + Yu (n 14) 6.

[88] This is the case of the errors that affect the E-Verify system, which is used in the US to verify if a new worker is legally eligible to work in the US. See Robinson + Yu (n 14) 12–14; National Immigration Law Center, 'Verification Nation' (2013) 6 <www.nilc.org/document.html?id=957> accessed 29 January 2015.

[89] See also Gandy (n 16) 1099–1100.

[90] See above para 1.

[91] See Jeremy Gorner, 'Chicago police use 'heat list' as strategy to prevent violence. Officials generate analysis to predict who will likely be involved in crime, as perpetrator or victim, and go door to door to issue warnings' *Chicago Tribune* (Chicago, 21 August 2013) <http://articles.chicagotribune.com/2013-08-21/news/ct-met-heat-list-20130821_1_chicago-police-commander-andrew-papachristos-heat-list> accessed 25 February 2015.

[92] See U.S. Department of Justice – Criminal Division, Office of the Assistant Attorney General, 'Annual letter' (2014) 6–7, 13 <http://www.justice.gov/criminal/foia/docs/2014annual-letter-final-072814.pdf> accessed 29 January 2015 ("This phenomenon ultimately raises constitutional questions because of the use of groupbased characteristics and suspect classifications in the analytics. Criminal accountability should be primarily about prior bad acts proven by the government before a court of law and not some future bad behavior predicted to occur by a risk assessment instrument. Second, experience and analysis of current risk assessment tools demonstrate that utilizing such tools for determining prison sentences to be served will have a disparate and adverse impact on offenders from poor communities already struggling with many social ills"). See also Administrative Office of the United States Courts – Office of Probation and Pretrial Services, 'An Overview of the Federal Post Conviction Risk Assessment' (2011) <http://www.uscourts.gov/uscourts/FederalCourts/PPS/PCRA_Sep_2011.pdf> accessed 29 January 2015; Underwood (n 7) 1409–1413.

Discrimination – the different treatment of different situations – also appears in commercial contexts to offer tailored services to consumers. In this case, where the interests are of a purely private nature, commercial practices may lead to price discrimination[93] or the adoption of differential terms and conditions depending on the assignment of consumers to a specific cluster.[94]

Thus consumers classified as "financially challenged" belong to a cluster "[i]n the prime working years of their lives [. . .] including many single parents, struggl[ing] with some of the lowest incomes and little accumulation of wealth". This implies the following predictive viewpoint, based on big data analytics and regarding all consumers in the cluster: "[n]ot particularly loyal to any one financial institution, [and] they feel uncomfortable borrowing money and believe they are better off having what they want today as they never know what tomorrow will bring".[95] It is not hard to imagine the potential discriminatory consequences of similar classifications with regard to individuals and groups.

It should be noted that these forms of discrimination are not necessarily against the law, especially when they are not based on individual profiles and only indirectly affect individuals as part of a category, without their direct identification.[96] For this reason, existing legal provisions against individual discrimination might not be effective in preventing the negative outcomes of these practices, if adopted on a collective basis. Still, such cases clearly show the importance of the collective dimension of the use of information about groups of individuals.

Within the EU, such data analysis focussing on clustered individuals may not represent a form of personal data processing,[97] since the categorical analytics methodology does not necessarily make it possible to identify a person.[98] Moreover,

[93] Price discrimination or "differential pricing" is the practice of charging customers different prices for the same product; see Executive Office of the President of the United States-Council of Economic Advisers (n 68), 4–5. The cases considered in this article mainly concern the so-called third-degree price differentiation, which occurs when sellers charge different prices to different segments of the market. See also Alex Rosenblat, Rob Randhava, danah boyd, Seeta Peña Gangadharan, and Corrine Yu , 'Data & Civil Rights: Consumer Finance Primer' (2014) <http://www.datacivilrights.org/pubs/2014-1030/Finance.pdf> accessed 15 March 2015.

[94] See Executive Office of the President of the United States-Council of Economic Advisers (n 68); Federal Trade Commission (n 2) 3, 19–21; Dixon and Gellman (n 9). See also Timothy C. Lambert, 'Fair Marketing: Challenging Pre-Application Lending Practices' (1999) 87 Geo. L. J. 2182.

[95] See Federal Trade Commission (n 2) 20, fn. 52.

[96] Article 4 (12b) of the EU Proposal defines "profile" as "set of data characterising a category of individuals that is intended to be applied to a natural person". See also Articles 4 (12a), 14 (1a) (h) and 20, EU Proposal; Article 29 Data Protection Working Party, 'Advice paper on essential elements of a definition and a provision on profiling within the EU General Data Protection Regulation' <http://ec.europa.eu/justice/data-protection/article-29/documentation/other-document/files/2013/20130513_advice-paper-on-profiling_en.pdf> accessed 29 March 2015; Article 29 Data Protection Working Party, 'Opinion 01/2012 on the data protection reform proposals' (2012) <http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2012/wp191_en.pdf> accessed 29 March 2015. Regarding the decisions that affect an individual as member of a specific cluster of people, it should be noted that in many cases these decisions are not based solely on automated processing; see Zarsky (n 17) 1518–1519. In this sense, credit scoring systems have reduced but not removed human intervention on credit evaluation. At the same time, classifications often regard identified or identifiable individuals. See Article 29 Data Protection Working Party, 'Opinion 2/2010 on online behavioural advertising' (2010) <http://ec.europa.eu/justice/policies/privacy/docs/wpdocs/2010/wp171_en.pdf> accessed 29 March 2015; Data Protection Working Party, 'Working Document 02/2013 providing guidance on obtaining consent for cookies' (2013) 5–6 <http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2013/wp208_en.pdf> accessed 29 March 2015. Regarding the applicability of the Data Protection Directive in case of automated profiling, see Lee A. Bygrave, 'Minding the Machine: Article 15 of the EC Data Protection Directive and Automated Profiling' in this Review (2001), vol. 17, issue 1, 17–24; Schreurs, Hildebrandt, Kindt and Vanfleteren (n 82) 241–257; Judith Rauhofer, 'Round and Round the Garden? Big Data, Small Government and the Balance of Power in the Information Age' (2014) University of Edinburgh School of Law, Research Paper Series 2014/06, 5, 10 <http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2389981> accessed 15 March 2015. See also Valeria Ferraris et al., 'Working Paper. Defining Profiling' (2014) 15–20 <http://www.unicri.it/special_topics/citizen_profiling/WP1_final_version_9_gennaio.pdf> accessed 15 June 2015; Douwe Korff, 'Data Protection Laws in the EU: The Difficulties in Meeting the Challenges Posed by Global Social and Technical Developments' (2010) 82–89 <http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1638949> accessed 7 April 2015.

[97] See Article 29 Data Protection Working Party, 'Opinion 4/2007 on the concept of personal data' (2007) <http://ec.europa.eu/justice/policies/privacy/docs/wpdocs/2007/wp136_en.pdf> accessed 25 January 2015.

[98] See also above fn. 96. On the blurring of the border between group profiles and personalised profiles, see also Hildebrandt, 'Profiling: From Data to Knowledge. The challenges of a crucial technology' (n 6).

group profiles can be made using anonymised data.[99] This reduces the chances of individuals taking action against biased representations of themselves within a group or having access to the data processing mechanisms, since the anonymised information used for group profiling cannot be linked to them.[100] Even so, group profiling does make it possible to take decisions affecting a multiplicity of individuals.[101] In this sense, the main target of the collective dimension of data processing is not the data subject, but the clusters of people created by big data gatherers.

The interests that assume relevance therefore have a supra-individual nature and a collective dimension,[102] which are not adequately addressed by the existing data protection legal framework. These interests may be shared by an entire group without conflicts between the views of its members (aggregative interests) or with conflicts between the opinions of its members (non-aggregative interests).[103] If the group is characterised by non-aggregative interests, the collective nature of the interest is represented by the fundamental values of a given society (e.g. environmental protection).

The notion of collective non-aggregative interests seems to be the best way to describe the collective dimension of data protection, which becomes important in the discrimination cases mentioned above. Although individuals may have different opinions about the balance between the conflicting interests,[104] there are some collective priorities concerning privacy and data-protection that are of relevance to the general interest. Here the rationale for collective data protection is mainly focussed on the potential harm to groups caused by extensive and invasive data processing.

3.2. *Collective interests in data protection and their representation*

Privacy and data protection are context-dependent notions, which vary from culture to culture and across historical periods.[105] In the same way, the related collective dimensions are necessarily influenced by historical and geographical variables and are the result of action by policymakers. For these reasons, it is impossible to define a common and fixed balance between collective data protection and conflicting interests.

There are jurisdictions that give greater priority to national and security interests, which in many cases prevail over individual and collective data protection; meanwhile, in some countries extensive forms of social surveillance are considered disproportionate and invasive. Therefore, any balancing test must focus on a specific social context in a given historical moment.[106] As has been pointed out in the literature,[107] defining prescriptive ethical guidelines concerning the values that should govern the use of big data analytics and the related balance of interests is problematic.

Given such variability, from a theoretical perspective a common framework for a balancing test can be found in the values recognised by international charters of fundamental

[99] On the limits of anonymisation in the big data context, see Arvind Narayanan, Joanna Huey, Edward W. Felten, 'A Precautionary Approach to Big Data Privacy' (2015) <http://randomwalker.info/publications/precautionary.pdf> accessed 4 April 2015; Arvind Narayanan, Edward W. Felten, 'No silver bullet: De-identification still doesn't work' (2014) <http://randomwalker.info/publications/no-silver-bullet-de-identification.pdf> accessed 25 March 2015; Ohm (n 67); United States General Accounting Office, 'Record Linkage and Privacy. Issues in creating New Federal Research and Statistical Information' (2011) 68–72 <http://www.gao.gov/assets/210/201699.pdf> accessed 14 December 2013; Golle (n 67); Latanya Sweeney, 'Only You, Your Doctor, and Many Others May Know' Technology Science. 2015092903. September 29, 2015 <http://techscience.org/a/2015092903> accessed 28 November 2015; Sweeney, 'Simple Demographics Often Identify People Uniquely' (n 67); Sweeney, 'Foundations of Privacy Protection from a Computer Science Perspective' (n 67).

[100] See Bygrave, *Data Protection Law. Approaching Its Rationale, Logic and Limits* (n 34) 319; Schreurs, Hildebrandt, Kindt and Vanfleteren (n 82) 252–253; Rauhofer (n 96). With regard to the EU Proposal, see also Bert-Jaap Koops, 'The trouble with European data protection law' (2014) 4(4) Int'l. Data Privacy Law 257–258. But see above fn. 99.

[101] This happens, for instance, in the management of smart cities or in the decisions adopted on the basis of credit scoring systems. Against this background, Mireille Hildebrandt observed that "once a profile is linked to an identifiable person – for instance in the case of credit scoring – it may turn into a personal data, thus reviving the applicability of data protection legislation", see Hildebrandt, 'Profiling: From Data to Knowledge. The challenges of a crucial technology' (n 6) 550. See also Vedder (n 9) ("Categorial privacy can be considered as relating to information (1) which was originally taken from the personal sphere of individuals, but which, after aggregation and processing according to statistical methods, is no longer accompanied by identifiers indicating individual natural persons, but, instead, by identifiers of groups of persons, and (2) which, when attached to identifiers of groups and when disclosed, is apt to cause the same kind of negative consequences to the members of those groups as it would for an individual person if the information were accompanied by identifiers of that individual").

[102] See Newman (n 76) 131.

[103] Newman (n 76) 131–132 makes this distinction and defines these two categories of interests respectively as "shared" and "collective" interests. As observed by Finnis, a collective interest in which the conflict is diminished may become a shared interest. See John Finnis, 'The Authority of Law in the Predicament of Contemporary Social Theory' (1984) 1 J.L. Ethics & Pub. Pol'y 115, 135–136.

[104] In this sense, an extensive group profiling for commercial purposes can be passively accepted, considered with favour or perceived as invasive and potentially discriminatory. The same divergence of opinions and interests exists with regard to government social surveillance for crime prevention and national security, where part of the population is in favour of surveillance, due to concerns about crime and terrorism.

[105] See Westin (n 25) 183–187; Whitman (n 48); Bygrave, *Data Protection Law. Approaching Its Rationale, Logic and Limits* (n 34) 327; Nissenbaum (n 72); Irwin Altman, 'Privacy Regulation: Culturally Universal or Culturally Specific?' (1977) Journal of Social Issues 33(3) 66–84.

[106] See in this sense the different attitudes of U.S. government with regard to surveillance, before and after the September 11, 2001, terrorist attacks. See also Bygrave, 'Privacy Protection in a Global Context. A Comparative Overview' (n 48) 329.

[107] See Wright (n 72) 200.

rights. These charters provide a baseline from which to identify the values that can serve to provide ethical guidance and define the existing relationships between these values.[108]

In addition, the context-dependent framework of values and the relationship between conflicting interests and rights needs to be specified with regard to the actual use of big data analytics. In Europe, for instance, commercial interests related to credit score systems can generally be considered compatible with the processing of personal information, providing that the data are adequate, relevant and not excessive in relation to the purposes for which it is collected.[109] Even so, specific big data analytics solutions adopted by some companies for credit scoring purposes may lead to a disproportionate scrutiny of a consumer's private life. The same reasoning can also be applied to smart mobility solutions, which can potentially lead to extensive social surveillance. This means a prior case-by-case risk-assessment is necessary to mitigate the potential impact of these solutions on data protection and individual freedoms.[110]

This "in-context" balance of conflicting interests based on an impact assessment of complex data collection and processing systems,[111] should not be conducted by consumers or companies, but must entail the active involvement of various stakeholders. Against this background, an important aspect of the protection of collective interests relating to personal information is an analysis of the existing conflicting interests and the representation of the issues regarding the individuals grouped in clusters by the data gatherers.[112]

Here it is useful to briefly consider the fields in which the group dimension of data protection is already known in more traditional contexts that are not characterised by extensive data collection and use of analytics. For instance, labour law recognises this collective dimension of rights and the dualism between individuals and groups.[113] Under certain circumstances, trade unions and employees' representatives may concur in taking decisions that affect the employees and have an impact on data protection in the workplace.[114]

Collective agreement on these decisions is based on the recognition that the power imbalance in the workplace means that, in some cases, the employee is unaware of the implications of employer's policies (e.g. employers' workplace surveillance practices). Moreover, in many cases, this imbalance makes it difficult for employees to object to the illegitimate processing of their data.

Entities representing collective interests (e.g. trade unions) are less vulnerable to power imbalance and have a broader vision of the impact of the employer's policies and decisions. It should also be noted that the employer's unfair policies and forms of control are often oriented towards discriminatory measures that affect individual workers, even though they are targeted at the whole group.

This collective representation of common interests is also adopted in other fields, such as consumer protection and environmental protection. These contexts are all characterised by a power imbalance affecting one of the parties directly involved (employees, consumers or citizens). Furthermore, in many cases the conflicting interests refer to contexts where the use of new technologies makes it hard for users to be aware of the potential negative implications.

The same situation of imbalance often exists in the big data context, where data subjects are not in a position to object to the discriminatory use of personal information by data

[108] See Wright (n 72) 201–202.

[109] See Articles 18 and 20 of the Directive 2014/17/EU. See also Article 8 of the Directive 2008/48/EC on credit agreements for consumers and repealing Council Directive 87/102/EEC.

[110] See e.g. Italian Data Protection Authority (Autorità Garante per la protezione dei dati perosnali, hereafter Gar.), 6 September 2006 (decision, doc. web n. 1339531) <http://www.garanteprivacy.it/web/guest/home/docweb/-/docweb-display/docweb/1339531> (text in Italian); Gar., 28 December 2006 (decision, doc. web n. 1413380) <http://www.garanteprivacy.it/web/guest/home/docweb/-/docweb-display/docweb/1413380> (text in Italian); Gar., 6 September 2006 (decision, doc. web n. 1339692) <http://www.garanteprivacy.it/web/guest/home/docweb/-/docweb-display/docweb/1339692> (text in Italian). See also Alessandro Mantelero, 'Data protection, e-ticketing and intelligent systems for public transport' (2015) 5(4) Int'l. Data Privacy Law 309–320.

[111] Moreover, these systems are influenced by lock-in effects. There are two different kinds of lock-ins: technological lock-in and social lock-in. The first is related to the technological standards and data formats that are adopted by service providers. This lock-in represents a limit to data portability and migration from one service to another. The second lock-in (social lock-in) is related to the dominant position held by some big players. This lock-in is evident, for example, in the social networks market, where there is an incentive to remain on a network, given the numbers of social relationships created by users.

[112] On group right, see also Julia Stapleton, *Group rights: perspectives since 1900* (Thoemmes Press 1995); Will Kymlicka, 'Individual and Community Rights' in Judith Baker (ed), *Group Rights* (University of Toronto Press 1994) 17–33; Eugene Schlossberger, A Holistic Approach to Rights: Affirmative Action, Reproductive Rights, Censorship, and Future Generations (University Press of America 2008) ch. 10; Helen O'Nions, *Minority Rights Protection in International Law: The Roma of Europe* (Ashgate Publishing 2007) 26–28; Peter Jones, 'Group Rights' (2008) in *Stanford Encyclopedia of Philosophy* http://stanford.library.usyd.edu.au/entries/rights-group/ accessed 4 July 2015; Newman (n 76); Will Kymlicka, 'The New Debate Over Minority Rights' in Wayne Norman and Ronald Beiner (eds.) *Canadian Political Philosophy: Contemporary Reflections* (Oxford University Press 2000) 159–176; John Packer, 'Problems in Defining Minorities' in Deirdre Fottrell and Bill Bowring (eds), *Minority and group rights in the new millennium* (M. Nijhoff Publishers 1999) 241 ff; Will Kymlicka, 'Individual and Community Rights' in Judith Baker (ed) *Group Rights* (University of Toronto Press 1994) 17–33.

[113] See Italian Articles 4 and 8, Act 300, 20 May 1970 (Statute of the Workers' Rights).

[114] See above at fn. 52. See also Bygrave and Wiese Schartum (n 41) 170.

gatherers.[115] Data subjects often do not know the basic steps of data processing,[116] and the complexity of the process means that they are unable to negotiate their information and are not aware of the potential collective prejudices that underlay its use.[117] This is why it is important to recognise the role of entities representing collective interests, as happens in the above cases.

Employees are part of a specific group, defined by their relationship with a single employer; therefore, they are aware of their common identity and have mutual relationships. By contrast, in the big data context, the common attributes of the group often only become evident in the hands of the data gatherer.[118]

Data subjects are not aware of the identity of the other members of the group, have no relationship with them and have a limited perception of their collective issues. Furthermore, these groups shaped by analytics have a variable geometry and individuals can shift from one group to another.

This does not undermine the idea of a representing collective data protection interests. On the contrary, this atomistic dimension makes the need for collective representation more urgent. However, it is hard to imagine representatives appointed by the members of these groups, as is instead the case in the workplace.

In this sense there are similarities with consumer law, where there are collective interests (e.g. product security, fair commercial practices), but the potential victims of harm have no relationship to one another. Thus, individual legal remedies must be combined with collective remedies.[119] Examples of possible complementary solutions are provided by consumer law, where independent authorities responsible for consumer protection, class action lawsuits and consumer associations play an important role.

In the field of big data analytics, the partially hidden nature of the processes and their complexity probably make timely class actions more difficult than in other fields. For instance, in the case of a product liability, the damages are often more evident making it easier for the injured people to react.[120] On the other hand, associations that protect collective interests can play an active role in facilitating reaction to unfair practices and, moreover, they can be involved in a multi-stakeholder risk assessment of the specific use of big data analytics.[121]

The involvement of such bodies requires specific procedural criteria to define the entities that may act in the collective interest.[122] This is more difficult in the context of big data, where the groups created by data gatherers do not have a stable character. In this case, an assessment of the social and ethical impact of analytics often provides the opportunity to discover how data processing affects collective interests and thus identify the potential stakeholders.[123]

3.3. *The role of data protection authorities*

How collective interests should be protected against discrimination and social surveillance in the use of big data analytics is largely a matter for the policymakers. Different legal systems and different balances between the components of society suggest differing solutions. Identifying the independent authority charged with protecting collective interests may therefore be difficult.

[115] In the digital economy, consumers often accept not having an effective negotiation of their personal information, due to market concentration and related social and technological lock-ins. See above fn. 111.

[116] See also Alessandro Acquisti, Laura Brandimarte and George Loewenstein, 'Privacy and human behavior in the age of information' (2015) 347(6221) Science 509–514.

[117] The complexity of data processing in the big data environment does not offer users a real chance to understand it and make their choice. See Pasquale (n 9) 143–144; Laura Brandimarte, Alessandro Acquisti, and George Loewenstein, 'Misplaced Confidences: Privacy and the Control Paradox' (2010), Ninth Annual Workshop on the Economics of Information Security <http://www.heinz.cmu.edu/~acquisti/papers/acquisti-SPPS.pdf> accessed 27 February 2014; Joseph Turow, Chris Jay Hoofnagle, Deirdre K. Mulligan, and Nathaniel Good, 'The Federal Trade Commission and Consumer Privacy in the Coming Decade' (2007) 3 ISJLP 723–749 <http://scholarship.law.berkeley.edu/facpubs/935> accessed 27 February 2014; Federal Trade Commission (n 2) 42. On the limits of the traditional notices, see also Rayan M. Calo, 'Against Notice Skepticism in Privacy (and Elsewhere)' (2013) 87(3) Notre Dame L. Rev. 1027, 1050–1055 <http://scholarship.law.nd.edu/ndlr/vol87/iss3/3> accessed 27 February 2014; Daniel J. Solove, 'Introduction: Privacy Self-management and The Consent Dilemma' (2013) 126 Harv. L. Rev. 1880, 1883–1888; World Economic Forum, 'Unlocking the Value of Personal Data: From Collection to Usage' (2013) 18 <http://www3.weforum.org/docs/WEF_IT_UnlockingValuePersonalData_CollectionUsage_Report_2013.pdf> accessed 27 February 2014.

[118] See also Bygrave, *Data Protection Law. Approaching Its Rationale, Logic and Limits* (n 34) 283–284.

[119] The same approach has been adopted in the realm of anti-discrimination laws; see European Commission, 'Developing Anti-Discrimination Law in Europe. The 28 EU Member States, the Former Yugoslav Republic of Macedonia, Iceland, Liechtenstein, Norway and Turkey compared' (n 82) 83–110. See also Lilla Farkas, 'Collective actions under European anti-discrimination law' (2014) 19 European Anti-Discrimination Law Rew. 25–40.

[120] As demonstrated by recent revelations on NSA case, in the big data context people are not usually aware of being under surveillance. Only leaks of information can disclose these practices, open a debate on their legitimacy and give the chance to individuals to bring legal actions. See also European Parliament, 'Resolution of 4 July 2013 on the US National Security Agency surveillance programme, surveillance bodies in various Member States and their impact on EU citizens' privacy' (2013) <http://www.europarl.europa.eu/sides/getDoc.do?pubRef=-//EP//TEXT+TA+P7-TA-2013-0322+0+DOC+XML+V0//EN> accessed 27 February 2014. On the role played by group actions in order to protect individual and collective interests concerning personal information, see Bygrave, *Data Protection Law. Approaching Its Rationale, Logic and Limits* (n 34), 288–290.

[121] See also para 3.3.

[122] See art. 76 of the EU Proposal.

[123] For these reasons, a preventive approach based on risk assessment seems to be more effective than *ex post* legal actions. Moreover, it also contributes to tackle the risks of hidden forms of data processing, which often create an asymmetric distribution of the control over information in our society. See also Alessandro Mantelero, 'Social Control, Transparency, and Participation in the Big Data World' (2014) April *Journal of Internet Law* 23–29.

Many countries have independent bodies responsible for supervising specific social surveillance activities, and other bodies focused on anti-discrimination actions.[124] In other countries, this responsibility is spread across various authorities, which take different approaches, use different remedies and do not necessarily cooperate in solving cases with multiple impacts.

Meanwhile, a central element in the risk-assessment of big data analytics is the analysis of data processing, the factor common to all these situations, regardless of the potential harm to collective interests. For this reason, data protection authorities can play a key role in the risk assessment processes, even if they are not focused on the specific social implications (e.g. discrimination).

On the other hand, if we take a different approach that takes into consideration the various negative effects generated by the use of big data (discrimination, unfair consumer practices, social control, etc.), we should involve multiple entities and authorities. Nevertheless, as we have seen the end result may be a fragmented and potentially conflicting decision-making process that may underestimate the use of data, which is the common core of all these situations.[125]

Furthermore, the data protection authorities are accustomed to addressing collective issues and have already demonstrated that they do consider both the individual and the wider collective dimension of data processing.[126] Focusing on data protection and fundamental rights, they are also well placed to balance the conflicting interests around the use of data.

The adequacy of the solution is also empirically demonstrated by important cases decided by data protection authorities concerning data processing projects with significant social and ethical impacts.[127] These cases show that decisions to assess the impact of innovative products, services and business solutions on data protection and society are not normally on the initiative of the data subjects, but primarily on that of the data protection authorities who are aware of the potential risks of such innovations. Based on their balancing tests, these authorities are in a position to suggest measures that companies should adopt to reduce the risks discussed here and to place these aspects within the more general framework of the rights of the individual, as a single person and as a member of a democratic society.

The risk assessment represents the opportunity for group issues to be identified and addressed. Thus, bodies representing collective interests should not only partially exercise traditional individual rights on behalf of data subjects,[128] but also exercise other autonomous rights relating to the collective dimension of data protection. These new rights mainly concern participation in the risk assessment process, which should take a multi-stakeholder approach.[129]

Against this background, data protection authorities may involve in the assessment process the various stakeholders, which represent the collective interests affected by specific data processing projects.[130] This would lead to the definition of a new model in which companies that intend to use big data analytics would undergo an assessment prior to collecting and processing data.

[124] See European Commission, 'Developing Anti-Discrimination Law in Europe. The 28 EU Member States, the Former Yugoslav Republic of Macedonia, Iceland, Liechtenstein, Norway and Turkey compared' (n 82) 113–125.

[125] See also Lerman (n 86) 60, who points out the limits of the U.S. equal protection doctrine in the context of big data analytics.

[126] See fn. 52. See also e.g. Article 29 Data Protection Working Party, 'Explanatory Document on the Processor Binding Corporate Rules' (2015) <http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2015/wp204.rev_en.pdf> accessed 29 November 2015; Article 29 Data Protection Working Party, 'Opinion 9/2014 on the application of Directive 2002/58/EC to device fingerprinting' (2014) <http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2014/wp224_en.pdf>; Article 29 Data Protection Working Party, 'Opinion 8/2014 on the Recent Developments on the Internet of Things' (2014) <http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2014/wp223_en.pdf>; Article 29 Data Protection Working Party, 'Opinion 03/2012 on developments in biometric technologies' (2012) <http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2012/wp193_en.pdf>; Article 29 Data Protection Working Party, 'Opinion 02/2012 on facial recognition in online and mobile services' (2012) <http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2012/wp192_en.pdf>.

[127] See above fns. 52 and 126.

[128] The stakeholders may have right of access to the documents that describe the specific structure and general purposes of big data processing. However, in order to protect the legitimate interests of companies and governments, the data protection authorities might limit this disclosure to third parties. See also art. 76 of the EU Proposal and Bygrave, *Data Protection Law. Approaching Its Rationale, Logic and Limits* (n 34) 274–282.

[129] Note that the extent of the rights conferred upon the different stakeholders in the protection of collective privacy is largely a matter for policymakers to decide and would depend on the nature and values of the different socio-legal contexts.

[130] See also Wright (n 72) 201–202, 215–220; Danielle Keats Citron, 'Technological Due Process' (2008) 85(6) Wash. U. L. Rev. 1249, 1312. A different assessment exclusively based on the adoption of security standards or corporate self-regulation would not have the same extent and independency. This does not mean that, in this framework, forms of standardisation or co-regulation cannot be adopted.

The assessment would not only focus on data security and data protection,[131] but also consider the social and ethical impacts relating to the collective dimension of data use in a given project.[132] This assessment should be conducted by third parties and supervised by the data protection authorities.[133] Once this multiple-impact assessment is approved by the data protection authorities, the ensuing data processing would be considered secure in protecting personal information and collective interests.[134]

Although data protection authorities are already engaged to some degree in addressing the collective dimension,[135] the suggested solution would lead to a broader and deeper assessment, which would become mandatory.[136] This proposal is therefore in line with the view that a licensing scheme might "prove to be the most effective means of ensuring that data protection principles do not remain 'law-in-book' with respect to profiling practices".[137] Finally, it should be noted that a different risk assessment model, which also takes into account the ethical and social effects of data use, directly affects data processing

[131] On traditional forms of privacy impact assessment, see Roger Clarke, 'Privacy impact assessment: Its origins and development' in this Review 2009, vol. 25, issue 2, 123–129; David Flaherty, 'Privacy impact assessments: an essential tool for data protection' (2000) 7(5) Priv. Law & Pol'y Rep. 45; David Wright, 'The state of the art in privacy impact assessment' in this Review 2012, vol. 28, issue 1 54–61; David Wright and Paul De Hert (eds), *Privacy Impact Assessment* (Springer 2012); David Wright, Michael Friedewald, and Raphael Gellert, 'Developing and Testing a Surveillance Impact Assessment Methodology' (2015) 5(1) Int'l. Data Privacy Law 40–53. See also Commission nationale de l'informatique et des libertés, 'Étude d'impact sur la vie privée (EIVP)' (2015) <http://www.cnil.fr/fileadmin/documents/Guides_pratiques/CNIL-PIA-1-Methode.pdf> accessed 16 July 2015.

[132] In the big data context, another important aspect is the transparency of the algorithms used by companies. See Citron and Pasquale (n 9) 5, 10–11, 25, 31; Pasquale (n 9) 193, 216–218. See also Viktor Mayer-Schönberger and Kenneth Cukier, *Big Data. A Revolution That Will Transform How We Live, Work and Think* (n 12) 179–182; they suggest a model based on independent internal and external audits. A wider access to the logic of the algorithms was required by Article 29 Data Protection Working Party, 'Opinion 03/2013 on purpose limitation' (2013) 47 <http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2013/wp203_en.pdf> accessed 27 February 2014. See also Tarleton Gillespie, 'The Relevance of Algorithms' in Tarleton Gillespie, Pablo J. Boczkowski, and Kirsten A. Foot (eds) *Media Technologies. Essays on Communication, Materiality, and Society* (MIT Press 2014) 167, 194; Dixon and Gellman (n 9), 7 ("Trade secrets have a place, but secrecy that hides racism, denies due process, undermines privacy rights, or prevents justice does not belong anywhere"). But see Recital n. 51 of EU Proposal, text adopted by the Council of the European Union, Brussels, 19 December 2014 ("Every data subject should therefore have the right to know and obtain communication in particular for what purposes the data are processed [. . .], what is the logic involved in any automatic data processing and what might be, at least when based on profiling, the consequences of such processing. This right should not adversely affect the rights and freedoms of others, including trade secrets or intellectual property and in particular the copyright protecting the software"). On the interest in knowing the logic of profiling, see also Schreurs, Hildebrandt, Kindt and Vanfleteren (n 82) 253–256. On transparency in decisional processes based on big data analytics, see also Zarsky (n 17) 1523–1530.

[133] The entire system will work only if the political and financial autonomy of data protection authorities from governments and corporations is guaranteed. Moreover, data protection authorities would need new competence and resources in order to bear the burden of the supervision and approval of these multiple-impact assessments. For these reasons, a model based on mandatory fees – paid by companies when they submit their requests for authorisation to data protection authorities – would be preferable. See Mantelero, 'The future of consumer data protection in the E.U. Rethinking the "notice and consent" paradigm in the new era of predictive analytics' (n 20). It should also be noted that, in cases of large scale and multinational data collection, forms of mutual assistance and cooperation may facilitate the role played by data protection authorities in addressing the problems related to the dimensions of data collection and data gatherers. See also Gwendal Le Grand and Emilie Barrau, 'Prior Checking, a Forerunner to Privacy Impact Assessments' in Wright and De Hert (n 131) 112–116.

[134] Therefore, in this scenario, companies can enlist users in the data processing without any prior consent, provided they give notice of the results of the assessment and provide an opt-out option. See more extensively Mantelero, 'The future of consumer data protection in the E.U. Rethinking the "notice and consent" paradigm in the new era of predictive analytics' (n 20), 654–659. In this case, although this assessment represents an economic burden for companies, it allows those who pass to use data for complex and multiple purposes, without requiring users to opt-in. At the same time, from the users' side, the assessment supervised by data protection authorities provides an effective evaluation of risks, while the option to opt-out allows users to choose to not be a part of the data collection. See also Citron and Pasquale (n 9) 24–28. The suggested model represents a significant change in the traditional approach to data protection, but this is in line with the approach adopted in other fields characterised by the presence of risks for individuals and society (e.g. authorisation procedure for human medicines, mandatory security standards provided by product liability laws, security standards for industrial activities). For this reason, it would be necessary to adopt a subset of rules for big data analytics, which focuses on multiple risk assessment and a deeper level of control by data protection authorities.

[135] See above fns. 52 and 126.

[136] See fns. 133, 134.

[137] See Bygrave, *Data Protection Law. Approaching Its Rationale, Logic and Limits* (n 34), 373–374. See also Lee A. Bygrave, *Data Privacy Law. An International Perspective* (Oxford University Press 2014) 183–186.

design. Literature on privacy by design[138] has clearly stressed the relevance of embedding privacy values in the device and services architecture. To achieve this goal, a preliminary analysis of the use of personal information in each specific application (good or service) is required to shape this use according to data protection purposes. Based on this assessment, suitable measures would be taken to reduce the potential negative outcomes of data use.

This strict relationship between risk assessment and solutions by design implies that any change in the nature of the assessment would affect the architectural solutions adopted. Thus, the multiple impact assessment suggested would necessarily lead companies to consider a broader range of by-design solutions to mitigate the additional ethical and social concerns.[139]

4. Conclusions

This article does not provide complete answers to the various issues concerning the collective dimension of privacy and data protection relating to the use of big data analytics. The analysis offers an introductory study of a new approach to group privacy that would appear necessary to adequately consider the non-aggregative interests arising from the data-driven society.

Big data analytics are creating a new digital landscape that cannot be described as a mere increase in the quantity of information processed. The predictive nature of the inferences extracted from databases, the complexity of data processing and its obscurity, as well as the categorical approach, distinguish it from previous profiling solutions.

From the group perspective, big data analytics create new kind of groups, which cannot be compared with the traditional concept of a group. They are the result of aggregations of information produced by data gatherers and have a variable geometry, setting them apart from the previous static categories used for group profiling. Moreover, data subjects are not aware of the identity of the other members of the group and have a limited or no perception of their collective issues, whereas in traditional groups there is an awareness of being part of a group and groups have external visibility.

The new scale of data processing, the pervasive diffusion of data-based applications, the evolution and complexity of group profiling represent important changes with respect to the previous scenario.[140] At the dawn of the data-driven society, the question arises whether it is necessary to reconsider the traditional approach to group privacy and data protection, which is mainly based on the model of individual rights.

This article gives an affirmative response to this question on the basis of the impact that big data analytics have on data processing and data-driven decisions. The shift in the data processing paradigm and the new forms of categorical approach have a disruptive effect on the traditional idea of group privacy and highlight its limits.

The new scale entails the recognition of a new layer, represented by groups' need for the protection of their collective data protection rights. In this scenario, data protection concerns not only individuals, but also the collective dimension, associated with potential harm to groups in terms of discriminatory and invasive forms of data processing.

However, collective interests require adequate forms of representation, as well as the involvement of a range of stakeholders in the balancing of conflicting interests. Specific procedural criteria must be laid down to define which entities may act in the collective interest, and this decision is made more difficult in the context of big data by the lack of stability in the nature of groups created by data gatherers.

[138] See Ann Cavoukian, 'Privacy by design. From rhetoric to reality' (2014) 12–18, 65–100 <http://www.ipc.on.ca/images/Resources/PbDBook-From-Rhetoric-to-Reality.pdf> accessed 27 February 2014; Ann Cavoukian, 'Privacy by Design: Leadership, Methods, and Results' in Serge Gutwirth, Ronald Leenes, Paul De Hert, Yves Poullet (eds), *European Data Protection: Coming of Age* (Springer 2013) 175–202; Ann Cavoukian, 'Privacy by Design: Origins, Meaning, and Prospects for Assuring Privacy and Trust in the Information Era' in Yee, G O M (ed), *Privacy Protection Measures and Technologies in Business Organizations: Aspects and Standards* (IGI Global 2012) 170–208; Ira S. Rubenstein, 'Regulating Privacy By Design' (2011) 26 Berkeley Tech. L. J. 1409–1456; Peter Schaar, 'Privacy by Design' (2010) 3(2) Identity in the Information Society 267–274. See also Article 29 Data Protection Working Party, 'Opinion 07/2013 on the Data Protection Impact Assessment Template for Smart Grid and Smart Metering Systems ('DPIA Template') prepared by Expert Group 2 of the Commission's Smart Grid Task Force' (2013) <http://ec.europa.eu/justice/dataprotection/article-29/documentation/opinion-recommendation/files/2013/wp209_en.pdf> accessed 27 February 2014; Article 29 Data Protection Working Party, 'Opinion 9/2011 on the revised Industry Proposal for a Privacy and Data Protection Impact Assessment Framework for RFID Applications' (2011) <http://ec.europa.eu/justice/policies/privacy/docs/wpdocs/2011/wp180_en.pdf> accessed 27 February 2014; Woodrow Hartzog and Frederic Stutzman, 'Obscurity by Design' (2013) 88 Wash. L. Rev. 385, 397; Federal Trade Commission, 'Protecting Consumer Privacy in an Era of Rapid Change. Recommendations for Business and Policymakers' (2012) 22–24 <http://www.ftc.gov/sites/default/files/documents/reports/federal-trade-commission-report-protectingconsumer-privacy-era-rapid-change-recommendations/120326privacyreport.pdf> accessed 25 June 2014.

[139] See Wright (n 72).

[140] In 2014, the 90% of the world's data were generated in the last two years, while the remaining 10% were produced through the rest of humanity history. In 2014, the speed of common computer processors was around 100–200 thousand MIPS (million instructions per second), in 1994, it was around 180. In 2012, there were 8.7 billion of connected devices (so-called Internet of Things); in 2020 they will be 50 billion. See CNN, 'The data rush: How information about you is 21st century "gold" ' (November 13, 2014) <http://edition.cnn.com/2014/11/04/tech/gallery/big-data-techonomics-graphs/>; Wikipedia, 'Instructions per second', *Wikipedia, the free encyclopedia* (2015) <https://en.wikipedia.org/wiki/Instructions_per_second>; Cisco, 'Seize New IoT Opportunities with the Cisco IoT System' <http://www.cisco.com/web/solutions/trends/iot/portfolio.html>. All these sources were accessed on 12 June 2015.

In this context, the assessment of the social and ethical impact of analytics may represent an opportunity to discover how data processing affects collective interests and similarly an opportunity to identify the potential stakeholders. Meanwhile, the assessment also represents the principal mechanism by which conflicting interests relating to the context-dependent notion of collective data protection can be balanced.

Finally, given the central role of data processing analysis in risk-assessment, data protection authorities can play a key role in the assessment process and licensing models can be reconsidered in the specific context of big data analytics.

Acknowledgements

I am indebted to all who provided feedback during the first presentations of my thoughts on this topic at the 6th International Conference on Information Law and Ethics (Thessaloniki, 30–31 May 2014) and at the 9th International Conference on Legal, Security and Privacy Issues in IT Law (Lisbon, 15–17 October 2014). I would like to thank Prof. Lee Bygrave, Prof. Leonardo Lenti and Dr. Giuseppe Vaciago for their helpful suggestions. I am grateful to the anonymous CLSR reviewers for their constructive comments.