Big data, privacy and information security in the European Union

Luigia Altieri[a*], Gianmarco Cifaldi[b]
[a] *University "G. d'Annunzio" of Chieti-Pescara, Italy*
[b] *University "G. d'Annunzio" of Chieti-Pescara, Italy*

Abstract
This article analyzes the concepts of "big data", "personal data", "maintaining data control", "the right to be forgotten", all this in the context of the new legislation imposed by the European Union. The question is whether the new European legislation has positive or negative effects on social development? The material emphasizes the fact that, in the context of the current impressive development of information technology, the recent Regulation (EU) 2016/679 of the European Parliament and of the Council, dated 27 April 2016, concerning the protection of natural persons with regard to the processing of personal data, as well as free movement of such data, seems to be in contrast, and may have negative effects on the development of the economy.

Keywords*: big data; privacy; security information; General Data Protection Regulation; European Union.*

1. About the Big Data

The Big Data "is an emerging area applied to manage datasets whose size is beyond the ability of commonly used software tools to capture, manage, and timely analyze that amount of data. The quantity of data to be analyzed is expected to double every two years" (IDC, 2012; Michael and Miller 2012: pp. 22-24; IEE BigData 2018). All these data are very often unstructured and "from various sources such as social media, sensors, scientific applications, surveillance, video and image archives, Internet search indexing, medical records, business transactions and system logs" (IEE BigData 2018).

Big data is "gaining more and more attention since the number of devices connected to the so-called "Internet of Things" (IoT) is still increasing to unforeseen levels, producing large amounts of data which needs to be transformed into valuable information" (MIT 2014; IEE BigData 2018). "Additionally, it is very popular to buy on-demand additional computing power and storage from public cloud providers to perform intensive data-parallel processing. In this way, security and privacy issues can be potentially boosted by the volume, variety, and wide area deployment of the system infrastructure to support Big Data applications" (IEE BigData 2018).

It has been repeating for many years the consideration that the one we live in is the information society, but this synthesis of the current relevance recognized in all

[*]Luigia Altieri. Tel.: +39 3913-286-753. *E-mail address: luigia.altieri@unich.it*.

areas to the processes of generation, processing and communication of information data is not an immutable acquisition, a definitive arrival point. On the contrary, the informative phenomenon is varied over time due to the different parameters concerning its distribution, size and the possibility of an effective analysis of the contents.

With regard to the first aspect, it should be noted that the information is not uniformly distributed as regards the usability of the same. First of all, it is necessary to distinguish between accessibility and data exploitation capacity; factors that concur then both to concentrate the information power in groups of subjects gradually more limited, up to what we could define the ”gentlemen of data” (The traditional notion of the individual's lordship on information concerning him seems weakened in favor of the holders of information and calculation resources, which have considerable information power deriving from the control over the management of data, such as to evoke the notion of lordship).

Again we are in the presence of a limited number of operators in whose hands is concentrated a great wealth of information and who have the opportunity to choose who can access it, very often in exchange for an economic benefit.

An element that completes the illustration of the power held by ”gentlemen of data” is psychological in nature. Recent studies have in fact shown how the awareness of the availability of multiple information online induces the subjects to memorize less than what is learned and to focus instead the memory on where the data are located. From this complex context, summarily summarized here, the ”gentlemen of data” and their immense power are born.

It is therefore evident that those who manage large amounts of data are thus able to acquire a predictive capacity on the future to the other foreclosed, constituting an undoubted advantage, both in competitive terms for the companies, both in terms of social control for the states and for the groups of power.

2. General Data Protection Regulation

The legal definition of ”personal data”, contained in the recent ”Regulation (EU) 2016/679 of the European Parliament and of the Council, of 27 April 2016, concerning the protection of natural persons with regard to the processing of personal data, as well as the free movement of such data and which repeals Directive 95/46/EC (General Data Protection Regulation)” - which is known to be directly applied in all EU Member States from 25 May 2018, highlights the complexity of legal issues, raised from collection and subsequent processing operations performed on personal data. According to the art. 4, par. 1, n. 1 of the aforementioned Regulation, ”personal data” means ”any information concerning an identified or identifiable natural person ("concerned"); an identifiable natural person can be identified, either directly or indirectly, with particular reference to an identifier such as a name, an identification number, location data, an online ID or one or more characteristic elements of his physical identity, physiological, genetic, psychological, economic, cultural or social”.

The amplitude of the definition is such that only the anonymous data cannot be traced back to it and once it is established that the information is directly or indirectly connected to a physical person it is necessary to ”reckon” with the complex architecture outlined by the Regulation , based on the principles of "privacy by design" and "privacy by default", as well as, more generally, by a general criterion of accountability. From

this point of view, the Project aims to analyze the constraints imposed by the Regulation underlying the processing of personal data, the obligations imposed on the data controller, with particular attention, considered the focus of the research, to the limits deriving from the communication of data, for the purposes underlying the original collection, to non-EU subjects (Moura and Serrão 2015).

In the awareness that personal data, although increasingly frequently placed under the lens of economic evaluation and consequently its central importance in the digital economy, represents, as a priority, a component of the personal identity of the person to whom the data refers. A person ”who maintains a control over that data, exercisable through various prerogatives recognized expressly by law”. In this regard, ”the right to delete personal data”, enshrined in art. 17 of the aforementioned Regulations and the ”right to be forgotten, declined in different meanings by national and Community jurisprudence”. The research, therefore, aims to ”reconstruct the rights recognized to the person (physical person to whom the personal data refer) with particular regard to the claim of the cancellation and the right to be forgotten, so as to detect if they are independent, distinct claims that is, one from the other and if, above all, we are dealing with claims that can be exercised ad libitum”. Finally, but not least, it is necessary to question the possible uses of the information gathered, on a large scale. Stringent, in fact, are the limits in the case in which the data are processed with automated processes in a particular way in the emerging cloud computing computational paradigm. Article. 22 of the Regulation establishes, in fact, the right of the interested party "not to be subjected to a decision, based solely on automated processing, including profiling, which produces legal effects that affect it or significantly affects its person". There are two key concepts on which this prediction is based and consequently, the recognition of the right of the data subject to escape the treatment: the automated treatment and the decision based solely on it. Two concepts on which the research will focus attention, so as to evaluate the limits of profiling, which in fact represents an automated treatment, increasingly widespread in the digital economy (Marques and Serrão 2013; Serrão, Rodriguez and Delgado 2011; pp. 129-139).

The power front deriving from big data and the largely hidden nature of the same, we must ask ourselves about the remedies that can be introduced in order to limit the asymmetries and the implications in terms of social control that derive from them. In this regard it does not seem appropriate to resort to drastic remedies such as the obligation to delete information after a certain time. Obligation that, in the context of free online access information, we would like to refer to the right to be forgotten. Without dwelling on the known difference between the right to delete data and the right to be forgotten, it should however be noted that such a solution seems scarcely feasible in relation to online communications and not very efficient with respect to the creation of large databases. To this we must add that in many cases the databases on which the elaborations under consideration are carried out do not necessarily contain personal data, but they are not therefore less relevant for predictive purposes (Serrão, Neves, Kudumakis, Barker and Balestri 2003: p. 648).

In other words, the methods of intervention seem to be more effective in reducing and redistributing the information power held by few and in limiting possible abuses. To do this, we need to act both on the market, stimulating competition and thus

favoring the pluralism of actors, as well as on accessibility to information. In fact, if the data held by the subjects in question were largely accessible, it would open to new interested parties the possibility to draw from the same inferences (not only large operators not yet present in the market, but also large groups of individuals able to aggregate beyond the critical threshold the limited resources of each).

The importance assumed in terms of information power from large concentrations of data, together with the strategic value also for the nations of the same, should then lead to evaluate the adoption of forms of control of such aggregations of data, providing for specific independent supranational authorities and introducing notification obligations (De Cristofaro, Soriente, Tsudik, and Williams 2012: p. 287; Jutla, Bodorik and Ali 2013: pp. 39-45). It is no coincidence that the notification of the establishment of the new databases was one of the obligations characterizing the first legislation on the processing of personal data, in a time when IT resources were the heritage of a few, centralized in specific places and porters (for those years) of an unimaginable power of control. The analogy between the era of the main frames and the current one of cloud computing and big data is significant, because once again (while remaining a distributed computer power) large IT resources are concentrated in the hands of a few subjects and are also physically aggregated into huge data centers. It is therefore once again possible to know who creates such large databases, who manages them and, therefore, put in place the control activities necessary to guarantee the security of information concerning citizens (Tankard 2012).

In this sense, the creation of supranational control authorities should on the one hand affect the standardization of services in terms of security, but should also serve as a tool to monitor and possibly contain both the invasive claims of governments and the possible abuses of the holders / big data managers (Juels and Oprea 2013).

Given the new "personal data" regulation contained in the recent Regulation (EU) 2016/679 of the European Parliament and of the Council, dated 27 April 2016, concerning the protection of natural persons with regard to the processing of personal data, as well as free movement of such data, this solution could appear to be in contrast with the spontaneous evolution of information and business processes, and to be seen as an undue interference of public powers in the development of economic systems and models, nevertheless it should be remembered that precisely the international dimension of internet has progressively led to a revision of the initial dogma centered on the simple self-regulation of users. In a society heavily influenced by information, by its appropriation and management, it does not seem so illogical that these resources, if they assume very significant proportions, are placed under limited control as is the case for other vital and strategic resources (from nuclear power plants to air spaces). This is obviously a very long and tortuous path because it requires international synergies, but it should be started as soon as possible, perhaps using existing organisms, to avoid introducing forms of regulation when it will be too late. Obviously it is appropriate to appropriately scale the intervention, addressing it not to any data farm made in any part of the globe, but only to those that have an absolutely relevant size or that, due to the data processed (eg national and military security) (Hand, Ton and Keller 2013), are of primary importance.

A final critical profile that needs to be overcome concerns the role of Europe in the context of the aggregation processes of the information taken into consideration. In

this regard, the comparison with the USA is important and a position of advantage is found on the latter. In terms of public data management, the United States boasts not only a structural homogeneity that is still unknown to the young European Union, but has also invested significant resources in the modernization of the same through the use of cloud computing technologies, encouraging the aggregation of the bases of data, a necessary precondition both for the increase and the exercise by the public subjects of their information power, and for the democratic accessibility to such information (Advantech 2013; Agrawal, Das, and El Abbadi 2011: p. 580,). On the private side, then, although it is undeniable that the big European companies are fully part of the big data managers, but the excellence of US companies in some strategic ICT sectors (search engines, cloud computing services, platforms UGC, social network) (Chen and Shi 2009: p. 95; Rodríguez, Rodríguez, Carreras and Delgado 2009; Feamster 2014; Gross and Acquisti 2005: pp. 71-80), put the latter in a position of advantage, given that precisely in the areas most closely related to ICT we are witnessing the largest data flows. In a geo-political and industrial policy perspective, this structure may prove to be a weakness for European countries, in terms of loss of control over citizens' data and the assignment of management of strategic information to foreign subjects. It is perhaps not by chance that the idea of a stronger personal data protection centered on the concept of belonging to the Union of the person to whom the information refers has recently come forward, as it is no coincidence that European industry is urged to take on a more relevant role in the implementation of new IT architectures. Disregarding such solicitations, can be a risk because it involves the use of services provided by companies linked to powers and conditioned by foreign legal rules, which can potentially pursue finality purposes from those of citizens and states of the Union. It is therefore necessary both to stimulate competition in the development of new ICT technologies, and to strengthen the regulatory framework for protecting information.

The power of big data and the largely hidden nature of it, we must build mechanisms that can be introduced in order to limit the asymmetries and the implications in terms of social control that derive from them. In this regard it does not seem appropriate to resort to drastic remedies such as the obligation to delete information after a certain time. Obligation that, in the context of free online access information, we would like to refer to the right to be forgotten.

To this we must add that in many cases the databases on which the elaborations under consideration are carried out do not necessarily contain personal data, but they are not therefore less relevant for predictive purposes.

In other words, the methods of intervention seem to be more effective in reducing and redistributing the information power held by few and in limiting possible abuses. To do this, we need to act both on the market, stimulating competition and thus favoring the pluralism of actors, as well as on accessibility to information (Goldwasser, Gordon, Goyal, Jain et. al. 2014: p. 580; Gentry 2009; Gentry 2010; Dohi and Uemura 2012: p. 1751). In fact, if the data held by the subjects in question were largely accessible, it would open to new interested parties the possibility to draw from the same inferences (not only large operators not yet present in the market, but also large groups of individuals able to aggregate beyond the critical threshold the limited resources of each).

The importance assumed in terms of information power from large concentrations of data, together with the strategic value also for the nations of the same, should then lead to evaluate the adoption of forms of control of such aggregations of data, providing for specific independent supranational authorities and introducing notification obligations. It is no coincidence that the notification of the establishment of the new databases was one of the obligations characterizing the first legislation on the processing of personal data, in a time when IT resources were the heritage of a few, centralized in specific places and porters (for those years) of an unimaginable power of control. The analogy between the era of the main frames and the current one of cloud computing and big data is significant, because once again (while remaining a distributed computer power) large IT resources are concentrated in the hands of a few subjects and are also physically aggregated into huge data centers. It is therefore once again possible to know who creates such large databases, who manages them and, therefore, put in place the control activities necessary to guarantee the security of information concerning citizens.

Personal data ”is any information that relates to an identified or identifiable living individual. Different pieces of information, which collected together can lead to the identification of a particular person, also constitute personal data” [Articles 2, 4(1) and(5) and Recitals (14), (15), (26), (27), (29) and (30) of the GDPR; WP 01245/07/EN, WP 136 Opinion 4/2007 on the concept of personal data; Article 29 Working Party Opinion 05/2014 on Anonymisation Techniques].

Personal data ”that has been de-identified, encrypted or pseudonymised but can be used to re-identify a person remains personal data and falls within the scope of the law,, [Articles 2, 4(1) and(5) and Recitals (14), (15), (26), (27), (29) and (30) of the GDPR; WP 01245/07/EN, WP 136 Opinion 4/2007 on the concept of personal data; Article 29 Working Party Opinion 05/2014 on Anonymisation Techniques].

”Personal data that has been rendered anonymous in such a way that the individual is not or no longer identifiable is no longer considered personal data. For data to be truly anonymised, the anonymisation must be irreversible” [Articles 2, 4(1) and(5) and Recitals (14), (15), (26), (27), (29) and (30) of the GDPR; WP 01245/07/EN, WP 136 Opinion 4/2007 on the concept of personal data; Article 29 Working Party Opinion 05/2014 on Anonymisation Techniques].

The law ”protects personal data regardless of the technology used for processing that data – it’s technology neutral and applies to both automated and manual processing, provided the data is organised in accordance with pre-defined criteria (for example alphabetical order). It also doesn’t matter how the data is stored – in an IT system, through video surveillance, or on paper; in all cases, personal data is subject to the protection requirements set out in the GDPR” [Articles 2, 4(1) and(5) and Recitals (14), (15), (26), (27), (29) and (30) of the GDPR; WP 01245/07/EN, WP 136 Opinion 4/2007 on the concept of personal data; Article 29 Working Party Opinion 05/2014 on Anonymisation Techniques].

”Processing covers a wide range of operations performed on personal data, including by manual or automated means. It includes the collection, recording, organisation, structuring, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, restriction, erasure or destruction of personal data. The General Data

Protection Regulation (GDPR) applies to the processing of personal data wholly or partly by automated means as well as to non-automated processing, if it is part of a structured filing system,, [Article 4(2) and(6) of the GDPR].

DPAs are ”independent public authorities that supervise, through investigative and corrective powers, the application of the data protection law. They provide expert advice on data protection issues and handle complaints lodged against violations of the General Data Protection Regulation and the relevant national laws. There is one in each EU Member State” (Regulation (EU) 2016/679).

Generally speaking, ”the main contact point for questions on data protection is the DPA in the EU Member State where your company/organisation is based. However, if your company/organisation processes data in different EU Member States or is part of a group of companies established in different EU Member States, that main contact point may be a DPA in another EU Member State” [Article 4(16), Chapter VI (Articles 51 to 59) and Recitals (117) to (123) of the GDPR; Article 29 Working Party Guidelines on the Lead Supervisory Authority, WP 244; Article 29 Working Party Guidelines for identifying a controller or processor’s lead supervisory authority, and Annex II].

References

Advantech. (2013). *Enhancing Big Data Security*. [online] Available: http://www.advantech.com.tw/nc/newsletter/whitepaper/big_data/big_data.pdf [accessed 18 March 2018].

Agrawal, D., Das, S., and El Abbadi, A. (2011) ”Big data and cloud computing”. In *Proceedings of the 14th International Conference on Extending Database Technology*. New York: ACM Press [online] Available: doi:10.1145/1951365.1951432 [accessed 10 April 2018].

Serrão, C., Neves, D., Kudumakis, P., Barker, T. and Balestri, M. (2003) ”OpenSDRM. An Open and Secure Digital Rights Management Solution”. In *Proceedings of the IADIS International Conference e- Society*, vol. 2, 647-650.

Chen, X. and Shi, S. (2009). ”A literature review of privacy research on social network sites”. In *Multimedia Information Networking and Security. MINES’09. International Conference*, Vol. 1, 93–97.

De Cristofaro, E., Soriente, C., Tsudik, G., and Williams, A. (2012) ”Hummingbird: Privacy at the time of twitter”. In *Security and Privacy (SP), 2012 IEEE Symposium*, 285–299.

Dohi, T., and Uemura, T. (2012). ”An adaptive mode control algorithm of a scalable intrusion tolerant architecture,,. *Journal of Computer and System Sciences*, Vol. 78, 1751–1754.

Feamster, N. (2014). *Software Defined Networking*. [online] Available: from https://www.coursera.org/course/sdn [accessed 22 May 2018].

General Data Protection Regulation (GDRP) (2018) *What does the General Data Protection Regulation (GDPR) govern?* [online] Available: https://ec.europa.eu/info/law/law-topic/data-protection/reform/what-does-general-data-protection-regulation-gdpr-govern_en [accessed 28 May 2018].

GDRP Regulation (2018) *Summary of articles contained in the GDRP* [online] Available: https://www.eugdpr.org/article-summaries.html [accessed 26 May 2018].

Gentry, C. (2009). *A fully homomorphic encryption scheme*. Stanford University, [online] Available: Retrieved from http://cs.au.dk/~stm/local-cache/gentry-thesis.pdf [accessed 21 April 2018].

Gentry, C. (2010). Computing arbitrary functions of encrypted data. *Communications of the ACM*. [online] Available**:** doi:10.1145/1666420.1666444 [accessed 21 April 2018].

Goldwasser, S., Gordon, S. D., Goyal, V., Jain, A., Katz, et. al. (2014). "Multi-input functional encryption". In *Advances in Cryptology. EUROCRYPT 2014,* Berlin: Springer, 578-602.

Gross, R. and Acquisti, A. (2005). "Information revelation and privacy in online social networks". In *Proceedings of the 2005 ACM workshop on Privacy in the electronic society*, 71–80.

Hand, R., Ton, M. and Keller, E. (2013) "Active security". In *Proceedings of the Twelfth ACM Workshop on Hot Topics in Networks* New York: ACM Press, 1-7.

IDC (2012). *Big Data in 2020*. [online] Available: http://www.emc.com/leadership/digital-universe/2012iview/big-data-2020.htm [accessed 25 April 2018].

IEE BigData (2018). *The 12th IEEE International Conference On Big Data Science And Engineering***,** [online] Available: http://www.cloud-conf.net/BigDataSE18/index.html [accessed 27 May 2018].

Juels, A., and Oprea, A. (2013). "New approaches to security and availability for cloud data". *Communications of the ACM*, *56*(2), 64.

Jutla, D.N., Bodorik, P. and Ali, S. (2013) "Engineering Privacy for Big Data Apps with the Unified Modeling Language". In *IEEE International Congress on Big Data 2013*, 38–45.

Marques, J. and Serrão, C. (2013). "Improving Content Privacy on Social Networks Using Open Digital Rights Management Solutions". *Procedia Technology*, *9*, 405–410.

Michael, K. and Miller, K.W. (2013). "Big Data: New Opportunities and New Challenges". *Computer*, 46(6), 22–24.

MIT (2014). *Big Data Privacy Workshop, Advancing the state of the art in Technology and Practice – Workshop summary report.* [online] Available: http://web.mit.edu/bigdata-priv/images/MITBigDataPrivacyWorkshop2014_final05142014.pdf [accessed 17 April 2018].

Moura, J. and Serrão, C. (2015) *Security and Privacy issues of big data, Handbook of Research on Trends and Future Directions in Big Data and Web Intelligence*, Hershey: IGI Global.

Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) [online] Available: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679 [accessed 25 April 2018].

Rodríguez, E., Rodríguez, V., Carreras, A. and Delgado, J. (2009) ”A Digital Rights Management approach to privacy in online social networks”. In *Workshop on Privacy and Protection in Web-based Social Networks (ICAIL’09), vol. 3, Barcelona: IDT Series*.

Serrão, C. (2008). *IDRM - Interoperable Digital Rights Management: Interoperability Mechanisms for Open Rights Management Platforms*. Universitat Politècnica de Catalunya. Retrieved from http://repositorio-iul.iscte.pt/handle/10071/1156

Serrão, C., Rodriguez, E. and Delgado, J. (2011). ”Approaching the rights management interoperability problem using intelligent brokerage mechanisms„. *Computer Communications*, *34*(2), 129–139.

Tankard, C. (2012). ”Big data security”. *Network Security*, 7, 5–8.