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.