

Big Data between privacy and copyright

AMELIA BONGARZONE

INTRODUCTION

The possibility provided by Big Data to develop interpretative, analytical and even predictive models of human events and behaviours, difficult to imagine until a few years ago, draws the attention of private law and prompts reflection on a problem which, due to its not only patrimonial implications, is assuming exponential proportions, in many ways out of control. The game is played between needs certainly deserving of protection, which however, in their meeting and not infrequently collide, call into question an eurythmy not easy to reach. The availability of a large quantity of information, while proving to be a source of progress and innovation in terms of public and private benefits, also highlights the exigency to protect compromised legal situations. Among these, in addition to the freedom of economic initiative and all that this entails, the rights of personality stand out.

Opportunities for improvement in the public sector (e.g. in the areas of security, prevention, strengthening of services, promotion of life's quality, health, culture, etc.), as well as the indisputable competitive advantages for private entrepreneurship (of launching, consolidation *et similia*, in the reference markets or in those not yet explored) are counterbalanced by the risks of abuse perpetrated against individuals who let their data more or less unconsciously and in exchange

for only apparently free services (Resta and Zeno-Zencovich 2018, 422). The benefits are certainly as many as the critical profiles and the sense of their proportion is given back by the sometimes insidious methods of collection, by the often uncertain places of conservation, by the not always transparent criteria of selection and analysis as even the frequently unpredictable treatments, whose description is made even more evanescent by the possible regeneration of the acquired data and by the continuous refinement of the technologies able to use them.

Faced with a phenomenon that is undergoing such rapid transformation and of transversal importance, the jurist is called upon to question himself on the adequacy and efficiency of the current regulatory framework, without indulging in dogmatic positions, in the awareness of the relativity of the institutes and the precariousness of the experiences. Reason can be found in the idea of law as a structure, “conditioned by economic-social relations” and “in turn conditioning the wider and more complex reality of which it is historically an integral part” (Perlingieri 2006, 161).

ABSENCE OF A CONVENTIONALLY ACCEPTED DEFINITION OF BIG DATA AND ATTEMPT TO DESCRIBE THEM

The perspective of investigation opens the field to a complex panorama whose difficulties can be seen, at first glance, where we reflect on the absence of a conventionally accepted definition of Big Data, an obvious symptom of a “chaotic state of the art” (De Mauro, Greco and Grimaldi 2016, 128).

In the heterogeneity of the proposals, the reference can be 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” (European Parliament resolution 2017, Recital A). It is, of course, a very composite process, whose phases (collection, storage, aggregation, analysis, comparison, use and reuse) are developed in a variety

of contexts and engage a plurality of stakeholders which, each for their own expertise, participate to the achievement of the final result, usually delivering an intermediate one.

The constant is constituted by the *data*, undisputed protagonists of a historical moment, threatened by the spread of the so-called *surveillance capitalism* (Zuboff 2019; Foster and McChesney 2014). The highly evocative syntagm alludes to a new economic order that, through the observation of the entirety of individual and collective behaviours, is capable of allowing anticipatory, probabilistic indications of the preferences of the subsidiaries and, even more so, of directing their interests and choices, in a perspective of reification of human experience, reduced to a precious commodity of exchange.

Upstream, there is a transfer of data by users, mainly consumers, who provide them immediately (e.g. by adhering to very common loyalty programs such as customer cards of the supermarket) or generate them, not infrequently passively, through the use of devices (Internet of things) or through online access, enticed in most cases by the use of services and benefits, otherwise unavailable, in the face of which the initial distrust is bound to succumb. Thus, with a mechanism as much underestimated as pervasive, often supported by a sort of coercion, not even particularly subtle, based on the *do ut des* (I give you the advantage if you give me the data), a considerable impact on the freedoms and fundamental rights of the person is consumed.

INCIDENCE OF BIG DATA ON PERSONALITY RIGHTS. OBSOLESCENCE OF THE DISTINCTION BETWEEN PERSONAL AND NON-PERSONAL DATA. RELEVANCE OF DATA

The European approach to the problem of Big Data, originally focused on promoting the transition process to a data-driven economy capable of competing with the world's giants, soon became aware of the danger looming over the existential sphere of the affiliates, highlighting the need for a new, adequate, empowerment able to contain otherwise devastating outcomes. In this direction is placed the European Regulation 2016/679, the GDPR, for the protection of personal data, issued in an attempt to combine the improvement possibilities,

on several fronts now indispensable, offered by the strategic asset of data, with the need to preserve human existence, in some of its essential values, from the invasive use of a certain technology.

The balancing is demonstrated by the definitive emancipation of the right to protection of personal data from proprietary paradigms through the confirmation that it is a fundamental right of natural persons (Rodotà 1991, 526, 531; Zeno-Zencovich and Giannone Codiglione 2016, 33; Directive EU 2019/770, Recital 24), according to the Article 1 of the GDPR, and by the simultaneous provision of free movement of data within the Union, in a prospect of cooperation and exchange, evidently functional to digital progress. In the same direction also leads Recital 4, stating that the “processing of personal data *should be* designed to serve mankind” and specifying that “the right to the protection of personal data is not an absolute right; it must be considered in relation to its function in society and be balanced against other fundamental rights, in accordance with the principle of proportionality”.

It does not seem superfluous to point out that the use of the conditional *should be* is significant, where, unless we want to consider its randomness, it reveals the dissatisfaction for a proposed objective but, evidently, not yet achieved. This issue has a primary importance, already clearly indicated in the choice of the legislative form adopted. The use of the Regulation, because of its general scope, the mandatory nature of its elements and its direct application in each Member State, clearly reflects the necessity to ensure uniform levels of protection in the EU.

In 2018, the European Regulation 2018/1807 on non-personal data joins the GDPR. Once again, the European legislature establishes the principle of free movement within the Union, unless public security reasons justify a restriction or prohibition, and points out that, in any event, “the legal framework on the protection of natural persons with regard to the processing of personal data, and on respect for private life and the protection of personal data in electronic communications and in particular Regulation (EU) 2016/679 [...] are not affected by this Regulation” (Regulation EU 2018/1807, Recital 8). A further demonstration of the ongoing balancing of interests.

Personal data and non-personal data, therefore, as the *summa divisio* from which there would appear to be no reason to deviate even for Big Data. However, what criteria for the distinction? The question, the importance of which is made evident by the fact that only the provision of personal data generally requires the consent of the person concerned, would seem not to generate perplexity and to find immediate confirmation in the combined reading of certain provisions of the cited Regulations.

The reference is to Article 4, in point (1), of the GDPR, according to which personal data “means any information relating to an identified or identifiable natural person (‘data subject’), with the specification that “an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person”. The notion of non-personal data, instead, is achieved by exclusion, since Article 3 of Regulation (EU) 2018/1807 defines such “data other than personal data as defined in point (1) of Article 4 of Regulation (EU) 2016/679”.

Anyway, if on the definition level the difference is rather easy, the same does not happen in the actual resolution of cases. An overlap, confusion or even a conversion of data is not at all remote, especially when one reflects on the exploratory capacity of technology, the constant development of which advises against the assumption of definitive positions. And indeed, if at present it cannot be radically excluded that anonymous or anonymized data - when involved in a complex process of analysis, perhaps crossed with additional data or simply processed with more refined devices - may lead to a result that allows the tracing or inferring of personal information, even more it is not unlikely that this will happen in the future. Leaving out the possibility that (apparently) neutral data, inasmuch as they are not immediately referable to an “identified or identifiable natural person”, reveal on the contrary their nature of personal data if analyzed from another subject or from another perspective. Consider, for example, of the image of a square that might interest some for the artistic pro-

files of the fountain that is in the center, others perhaps to identify the people present there (AGCOM 2018, 35), or even think of the information obtained from sensors installed on cars which, normally instrumental to an investigation relevant to the object, much can actually reveal about habits, inclinations, preferences of the driver, for example according to the places frequented.

The transversal nature of the contents, which is exponentially amplified for Big Data, therefore undermines the traditional distinction by type of data, voting for its obsolescence. Thus, the opportunity arises to reconsider the issue, starting from data *tout court*.

EXTENSIBILITY TO BIG DATA OF THE PRECAUTIONS FORESEEN IN THE GDPR: CRITIQUE

The proliferation of sources of (personal) data collection, and the intertwining of relations that develop around them, lead us to dwell on one of the possible presuppositions of their circulation: consent (Regulation EU 2016/679, Article 6; Nazzaro 2018, 1248).

Defined in Article 4, in point (11), of the GDPR, as “any freely given, specific, informed and unambiguous indication of the data subject’s wishes by which he or she, by a statement or by a clear affirmative action, signifies agreement to the processing of personal data relating to him or her”, many other provisions and recitals of the same body of legislation deal with it. In this way, a regulatory plan is outlined to ensure requirements and conditions for a conscious transfer by settlor, based on a genuinely free choice, the recurrence of which should be doubtful if, for example, the person concerned is “unable to refuse or withdraw consent without detriment” (Ibid., Recital 42). The picture is completed by the attention paid to the presupposed act, *id est* the information, which must be expressed “in a concise, transparent, intelligible and easily accessible form, using clear and plain language” (Ibid., Article 12, paragraph 1) on the basis of the principle according to which “it should be transparent to natural persons that personal data concerning them are collected, used, consulted or otherwise processed and to what extent the personal data are or will be processed” (Ibid., Recital 39) and in the wake of which a series of precautions are envisaged.

The discipline certainly seems suitable for a binary model, in which there is a well-identified titular subject, who collects the data of a natural person and processes them, or has them processed by his auxiliaries, for purposes punctually established. A system, in short, in which the owner is immediately able to inform the person concerned of the destination of the data, who, in turn, is in the position of being able to fully accept or disagree. Otherwise, however, it happens in the Big Data scenario, where, to simplify matters extremely much, it can be said that the data rise up a long chain, only as a result of which they are aggregated and examined by an 'upstream' holder who may not coincide, and usually does not coincide, with the one who initially acquired them. Moreover, before the data are analyzed, it is not certain what correlations and results their processing will reveal, nor can it be reliably established how and for what purpose they will be used in the end. As a result, it is more or less impossible for the settlor to make a conscious choice, as he often lacks even a definite perception of the data he is giving up and although he is neither in a position to ascertain *ex ante* what the destination will be, since, as has been said, it may happen that not even the actors called upon to deal with the relative process of decomposition and composition of the accumulated material know this.

The hypothesis is accredited by the increasingly frequent use of automated decision-making processes, including profiling, in which case the intervention of the artificial intelligence and algorithms, able to handle potentially infinite dimensions, often makes it difficult to identify the interlocutors of the treatment process. In the absence of such essential coordinates, the interested party is effectively precluded from using the remedies predisposed by the GDPR (access, rectification, cancellation, oblivion, limitation, etc.), so that the rules on information and consent show all their inadequacy. What emerges is an asymmetrical knowledge, stigmatized from many sides, fed by the opacity of the processes used for data processing, in place of the much celebrated transparency, and further aggravated by the lack of adequate literacy of the data subject, devoid of the cognitive baggage necessary to understand the complex dynamics involved, certainly beyond the ordinary skills widespread among the population.

A remedy would appear to be provided by Article 35, paragraph 1, of the GDPR, which requires a “data protection impact assessment” to be carried out if it is assumed that “a high risk to the rights and freedoms of natural persons” may result from a type of processing. The aim is to prepare the most suitable measures to avoid the danger, with the intervention, in case of inability to contain it, of the supervisory authority (Ibid., Article 36). However, it should be noted that the application of the mentioned provision, as of the entire GDPR, is limited to personal data concerning natural persons, while very often the trap is consumed in the unpredictable transformation of anonymous or non-personal data that, for example, as a result of combinations with other data become specifically referable to individuals (or small groups, such as the family), sometimes even revealing sensitive aspects (political, religious or sexual orientation, etc.). Among the different risk profiles, this is perhaps the most felt and it is in such cases that the current protection systems show their inefficiency, with worrying repercussions not only on the level of self-determination of the individual - based on a consent that continues to preach, but only formally, as free, specific, informed and unequivocal - but also on the power of control and inhibitory of unauthorized exploitation.

On this point, one reflects, once again, on the various digital interactions where the usability of the contents, allegedly free, is instead radically conditioned to the authorization to the processing of data (Di Porto 2016, 13; Resta and Zeno-Zencovich 2018, 416, 422; Nazario 2018, 1250) which is usually done in a mechanical way and, often, without it being even clear what data is, in fact, available. It should also be borne in mind that in assessing “whether consent is freely given, utmost account shall be taken of whether, *inter alia*, the performance of a contract, including the provision of a service, is conditional on consent to the processing of personal data that is not necessary for the performance of that contract” (Regulation EU 2016/679, Article 7, paragraph 4). It is legitimate to ask oneself, in how many cases a thoughtful investigation would lead to the conclusion of a truly free and conscious transfer.

PROCESSING OF BIG DATA AND OWNERSHIP OF THE RESULT:
COPYRIGHT AND ACQUISITION BY SPECIFICATION

It is just the case to point out that the generic nature of the term *data*, and in particular its ability to tendentially describe every immaterial representation of events and/or conducts, pushes to circumscribe its reference to those elements that are subject to computational use, *id est* suitable to be object of automated or artificial elaboration. In this way, the data constitute an observable entity of which the free movement, provided for by European legislation, participates in the process of legal objectification. Since, in fact, a circulatory event can concern only persons, services or goods, tangible or intangible, it does not appear risky to note that guaranteeing the possibility with regard to the data presupposes their qualification of goods, precisely of intangible goods (Resta 2010, 20).

It is worth clarifying that the single data, considered in isolation, enjoy a value certainly different from the sum of the data of which they are a possible element; just as the sum of the data collected, as components of a whole, is not equal to the value of the whole in its complexity, since the whole is something more than the sum of its parts.

In this perspective, it should be pointed out that a huge amount of data stored is still only the raw material, destined to be processed within a complex process. It is here that the use of the algorithmic instrumentation allows the generation from the inputs received of an *aliquid novi*, different from the single goods which have originated it, with respect to which manifests the problem of ownership and the individuation of the protection mechanisms. In particular, the impossibility of immediately bringing the result obtained back to the original expression of an idea, and therefore to human creativity, raises perplexity as to its inclusion in the scope of works of intellect. It is argued that while the structuring of a database, by virtue of the selection and organization of the content, can lead to the recognition of a copyright of the founder, not so for unstructured databases, typical of Big Data, which would lack the 'creative touch'.

Nevertheless, beyond the reservations aroused by radically negatory positions, it must be considered that the composition of a databa-

se involves substantial investments, not even neglected by the Italian law on copyright (l. 633 of 1941). From an antiparasitic point of view, Article 102 *bis* grants the maker of the database the right to prohibit, for a certain period, “the operations of extraction or re-utilization of all or a substantial part of the same”, “regardless of whether the database can be protected under copyright or other rights and without prejudice to the rights to the content or parts thereof”.

It is questionable whether, and in what terms, such cautions can be extended to Big Data, without affecting the superior observation according to which one thing is the database, what set of data, another thing is the product that from the processing of those data is obtained. The transformation of the data, constituting the original material, into a *res nova* having a distinct socio-economic individuality, and therefore the transformation of goods into goods with added value, evokes in some way the rule contained in Article 940 of Italian Civil Code, whose foundation is found in the principle of ownership of the goods to those who produce them, even if the material used belongs to others, unless the value of the latter greatly exceed that of the processing. The point would deserve to be examined in more detail, nevertheless, the boundaries entrusted to the present reflections only allow to mention it.

ESSENTIAL BIBLIOGRAPHY

AGCOM, “Big data. Interim report nell’ambito dell’indagine conoscitiva di cui alla delibera n. 217/17/CONS”, June 2018. <http://www.agcom.it>.

De Mauro, Andrea, Marco Greco, and Michele Grimaldi. “A formal definition of Big Data based on its essential features”. *Library Review*, 65 (March 2016): 122-135. doi: 10.1108/LR-06-2015-0061.

Di Porto, Fabiana, “La rivoluzione big data. Un’introduzione”. In *Big data e concorrenza* edited by Di Porto Fabiana, *Concorrenza e mercato*, no. 23 (2016): 5-14.

Directive (EU) 2019/770 of the European Parliament and of the Council of 20 May 2019 on certain aspects concerning contracts for the supply of digital content and digital services. *Official Journal of the European Union*, May 22, 2019: L 136/1-L 136/27.

European Parliament resolution of 14 March 2017 on fundamental rights implications of big data: privacy, data protection, non-discrimination, security and law-enforcement (2016/2225(INI)). <http://www.europarl.europa.eu>.

Foster, John Bellamy, and Robert W. McChesney, "Surveillance Capitalism. Monopoly-Finance Capital, the Military-Industrial Complex, and the Digital Age". *Monthly Review. An independent socialist magazine*, Jul 01, 2014. <http://www.monthlyreview.org>.

Nazzaro, Anna Carla, "L'utilizzo dei *Big data* e i problemi di tutela della persona". *Rassegna di diritto civile*, no. 4 (2018): 1239-1260.

Perlingieri, Pietro. *Il diritto civile nella legalità costituzionale secondo il sistema italo-comunitario delle fonti*. 3th ed., Napoli: Edizioni Scientifiche Italiane, 2006.

Regulation (EU) 2018/1807 of the European Parliament and of the Council of 14 November 2018 on a framework for the free flow of non-personal data in the European Union. *Official Journal of the European Union*, November 28, 2018: L 303/59-L 303/68.

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). *Official Journal of the European Union*, May 4, 2016: L 119/1-L 119/88.

Intelligenza artificiale y datos masivos...

Resta, Giorgio and Vincenzo Zeno-Zencovich, “Volontà e consenso nella fruizione dei servizi in rete”. *Rivista trimestrale di diritto e procedura civile*, no. 2 (2018): 411- 440.

_____. “Nuovi beni immateriali e *numerus clausus* dei diritti esclusivi”. In *Diritti esclusivi e nuovi beni immateriali* edited by Resta Giorgio, Milanofiori Assago (MI): Wolters Kluwer Italia S.r.l, 2010: 3-72.

Rodotà, Stefano, “Privacy e costruzione della sfera privata. Ipotesi e prospettive”. *Politica del diritto*, no. 4 (December 1991): 521-546.

Zeno-Zencovich, Vincenzo and Giorgio Giannone Codiglione, “Ten legal perspectives on the “big data revolution””. In *Big data e concorrenza* edited by Di Porto Fabiana, Concorrenza e mercato, no. 23 (2016): 29-57.

Zuboff, Shoshana. *The age of surveillance capitalism. The Fight for the Future at the New Frontier of Power*. London: Profile Books Ltd, 2019.