TIZIANA CROCE*

*Big data? A question of balance between privacy, security and information power***

Introduction

The investigation and research techniques developed by the Network imply, in particular, the exploitation of the multiple information potentials arising from big data, or the huge amount of data generated by the most modern IT systems for communication, transaction and localization[1]. The terms retargeting or remarketing, data mining, web crawling and data strategies indicate activities aimed at transmitting targeted promotional messages that are supposedly more effective in maximizing the usefulness of commercials while minimizing the effort. The growing use of big data is a phenomenon that primarily affects the fundamental right to the protection of personal data, because the techniques underlying the proliferation of data are aimed at predicting the future behaviour of individuals and the study of their habits, preferences and relationships through the analysis of the multiple digital traces generated by localization, transaction and digital interaction systems. In gen-

* *Professore Aggregato di Informatica giuridica presso l'Università degli Studi di Camerino.*

** *Contributo sottoposto positivamente al referaggio secondo le regole del double blind peer-review.*

[1] An opinion of the European Data Protection Supervisor defines big data as "… the practice of combining huge volumes of diversely sourced information and analysing them, using more sophisticated algorithms to inform decisions", European Data Protection Supervisor, *Meeting the Challenges of Big Data* (Opinion 7/2015), 17 November 2015, p. 7.
http://secure.edps.europa.eu/EDPSWEB/webdav/site/mySite/shared/Documents/Consultation/Opinions/2015/15-11-19_Big_Data_EN.pdf.

eral terms, the activities related to the use of big data and commercial proliferation can highlight clear contrast profiles related to the violation of the principle of finality, the absence of consent of the parties concerned, lacking or inadequate information about the processing of individuals' personal data. Added to this are doubts concerning the conditions of legitimacy of real private databases, the levels of security, understood as irreversibility of anonymisation techniques, and the compliance with the erasure obligations. The idea that the protection of personal data is just a formal problem is now old and outdated, and it is misleading to assume that there is no longer any possibility or usefulness in claiming to protect personal data from undue processing. Technologies are clearly limited and there are many inevitable risks related to digital technology, albeit not unmanageable. Legal rules are also limited in tackling and resolving the concrete problems of a world in which the digital dimension is an integral part of everyday life in every sector[2]. It is worth highlighting that nowadays the concentration of information is in the hands of a few actors in the network. Although this is no news, it is true that the size of big data and their management further increases the centralization of most data flows in the hands of a few operators. The analysis of such data aggregates may have a significant strategic, socio-political and patrimonial value, such that the dominion of individuals over information about them now makes room for the dominion of information holders over information, with all the consequences resulting from it. It is therefore clearly necessary, for the completeness of the contribution, to mention the innovations introduced by the new European regulatory framework to protect the processing of personal data and the provisions that the general regulation 2016/679 provides for the proliferation of data.

1. *General Data Protection Regulation*

The general regulation on the protection of personal data 2016/679, which will become fully applicable in May 2018, and the proposal for a Directive on e-privacy, which will replace Directive 2002/58/EC[3], reveal even more the need for a Community acquis

[2] F. PIZZETTI, *Big Data e Privacy by Design*, Turin, 2017, p. 18 and ff.
[3] Commission proposal, COM (2017) 10 final of 10 January 2017, *Respect for private life and the protection of personal data in electronic communications*. The European e-privacy regulation will form a special law within the Gen-

that adapts data protection mechanisms to the digital environment, which of course is no longer the elective framework of reference for the period in which Directive 1995/46 was adopted[4]. Evidently, the development of electronic communications determines the ever-increasing need for legal certainty as people increasingly use text messages instead of traditional telephony, and voice-over-IP services, web-based email services and messaging services instead of email services, thus creating the conditions to extract from the information further information related to behaviours, habits and movements, used to define personal aspects related to social relationships, interests, tastes of the people involved in electronic communications. The new rules should also cover machine-to-machine communications in the context of the Internet of Things regardless of the type of network service or communication used[5].

The awareness that the evolution of the digital society is unstoppable has forced the European Union to reconcile a high level of protection of the right to the protection of personal data with a society that increasingly lives on data and builds its future with them. Some important innovations including Web 2.0, which allows everyone to use the network to exchange information originating the explosion of social networks, and the evolution of the cloud technology, connected to increasingly powerful data transmission networks and

---

eral Data Protection Regulation, and it will regulate and integrate data relating to electronic communications of the general regulation having the character of personal data.

[4] In *Recommendation 1/99 on invisible and automatic processing of personal data on the Internet performed by hardware and software* adopted on 23 February 2009, the European Group of Supervisors formulated the criterion whereby software and hardware companies had to configure devices and technical tools developed so as to make them, before any use thereof, compliant with the data protection rules that have their source in the European directives.

[5] In this regard, the Data Protection Authority, *Opinion of 14 October 2016 on the revision of the e-privacy directive*, OJEU C 378 14 October 2016, "The scope of the new legal framework must be extended. This is to take account of technological and societal changes and to ensure that individuals be afforded the same level of protection for all functionally equivalent services, irrespective whether they are provided, for example, by traditional telephone companies, by Voice over IP services or via mobile phone messaging apps … protect not only 'functionally equivalent' services, but also those services that offer new opportunities for communication … ensure that the confidentiality of users' communications will be protected on all publicly accessible networks, including Wi-Fi services in hotels, coffee shops, shops, airports and networks offered by hospitals to patients, universities to students, and hotspots created by public administrations ... no communications should be subject to tracking and monitoring without freely given consent, whether by cookies, device-fingerprinting, or other technological means. Users must also have user-friendly and effective mechanisms to provide and revoke their consent within the browser (or other software or operating system) … the current consent requirement for traffic and location data must also be maintained and strengthened ... allow users to use end-to-end encryption to protect their electronic communications … Finally, the new rules on e-privacy should protect against unsolicited communications and should be updated and strengthened, requiring prior consent of recipients for all types of unsolicited electronic communications, independent of the means.

the development of mobile technologies, determine an unstoppable growth in data production, the possibility to process and cross-check them, a proliferation of information unimaginable to most people[6].

To regulate the new reality, the legislative instrument used by the European Union is the Regulation as a binding act for all member states, a normative act that allows for a broader, more structured and uniform regulatory framework throughout the territory of the Union. The regulation recognizes the fundamental right of the person to the protection of personal data as a European public interest; in other words, the protection of personal data no longer responds only to the protection of the person to whom the data refer, but extends to the protection of the society to which the person belongs, and the protection of personal data was already recognized as a fundamental right of the Union in the Treaty of Lisbon and the Charter of Nice[7]. The regulation ensures protection is both effective and not such as to hinder the evolution of a society that cannot renounce digital technology and the prospects it offers. The playing field is characterized by the deterritorialization, denationalization and dematerialization processes which are perhaps the most immediate and, paradoxically, most tangible result of the digital revolution[8].

The new European provisions underpin the strengthening of the principles referred to, through the introduction of new rules and institutions aimed precisely at ensuring a better ability to govern the phenomenon. A new principle provides for the application of European law also to the processing of personal data not carried out in the EU if related to the supply of goods and services to EU citizens that would entail their monitoring, in con-

[6] F. PIZZETTI, *op.cit,* p. 1, "…In this new world, the enormous growth in data production and the possibility of acquiring, storing, processing and cross-checking them has triggered the phenomenon of big data, data also increased by technologies related to artificial intelligence and the Internet of things".

[7] Pursuant to art. 7: "Every person has the right to respect for his or her private and family life, home and communications". Pursuant to art. 8: "1. Everyone has the right to the protection of personal data concerning him or her. 2. Such data must be processed fairly for specified purposes and on the basis of the consent of the person concerned or some other legitimate basis laid down by law. Everyone has the right of access to data which has been collected concerning him or her, and the right to have it rectified. 3. Compliance with these rules shall be subject to control by an independent authority". The articles mentioned were, in the rulings of the Court of Justice, their reasoning, highlighting in particular the role played by the provisions of the Charter in the argumentative process and in the final outcome of the decisions, in relation to the impact of the technological factor on the level of protection of fundamental rights and the possible limitations that the latter ones can undergo through the new methods of monitoring and indexing provided by the development of digital technology.

[8] Cf. O. POLLICINO, *Interpretazione o manipolazione? La Corte di giustizia definisce un nuovo diritto alla privacy digitale*, in *Federalismi.it*, 24 November 2014.

trast to the European Directive 1995/46 which provided that the applicable legislation was the one relating to the place of establishment of the undertaking[9]. A new right is provided for by art. 17 or the right to be forgotten[10], that is the right for the data subject to decide that personal data concerning him or her that are no longer necessary for the purposes for which they are collected are erased and not processed, in addition to the right of erasure in the case of withdrawal of consent or when the data subject has opposed the processing of personal data concerning him or her or when the processing of his or her personal data does not comply with the regulation. The right of data erasure is limited, with a view to balancing interests, with respect to the existence of reasons of public interest in the health sector, in the need to safeguard the purpose of processing consisting in archiving data for public interest or for the purposes of scientific, historical and statistical research, and to protect the assessment and proper defence in court.

Indeed, the Court of Justice[11], pending the new European regulatory provisions, has

---

[9] V. Z. ZENCOVICH, *Intorno alla decisione Schrems: La sovranità digitale e il governo internazionale delle reti di telecomunicazione* in *Diritto dell'informazione e dell'informatica*, 2015, p. 683 and ff. ... "The decision of the Court of Justice in the Schrems case is a step further towards the affirmation of a digital sovereignty of the European Union, with sovereignty understood as the power to control, de jure and de facto, a certain space, the activities taking place in there, how the space is organized, to administer police, judicial and safety powers in this space … and "the Court of Justice in the Google Spain case has stated that Google must be considered established in the European Union and therefore subject to European law, thus affirming the sovereignty over economic entities operating within the European space, albeit through telecommunication networks that allow for the use of the Internet" … "… by making clear that the transfer of personal data of European citizens to the United States is not lawful, it essentially states that the processing of personal data is governed by EU law and not by the law of another State … and consequently the abolition of the Safe Harbor agreement concluded by the European Commission with the United States".

[10] T.E. FROSINI, *Il Diritto all'oblio e la libertà informatica* in *Danno e responsabilità,* 2012, p. 720; F. PIZZETTI, *Il caso del diritto all'oblio,* Torino, 2013, p.23, G. FINOCCHIARO, *Il diritto all'oblio nel diritto alla personalità* in *Diritto dell'informazione e dell'informatica,* 4/5, 2014.

[11] The Court of Justice of the European Communities, on the occasion of the Judgment no. 131/12 of 13 May 2014, so-called Google Spain, had established that, in relation to the rights deriving from articles 7 and 8 of the Charter of fundamental rights of the European Union, the data subject can ask that a given piece of information, published on the web, be no longer made available to the general public, prevailing both on the economic interest of the search engine operator and on the public's interest in accessing this information when searching for the name of the person concerned. It also stated that such prevalence fails if it turns out that interference with the fundamental rights of the data subject is justified by the pre-eminent interest of web users in having access, by virtue of the aforementioned inclusion, to the information in question. In the aftermath of the aforementioned judgment by the EU Court of Justice, on the right to be forgotten which recognized for the first time the right to be "de-indexed" by the search engine, thus actually requiring Google to meet users' requests, the EU Privacy Supervising Authorities took steps to work out common criteria for handling appeals and complaints made by users whose request was rejected. This was the start of a process to harmonize the procedural and substantial criteria already in force in every single system concerned, to handle cases in which the search engine rejects the deindexing request. At the same time, the abovementioned authorities reaffirmed the duty for search engines to fulfil the obligations arising from the aforementioned

tried to balance the right to inform and to be informed on the web and other personality rights: reputation, confidentiality, data protection, identity, and right to be forgotten, thus filling the blank spaces or modifying the previously achieved balances, due to the progressive technological change, defining a right to digital identity understood as the right of the individual to obtain rectification, contextualization, progressive updating over time and de-indexing and erasure of personal data from the web, in order to ensure a correct and current representation of their identity and to guarantee the right to be forgotten[12]. In article 17, the European legislator understood the right to be forgotten as a mere cessation of the processing[13], as it did not include all the features that the notion of right to be forgotten has taken on in recent years with protection of personal identity and protection of personal data of the individual, and although they were provided for in the recitals of the regula-

judgment by the European Court. O. POLLICINO, *Interpretazione o manipolazione? La Corte di giustizia definisce un nuovo diritto alla privacy digitale* in *Federalismi.it*, November 2014 ... "Whenever Member States prove to be unwilling to progress in the community acquis by law, the Court of Justice follows the line of judge-made law and accelerates on the basis of case-law … it had been decided for years that an act of general application, immediately binding and mandatory was essential, which would lead to greater standardization of the regulatory and mandatory data ... which would adapt the mechanisms of data protection to the digital context which, obviously, was not the reference elective framework when Directive 46/95 was adopted ... even so, the general regulation has not surprisingly accelerated as a result of the intervention of the Court of Justice in the Google Spain judgment … A right of privacy based on the two pillars that are the rights to respect for private life and to the processing of personal data provided for by articles 7 and 8 respectively of the Charter of Fundamental Rights of the European Union ... The judges of Luxembourg have imposed the obligation on search engines to remove, under certain circumstances and at the express request of the applicant, links to Internet pages containing information that could be prejudicial to the so-called right to be forgotten of the individual whose personal, and often sensitive data remain for a significant period of time on the web ...".

[12] Thus G. E. VIGEVANI, *Identità, oblio, informazione e memoria in viaggio da Strasburgo a Lussemburgo, passando per Milano*, in *Federalismi.it*, 19 September 2014.

[13] "The data subject has the right to obtain the erasure of personal data concerning him or her from the data controller without undue delay and the data controller is required to erase such personal data without undue delay", subject to certain conditions: a) personal data are no longer necessary in relation to the purposes for which they are collected or otherwise processed; b) the data subject withdraws his or her consent on which the processing is based, and there is no other legal basis for the processing; c) the data subject objects to the processing due to his or her particular situation (art. 21, par. 1) and there is no other legitimate overriding reason to proceed with the processing, or objects in relation to personal data that are processed for direct marketing purposes (art. 21, section 2); d) personal data have been processed unlawfully; e) personal data must be erased to fulfill a legal obligation under Union law or the law of the Member State to which the data controller is subject; f) personal data have been collected in relation to the offer of services of the information company in accordance with the provisions of article 8 about consent given by minors. The second section of the provision provides that the data controller who, in the presence of the conditions described, is required to erase personal data, "taking into account the available technology and implementation costs shall take reasonable measures, including technical ones, to inform data controllers who are processing personal data of the data subject's request to erase any link, copy or reproduction of his or her personal data".

tion[14]. In fact, scholars have repeatedly asserted a sharp distinction between erasure and oblivion, understanding data erasure as the operation that excludes any further conservation of data, while oblivion is aimed at both erasing and blocking data[15].

The Regulation establishes the right to data portability[16] by virtue of which the data subject has the right to receive, in a structured format that is commonly used and readable by automatic device, the personal data that concern him or her provided to a data controller, and the right to transmit such data to another data controller, without any impediments, if the data subject has given his or her consent to the processing or if this is necessary for the execution of a contract; for example, it will be possible to change the electronic mail provider without losing contacts and saved messages[17]. The portability right does not apply to the processing necessary for the performance of a task carried out in the public interest or in connection with the exercise of official authority appointed to the data controller.

The objective of the Community legislator is to further strengthen the data subject's control over their personal data, but above all to avoid any commercial abuse or illegal manipulation of personal data and therefore not to leave the way open to the market. It is important that the consumer is given the right to change service and take away their own data, and that the upload of personal contents does not result in a mere commercial advantage for personal data processing companies.

---

[14] RGPD 679/ 2016 recital 65 "… In particular, a data subject should have the right to have his or her personal data erased and no longer processed where the personal data are no longer necessary in relation to the purposes for which they are collected or otherwise processed, where a data subject has withdrawn his or her consent or objects to the processing of personal data concerning him or her, or where the processing of his or her personal data does not otherwise comply with this Regulation".

[15] Cf. G. FINOCCHIARO, *Il diritto all'oblio nel quadro dei diritti della personalità* in *Diritto dell'informazione e dell'informatica*, 4/5 2014, p. 643.

[16] Data are no longer hostage to the online service provider and those wishing to close their accounts or migrate to another provider have the right to bring along their story and restart with a new provider if they had stopped receiving services from the old provider. The European legislator supports the interoperability between different systems of different providers and a process of standardization of personal data formats so that, on the one hand, the reference market becomes increasingly competitive and, on the other, the prevention of worrying lock-in phenomena is consolidated.

[17] Thus the Personal Data Protection Supervisor, *Cloud computing: indicazioni per l'utilizzo consapevole dei servizi,* 2 011, doc. web. 1819993: the Authority had denounced that the adoption of own technologies by the service provider could make the transition of data and documents from one cloud system to another or the exchange of information with subjects using cloud services of different providers complex for the user, jeopardizing data portability or interoperability. Therefore, the Authority asked to privilege services based on open formats and standards capable of simplifying the transitions from one cloud system to another.

2. *General Personal Data Protection Regulation*: *continuation*

New responsibilities rest with the data controller and the person in charge of personal data processing. In particular, articles 24[18], 25[19] and 28[20] of the Regulation provide for data protection right from the design of the processing with standard protection modes, namely privacy by design[21] and privacy by default[22], and the adoption of suitable organiza-

[18] Article 24 Responsibility of the controller: 1. Taking into account the nature, scope, context and purposes of processing as well as the risks of varying likelihood and severity for the rights and freedoms of natural persons, the controller shall implement appropriate technical and organisational measures to ensure and to be able to demonstrate that processing is performed in accordance with this Regulation. Those measures shall be reviewed and updated where necessary. 2. Where proportionate in relation to processing activities, the measures referred to in paragraph 1 shall include the implementation of appropriate data protection policies by the controller.

[19] Article 25 Data protection by design and by default: 1. Taking into account the state of the art, the cost of implementation and the nature, scope, context and purposes of processing as well as the risks of varying likelihood and severity for rights and freedoms of natural persons posed by the processing, the controller shall, both at the time of the determination of the means for processing and at the time of the processing itself, implement appropriate technical and organisational measures, such as pseudonymisation, which are designed to implement data-protection principles, such as data minimisation, in an effective manner and to integrate the necessary safeguards into the processing in order to meet the requirements of this Regulation and protect the rights of data subjects. 2. The controller shall implement appropriate technical and organisational measures for ensuring that, by default, only personal data which are necessary for each specific purpose of the processing are processed. That obligation applies to the amount of personal data collected, the extent of their processing, the period of their storage and their accessibility. In particular, such measures shall ensure that by default personal data are not made accessible without the individual's intervention to an indefinite number of natural persons.

[20] Article 28 Processor: 1. Where processing is to be carried out on behalf of a controller, the controller shall use only processors providing sufficient guarantees to implement appropriate technical and organisational measures in such a manner that processing will meet the requirements of this Regulation and ensure the protection of the rights of the data subject. 2. The processor shall not engage another processor without prior specific or general written authorisation of the controller. In the case of general written authorisation, the processor shall inform the controller of any intended changes concerning the addition or replacement of other processors, thereby giving the controller the opportunity to object to such changes. 3. Processing by a processor shall be governed by a contract or other legal act under Union or Member State law, that is binding on the processor with regard to the controller and that sets out the subject-matter and duration of the processing, the nature and purpose of the processing, the type of personal data and categories of data subjects and the obligations and rights of the controller. That contract or other legal act shall stipulate, in particular, that the processor.

[21] The first group of measures concerns the preparation and planning of personal data processing activities, in which the configuration of tools and methods must be pre-arranged by the data controller, and be bound to the performance of operations compliant with data protection principles, processing requirements and the protection of data subjects. In this regard, one of the adoptable measures is pseudonymisation: the result obtained when depriving data of their traceability to an identified or identifiable specification to which they cannot be irreversibly attributed once the relevant processing operation has been completed, unless additional information is used, stored separately and be subjected to appropriate guarantee measures.

[22] This group consists of measures concerning appropriate technical and organisational solutions implemented by the controller to ensure that, based on default settings and in relation to each specific purpose of the processing, the processing itself is limited to necessary data. The provision applies to the quantitative and qualitative aspects of collection, the duration of storage and their accessibility, so that the default configuration of the systems can ensure the legal compliance of the processing and preclude the dissemination of data.

tional technical measures for ensuring that the processing meets the requirements under the regulation and guarantee adequate protection of the data subject. One of the most significant innovations is that the controller must identify the means to be used and design the methods of processing so that the guarantees required are integrated right from the beginning of the processing. The controller and the person in charge of processing are no longer responsible for the mere compliance of processing with the personal data protection regulation, but are required to prove that they have taken all appropriate technical and organizational measures, which are consistent with the structural need for a systematic assessment by the data controller and the person in charge of processing with reference to both the protection of the rights of the data subject and data security and, in this context, it is essential to carry out a risk analysis that is strictly linked to the profiles concerning the legal responsibility for the processing and to the operations of other institutions introduced by the regulation, such as impact assessment[23], a flexible and dynamic concept of responsibility that must be parameterized from time to time on the methods of processing and the risks involved. Constant monitoring of the measures taken is needed to make sure that they are immediately appropriate in case technological innovations or other processing-related aspects require so. This important innovation contained in the regulation is defined "data protection impact assessment"[24], that is a risk analysis intended to weigh, ex ante, the impact that a certain technical solution will have on the protection of processed data, and to identify, in relation to the various processing phases, the related risks and the measures suitable to contain or neutralize them. Sometimes privacy by design is identified with data protection impact assessment, but that is not correct since first of all, impact assessment comes in a preliminary stage of the service development, when service design is not out-

[23] Cf. S. CALZOLAIO, *Privacy by design,* in *Federalismi.it*, 2017, " … the regulation is intended to qualify the level of risk, distinguishing between generic risk and high risk … The risk assessment parameter takes into account the likelihood and seriousness of an infringement of the rights and freedoms of data subjects due to or within the scope of the processing, and is not left to the mere sensitiveness of the data controller, but is objectified by the dynamics of evaluation of approved codes of conduct and/or approved certifications and/or the guidelines provided by the European committee for data protection and/or guidelines provided by a protection officer".

[24] A. MANTELERO, *Riforma della direttiva comunitaria sulla data protection e privacy impact assessment, verso una maggiore responsabilità dell'autore del trattamento*, in *Diritto dell'informazione e dell'informatica*, 2012, 145 and ff. … "The procedures aimed at defining the privacy impact assessment have been customary for several years in various countries, especially in relation to the activity of public entities … in order to encourage producers to develop technologies that are privacy-compliant from the very beginning".

lined definitively. And it is exactly in this stage that an assessment of the specific envisaged solution implying the processing of data must be carried out. In case of compliance with the regulatory provisions for the protection of personal data, the service will be developed incorporating specifications and solutions of privacy by design. The preliminary assessment of the critical issues related to the acquisition and processing of data is the peculiarity of the privacy impact assessment, which differs from other risk analysis processes that do not intervene ex ante, but ex post, as in the case of the data security program already provided for by Legislative Decree 196/2003, privacy code, but deleted by Legislative Decree D.L. 5/2012 converted by Law 35/2012[25].

In terms of interpretation, the regulation draws a line that allows introducing a type of processing that appears to fully integrate the extremes of relevant and persistent riskiness into the new European discipline. This is so-called proliferation defined in article 4 as any form of automated processing of personal data consisting in the use of such personal data to evaluate certain aspects relating to a natural person. Proliferation is a new form of knowledge due to the correlation of data contained in one or more databases aimed at defining the profile of a person or a group. This activity may include the creation of big data, against which the regulation, article 21, provides for the data subject's right of opposition and right of explanation. This entails acknowledging the data subject's right to know the mathematical and statistical procedures used by the data controller for the proliferation of data. Pseudonymisation is the remedy that the general regulation identifies in order to systematically address the risks or in any case minimize the impact of automated processing on the personal sphere. It is a technical measure, an operation whereby personal data cannot be referred to a data subject without the use of additional information.

The European legislator, with a view to greater protection of the data subject concerned, has provided a centre of imputation of responsibility called to respond for any offenses which, in addition to providing the already known figures of the owner and the person responsible for processing personal data, has introduced the figure of the data protec-

[25] A. MANTELERO, *op. cit.*, p. 147.

tion officer[26], that is the person responsible for personal data protection, who is required to ensure internal control of the company, in order to minimize the risks of violation of the regulation and to allow the Authorities responsible for monitoring to refer to a single figure that must be the terminal for the companies themselves in case of enquiries or sanctions. The new figure established by the regulation has a direct impact both on the verification of the processing carried out by the controller and the processor in relation to the technical and legal compliance with data protection regulations, and on the relationships between these and the control authorities. The data protection officer is required to monitor all personal data processing carried out by the organisation to which he or she belongs and the regulation establishes for this function that the data protection officer is given absolute independence from the same facility that has designated him or her and for which he or she works, and that the availability of human, instrumental, technical and organizational resources is also ensured[27].

3. *Big data*?

The expression of Big Data refers to large volumes of data which, thanks to the use new techniques and advanced technologies, allow for the collection, storage, distribution, management and constant production of information[28]. There is intense debate about the origin of the expression Big Data and how to define it properly. The two words appeared together occasionally for decades. A research project published in 2001 by Doug Laney at the American consulting firm Gartner Group highlighted three characteristics that define the phenomenon of big data and named them the three Vs, volume, velocity and variety.

[26] The data protection officer had already been introduced as mandatory in certain European systems, including Germany, Austria and Czech Republic; instead, in other systems like France, the appointment of this figure was optional.

[27] F. PIZZETTI, *Privacy e il diritto europeo alla protezione dei dati personali*, Turin 2016, p. 301 "… The data protection officer has a threefold nature: a) fully independent subject that must control every data processing within the scope of his or her competence without encountering any obstacles; b) point of contact between data subject and the facility to which he or she belongs for processing; c) single point of contact between the supervisory authorities and the facility to which he or she belongs for everything concerning personal data processing carried out by the facility".

[28] Motion for the *European Parliament resolution on fundamental rights implications of big data*: privacy, data protection, non-discrimination, security and law-enforcement, recital A: big data refers to the collection, analysis and the recurring accumulation of large amounts of data, including personal data, from a variety of sources, which are subject to automatic processing by computer algorithms and advanced data-processing techniques using both stored and streamed data in order to generate certain correlations, trends and patterns.

This characterization was useful for that time, but imperfect at the same time[29]. The first characteristic defined Volume indicates that the huge number of devices connected in the network provides a large amount of data that had never been available in the entire history of mankind. However, this feature does not only refer to the quantity of data, but also to the ease and cheapness of collection. The higher data processing power allows any phenomenon to be analysed globally, and therefore data to be reused for any purpose, unlike what happened in the past, when data samples were chosen depending on the purpose to be achieved. The second characteristic, Velocity, indicates that data are produced continuously in a dynamic and not static manner; the data provided in any context determine a flow released at a certain velocity and the processing must be updated continuously based on the new data[30]. The third characteristic, Variety, indicates both the variety of formats and sources. It refers to data published on Facebook, videos on YouTube or data generated by sensors, that is the technology of the Internet of things[31] or Twitter tweets. Associated with the three Vs mentioned above is the fourth V, which indicates the Veracity of data. In fact, the heterogeneity of the sources makes it more complex to verify the correctness of data. The essential characteristic for data to constitute big data is that the bigger the number of attributes with which the data is described and the number of phenomena that it is

[29] Wikipedia: … In 2011, Teradata stated that "Big data exceeds the reach of commonly used hardware environments and software tools to capture, manage and process it within a tolerable elapsed time for its user population". A further definition of big data was given by the McKinsey Global Institute: "Big data refers to data sets whose size is beyond the ability of typical database software tools to capture, store, manage and analyse".

[30] I cite a case that is by now for books: in 2009 a new influenza virus was discovered. A combination of viruses that cause bird flu and swine flu, the new disease called H1N1 has spread rapidly. Within a few weeks, health agencies around the world began to think that a terrible pandemic was going on … they thought of a tragedy comparable to the Spanish flu, which spread in the last century and killed twenty million people … there was no vaccine … hoping to limit the spread, it was first necessary to know the exact location of the outbreaks … Weeks went by … health authorities had no data on the actual facts … sick people would not call the doctor immediately … Google then identified a prediction technique based on all the requests for information sent to the search engine concerning the flu, and on the possibility of trying out a large number of different models to describe the evolution of the phenomenon, then selecting the best one; the model proved capable of pinpointing the flu peak.

[31] The Working Group, pursuant to article 29, in *Statement on Statement of the WP29 on the impact of development of big data on the protection of individuals with regard to the processing of their personal data in the* EU (WP221) 16 September 2014, identifies three types of Internet of things devices: wearables, quantified self devices, namely devices that allow having data related to activities, hours and daily habits of the person, and home automation devices, such as fridges and lamps. This distinction should be used by IoT producers to develop a proactive and non-reactive approach that can anticipate possible invasions of a person's privacy, as early as in the product design phase, i.e. what is defined in the regulation as privacy by design.

potentially able to explain, the bigger the data itself. Each attribute added to the description of a data immediately becomes a new dimension to be explored for connections between that data and other data, between the phenomenon represented by that data and other phenomena represented by other data. The greater the number of descriptors, the more likely the connection of a data with other data[32], each made in turn bigger by a richer set of attributes. This way, connections can be found between phenomena which were previously hidden[33]. One of the ways to make data into big data is the semantic classification of contents by keywords or indicators. The richer the description of the data, the more suitable it will be for a connection with other data. The other way is crowdsourcing[34], i.e. the set of information about data that is provided by the users of such data, or the extraction of the keywords entered by the users, or the analysis of suggestions and feedback provided on automatic translations. Hence the consideration that the connection surfacing scheme can be applied to any data in order to create a net that is incomparably thicker than the current structure of the web, namely on connections between data chosen unilaterally by those who upload them[35]. The Internet of things[36] can generate big data – in fact, each data refers to one thing and each thing bears a meaning: a particular condition of the thing, its history,

[32] This connection surfacing scheme can be applied to any data, even the smallest and apparently meaningless ones, in order to create a net that is incomparably thicker than the current structure of the Web, that is, on the connections between data chosen unilaterally by those who upload them.

[33] To extract descriptors from a data there are many techniques that can be used in combination with each other to enrich the description of a data. Some can be automatic, others need human intervention. The passive ones include machine learning, i.e. the automatic identification of the categories to which the data belongs; another one is hashing, or the creation of univocal digital prints of a data, which can be obtained independently of the meaning of the data, by applying cryptographic techniques. For example, it is also used for non-text data such as audio tracks, videos and images.

[34] Wikipedia: Crowdsourcing is the collective development of a project by several people outside the entity that has conceived the project. The people who collaborate usually do it voluntarily, responding to an invitation to collaborate. This model of project implementation is generally made possible by the Internet and does not necessarily concern the writing of codes in programming languages, but the variety of projects may be different. Just think of Wikipedia itself, which is written by its readers.

[35] G. D'ACQUISTO, M. NALDI, *Big data e Privacy by design*, op. cit. p. 17, "… It is to be expected that the full deployment of this new knowledge generation scheme will be the result of an evolution rather than a revolution … We already see the first examples of it … From the automatic completion of search queries to the disambiguation of the results of a search, from searching by images to the possibility of finding the title of a song directly".

[36] The expression Internet of Things (IOT) indicates a network between physical objects connected through electronic systems, software and sensors, which can be accessed so that the system can be started, corrected and oriented remotely. Basically, it is a set of devices based on the infrastructure of the International Union's Global standards Initiative. The system creates an interconnection between physical subjects and computer-based systems aimed at increasing the efficiency, quality and cheapness of the activity carried out by interconnected things.

its possible use, the set of experiences that have led to its creation, the set of other things that make up one thing, with their conditions, histories, uses and experiences, in a play of references that can be potentially repeated endlessly. Things equipped with sensors could generate accurate state descriptors about place, time, operating conditions, and surrounding environment. Therefore, they would be open to a standardisation of formats and meanings.

4. *Big data and privacy*

The peculiarity in the big data system is the find engine, which, unlike the search engine, has a decisive role in identifying connections between data and things, and is accountable in case data and things refer to people[37]. As a matter of fact, when responding to a query, the search engine displays results based on the number of visits to the websites, and therefore it is users who affect the outcome of the response to the request. On the other hand, with the find engine, the outcome of the response will be conditioned not only by the positioning of a result but also by the final relevance for each of the contents associated to a result. The richer the description of data and things form a semantic point of view, the more precise the search. If things and data refer to people in the big data system, greater responsibility will be required on the part of the mediator, whose activity also implies leaving to the people to whom data and things refer the right to control such information, not only to comply with the provisions on the protection of personal data but also to prevent people from disseminating false data on the web to compensate for the information asymmetry[38].

The use of big data clearly highlights contrasts with the regulations on the protection of personal data in relation to the violation of the principle of purpose, the absence of consent of the parties concerned, lacking or inadequate information about the processing of personal data of the subjects to whom the data refer.

---

[37] The search engine *searches*, while we then *find* what we need. In order for the search engine to carry out this last step, the representation of data must be adequate: bigger and not only more data are needed. The bigger the number of phenomena a data can explain, the bigger the data itself. Two phenomena can be related to each other because the data that represent them show commonalities, expressed by the presence in both of the same descriptors, which will connect both.

[38] Cf. G. D'ACQUISTO, M. NALDI, *Big data e Privacy by design*, op. cit. p. 28 and ff. "… it is in the interest of the find engine to promote inclusion and avoid fakes that would distort reality, causing us to lose this huge opportunity for knowledge …".

It should be noted that the notion of personal data provided for in the regulation is rather wide. In addition to data that directly or indirectly identify the person, it includes online identifications provided by digital devices, such as IP addresses, temporary markers, so-called cookies, or radiofrequency identification tags (RFID).

Big data processing may involve personal and non-personal data, data relating to one or more persons collected on the basis of the informed consent of each of them, and data that may concern those same persons but which are found and collected freely; data relating to identified or identifiable persons, obtained from information on the network or in any case public information; data identified and collected by exercising the right of access and the rules of administrative transparency. We can clearly talk about personal data processing even when the data that is being processed can be connected to data relating to the same subject present in several databases. The responsibility of the data controller, who determines the purposes and tools used for processing personal data, is also clear and further confirmed by the general data protection regulation. The data controller has specific obligations related to data quality: data must be processed correctly, which means they must be exact, complete relevant and not going beyond the purposes for which they are processed. Above all, the must never be processed without the subject to whom they refer being really aware of it. The data controller must respect the principle of purpose: in other words, personal data may be processed for specific, explicit and defined purposes before data processing takes place. This means that the pursuit of purposes other than those provided for originally determines the unlawfulness of the processing. Furthermore, the principle of necessity establishes that data being processed are only those necessary for the purpose pursued and also used for the time necessary to pursue this purpose, after which processing must cease. Another requirement for the lawfulness of processing is the consent of the subject to whom the data refer.

Big data systems must necessarily comply with the principles just mentioned: the consent must be *given freely* in order to be valid and the data subject must have the possibility to accept or refuse the processing of his or her personal data; it must be *informed*, so the person must have the necessary information on the processing in order to form a precise judgment; it must be *specific*, or requested only for the purposes for which data are pro-

cessed; it must be *unequivocal*, which means a positive action is required, which unambiguously indicates the will of the data subject before processing begins. In addition to these principles, the data controller has transparency obligations: on his or her identity, on the purposes of processing, on the prediction of possible and further recipients of data, on the existence of rights to access data and on the right to oppose processing. The availability and clarity of such information are a prerequisite for the validity of the consent; moreover, the data controller is responsible for data security, which is ensured by implementing adequate technical and organizational measures for monitoring and limiting access to data. As mentioned above, the general regulation provides for new rights for the data subject: the right to data portability aimed at strengthening the right of access for data subjects; the right to be forgotten which provides for the obligation for the data controller who has communicated data to third parties to take all necessary steps to inform them of the erasure request by the data subject; the obligation for data controllers to implement a data protection impact assessment, which must contain a description of processing, an assessment of the risks to privacy and the measures envisaged to prevent such risks. In relation to the complexity of big data, new co-controllership mechanisms have been introduced: in case two or more controllers jointly determine the purposes and tools of processing, only one must be the reference person for the exercise of rights, while the respective responsibilities in terms of privacy can be established in a transparent manner by the various controllers.

5. *Big data and privacy: continuation*

The regulation reaffirms the principle of purpose as a fundamental element of the legitimacy of personal data processing and, along with the informed consent, a prerequisite for the implementation of big data. Likewise, it confirms that its principles must not be applied if the information is made anonymous[39]. In this sense, it is therefore possible to process data with big data techniques for a plurality of purposes when the data are anonymous or in any case made as such; on the other hand, personal data, even pseudoanonymised, cannot be processed with big data techniques if the processing pursues purposes other than

[39] Recital 26: "… the principles of data protection should therefore not apply to anonymous information, namely information which does not relate to an identified or identifiable natural person or to personal data rendered anonymous in such a manner that the data subject is not or no longer identifiable".

those for which data subjects have given their informed consent, and are not among those purposes for which there are specific legislative provisions that authorise processing even without consent[40].

The problem is that data processing through big data techniques hardly ever remains confined to a single purpose, or it can only concern the purposes that can already be envisaged when the data is collected and with respect to which it is therefore possible to request informed consent. Big data processing often concerns personal and non-personal data; data relating to one or more specific persons collected on the basis of the informed consent by each of them, and data that may also concern those same persons but are found on the network; data relating to identified or identifiable persons; data identified and collected by exercising the right of access and the rules of administrative transparency.

In most cases, the use of big data techniques is aimed at acquiring new information about the behaviour of natural persons, for the purpose of predicting their behaviour or identifying preferences, relationships and habits. In these cases it is not possible, in principle, to process data only in an anonymised form; it is surely possible to carry out pseudo-anonymised processing, but in any case the creator of the big data system has interest in using data for identifying the persons whose behaviour is to be studied or influenced. In this regard, the regulation is very clear: the data must be processed anonymously and it is considered as such if the identification of the data subject is prevented and no longer allowed. In conclusion, if the data in the big data system are processed in an anonymised form, there are no problems relating to the informed consent or to the information of the purposes. If the data are pseudoanonymised, data processing is subject to the rules set by the general regulation; on the other hand, in order to obtain information from information, big data use different algorithms with different and often changing purposes, even within the scope of the same search.

[40] General regulation art. 6, paragraph 4: ... where the processing for a purpose other than that for which the personal data have been collected is not based on the data subject's consent or on a Union or Member State law ... in order to ascertain whether processing for another purpose is compatible with the purpose for which the personal data are initially collected, take into account, inter alia … specific elements follow that the controller can assess to establish under his or her responsibility that the processing to be carried out is compatible with the purposes for which consent has been requested, or with the European or national regulation legitimizing such processing … The responsibility for the assessment lies with the controller, who may share the responsibility for the choice with the competent data protection authority, whenever deeming an adequate impact assessment necessary.

6. *Big data application examples*

The big data technology has recently been used to create a reputational database, through which reputational profiles of natural and legal persons, contractors and subcontractors, suppliers, distributors, business partners, aspiring employees and customers can be developed and from which a precise reputational rating score of subjects surveyed can be obtained, through a sophisticated software program, in order to allow customers to verify their real credibility. The mechanism consists in uploading on a web platform data relating to a specific person, such as judicial records, tax regularity certificates, qualification certificates, diplomas, complaints, lawsuits, judicial measures, so that reputational consultants, such as lawyers and accountants specially hired by the data controller, can verify and ensure their authenticity and integrity. This data can be updated either directly by the data subject or by third parties: think of customers interested in the credibility of contractors, of employers interested in the credibility of workers, of suppliers interested in the credibility of customers. The Guarantor has adopted an injunction prohibiting the type of processing proposed, stating that it does not comply with various provisions of the privacy code: from the violation of the principles of necessity, proportionality and purpose of processing to the failure to comply with the rules on disclosure and consent[41]. Another example of the implementation of big data is in the political and electoral fields, where there are certainly other elements of legal assessment. In fact, the profiling of elector citizens raises further problems, if we just consider processing operations on personal data that have a sensitive nature as they are capable of revealing the political orientation of those concerned. The collection, analysis and reprocessing of voters' personal data may, in fact, concern the membership or affiliation to a political party as well as political opinions expressed on the profiles of social networks; in turn, these sensitive data can then be crosschecked with personal and demographic data, which fall within the category of common data: age, income, and marital status. These types of operations are used to isolate and understand the social preferences of citizens, or what citizens want their representatives to do and therefore what

[41] Provision dated 24 November 2016, web doc. no. 5796783. The Guarantor considered the proposed processing not compliant with the provisions of the Privacy Code, pursuant to Legislative Decree no. 196 of 2003.

candidates should propose in order to be elected[42]. The principle of finality established by the regulation excludes that the easy availability of such data on the web can authorize processing for any purpose, but imposes that such data are processed only for the purposes underlying their initial and original publication[43]. And yet the application of big data to the electoral field means that the voter is no longer simply a free citizen, since he or she is increasingly assimilated in all respects to a consumer whose tastes, preferences and needs are to be anticipated, with inevitable consequences both on the concept of citizenship and on that of political participation[44].

Personal data are victims of a sort of two-way vortex: the citizen who wants to get informed and participate does it more and more through the web and searches, navigates, leaves digital traces, and spreads, even unknowingly, his or her personal data: by so doing, he or she increasingly feeds the already impressive amount of big data, continuing to actually offer information, contents and data for the profiling operations that have a broad and inevitable impact on the dignity of the human person.

---

[42] The Guarantor for the protection of personal data intervened precisely in the matter of data processing at political parties and in electoral propaganda activities, addressing parties, political movements, committees of promoters and supporters and individual candidates; with a general provision it was decided, among other things, to establish a ban on use of all data found freely on the web for electoral propaganda and related political communication, referring in particular to data collected automatically on the Internet through special software or data obtained from social networks, forums or newsgroups.

[43] Provision on the processing of data at political parties and exemption of information for the purpose of electoral propaganda of 6 March 2014, web doc. no. 3013267.

[44] L. CALIFANO, *Brevi riflessioni su privacy e costituzionalismo al tempo dei big data* in *Federalismi.it*, no. 9, 2017. "Thus one wonders again what idea of democracy the use of such instruments presupposes and, therefore, what new forms popular sovereignty and political representation take on within contemporary society at the time of big data … More recently, thanks to the evolution of the media and mainly to the dissemination of the Internet and ICTs, it has become a hybrid democracy, characterized by a form of communication and political participation that 'crosses the boundaries between network, TV, newspapers, old and new media' and by a high degree of dis-intermediation. In this context, citizens demand more and more information and data to know and control political power with awareness and use the web to participate and, in turn, get organized and mobilized. All this can certainly be a resource for representative democracy, if and only if we limit ourselves to seeing the network as an instrument and not an end in itself and for itself, as the overcoming of the democratic intermediation offered by parties and other intermediate subjects such as unions and associations … In short, with the advent of the web and the idea, albeit overshadowed, of a perennial consultation of the electorate, the hypothetical risk is the genetic transformation of citizens' control power into an automatic mechanism of delegitimising institutions, thus outlining a paradoxical result precisely in relation to the original facts of popular sovereignty and representation. The outcomes, both possible and paradoxical, generated by an abuse of big data involve the debate on the developments of contemporary democratic constitutionalism and the real exercise of the fundamental right to the protection of personal data".

7. *Big data and economy*

The silent transformation of the twenty-first century is the revolution of data, as well as the revolution of statistical analysis, macro-social studies, systems to protect the privacy of individuals, knowledge. Data silently turn into information and such information silently revolutionizes the sectors where it is used. This percentage of quantifiable knowledge is today one of the pillars on which companies must implement their business strategies; since data are easily transformed from knowledge into money, numbers become real expansionist weapons for the businesses of the twenty-first century. The phenomenon of big data must first of all be inserted within a broader framework that can be defined as the fourth industrial revolution or revolution 4.0. This transformation affects the entire global socio-economic system, through the creation of a true digital ecosystem, made up of digital identities, digital relationships, digital contents, big data and multi-screens. Companies must be ready to change from within to react quickly to this digital tsunami and turn it from an obstacle into an opportunity. There is no more space for a pillar organization, but more flexibility and real-time adaptation will have to be ensured. Relations between companies and their stakeholders are not exempt from this change; the relationship between the customer and the company becomes increasingly bilateral and immediate, determining increasing value to the final consumer and collecting an infinite amount of data to generate useful knowledge for satisfying him or her. The phenomenon of big data is therefore to be intended as a powerful tool for analyzing and understanding the complexity of the ecosystem and exploiting it. Every company, large or small, organization or government capable of gathering relevant information on players and the surrounding context can adapt its strategy, organization, objectives and next moves in an appropriate, conscious and probably winning way[45]. And yet the use of big data is not limited only to behavioral advertising and profiling, but always affects new and different sectors.

[45] Antonello SORO *Big Data e Privacy - La nuova geografia dei poteri*, Records of conference on 30 January 2017 "…The new economy made of increasingly interconnected technology, favored by the expansion of mobile Internet, fueled by the widespread presence of intelligent sensors, is characterized by large volumes of data, the infinite heterogeneity of sources from which they come and the speed of the systems that analyze them. The ability to extract meaningful and functional information from data requires the development of sophisticated technologies and interdisciplinary skills that work closely with each other. In this framework, advances in computing power play a central role in the analysis of Big Data and the acquisition of knowledge. And in

8. *Big Data*: *information asymmetry*

The concentration of information in the hands of a few operators is no news[46], but in the case of big data, in addition to proliferation or mass indexing, another important feature is the predictive ability that analyses conducted with sophisticated tools on such large aggregations can achieve, and then rise to a great strategic and socio-political value. In fact, they allow the emergence of unpredictable inferences and unsearched phenomena, although they do not have a sampling strategy such as statistical analyses and could consequently distort the results, while avoiding the erasure of useful and relevant data. This predictive ability is an undoubted advantage in economic terms, as well as in the social context for power groups. In fact, data are not accessible to everyone, free access information is not available and those who do not have the technology to analyse this data cannot obtain the appropriate predictive results. This means that in a society that is increasingly in need of data functional to the decision-making processes, data sets and their processing capacity are means for acquiring significant information power, also in relation to the so-called open data. Another important peculiarity is that most of the time large databases are managed with cloud computing systems and so data controllers sometimes reside in locations other than the location of the cloud; geolocation involves consequences in terms of applicable laws, greater or lesser protection of data and last but not least the conditioning of local political power. It is no coincidence that the new European regulatory framework provides for the protection of European data wherever they are even outside the European borders. Besides their supervising tasks, the independent supervisory authorities could plan actions aimed at affecting above all the security and uniformity of the standards, to affirm the transparency of the information society according to the accessibility of data and the sharing of information. It is also true that if the society in which we live is the society of information, it does not constitute a definitive end point since the information phenomenon varies over time in relation to the different parameters concerning its distribution, size and

---

the not too distant future, artificial intelligence will offer effective solutions to satisfy the most diverse needs, thanks to algorithms able to learn and improve their skills autonomously.

[46] The analogy between the era of the main frames and the current one of cloud computing and big data is significant, because once again the large IT resources are concentrated in the hands of a few subjects and are also physically aggregated in huge data centers.

possibility of an effective analysis of contents. It is equally indisputable that the asymmetric relationships between those who provide data and those who exploit them are resolved in favor of those who manage digital platforms.

Abstract

The traditional activity of searching for one or more information online has turned into the use of the web to find new information, cross-checking the available data according to algorithms aimed at extracting data from data, information from information. In this new reality, the approach adopted to ensure effective personal data protection has changed, as confirmed by the new regulation.

Camerino, aprile 2018.