Personal Data v. Big Data in the EU: Control Lost, Discrimination Found

Maria Bottis, George Bouchagiar
Ionian University, Corfu, Greece
Email: botti@otenet.gr, georgebouchayar@yahoo.gr

Abstract
We live in the Big Data age. Firms process an enormous amount of raw, unstructured and personal data derived from innumerable sources. Users consent to this processing by ticking boxes when using movable or immovable devices and things. The users’ control over the processing of their data appears today mostly lost. As algorithms sort people into groups for various causes, both legitimate and illegitimate, fundamental rights are endangered. This article examines the lawfulness of the data subject’s consent to the processing of their data under the new EU General Data Protection Regulation. It also explores the possible inability to fully anonymize personal data and provides an overview of specific “private networks of knowledge”, which firms may construct, in violation of people’s fundamental rights to data protection and to non-discrimination. As the Big Data age is here to stay, both law and technology must together reinforce, in the future, the beneficent use of Big Data, to promote the public good, but also, people’s control on their personal data, the foundation of their individual right to privacy.

Keywords
Personal Data, Consent, Control, Discrimination

1. Introduction
In the age of Big Data (King & Forder, 2016: p. 698; Giannakaki, 2014: p. 262), information (Lessig, 2006: pp. 180-185; Summers & DeLong, 2001) fully confirms its etymological origin (Araka, Koutras, & Makridou, 2014: pp. 398-399) and becomes abundantly available (Himma, 2007). It constitutes a mass-produced good (Battelle, 2005), consumed as a commodity, rather than leveraged as a tool for personal growth of the individual or the development of democratic societies (Koelman, 2006). Information, including personal data (i.e. “any information
relating to an identified or identifiable natural person (‘data subject’); an identifiable natural person is one who can be identified, directly or indirectly”, see Article 4(1) of GDPR, has acquired independent economic value (Pasquale, 2015: p. 141; Hugenholtz & Guibault, 2006) and, thus, new and innovative business models constantly emerge and dominate the market. For instance, a business that owns no vehicles (such as Uber) may dominate the “taxi market”, while large “hoteliers” (e.g. Airbnb) may own no property at all (Chesterman, 2017). Firms, thus, process raw (Giannakaki, 2014), unstructured (Mayer-Schönberger & Cukier, 2013: p. 47) and personal (Picker, 2008; Tene, 2008) data (Scholz, 2017: pp. 9-12) from a multiplicity of sources (Tene, 2011). The Internet of Things (Panagopoulou-Koutnatzi, 2015a) only dramatically accentuates the huge potential of these vast collections of information (Petrovic, 2017: p. 187).

How do firms obtain data from people? A way to extract them is a peculiar quid pro quo: Data constitute the “fee” that users “pay” for multiple “free” digital services. This “deal” has not only been accepted by a number of institutions (European Commission, 2017), but has also become both a global phenomenon and an everyday business practice. While providing a service, e.g. an e-mail service, a firm can collect and process personal information contained in the e-mail (Prins, 2006: p. 229). Data collected may also concern e.g. the language the user speaks or her mobile phone or her real location (or even device-specific information, such as hardware model, operating system version, unique device identifiers and mobile network information). In addition, when a user stores her digital, and sometimes personal, files using Cloud Computing (for instance, Dropbox, Google Drive, Sky Drive, i-Cloud), the provider, i.e. the company that offers the cloud service, may process data contained in the user’s (Lidstone, 2014) files stored in the “clouds” (Morozov, 2011: p. 286). Finally, a cornucopia of data that relate to a user’s health, movement or just living patterns (e.g. heart rate, blood pressure, or even sleep times) may be collected and processed as long as users, accompanied by smart devices (Brabazon, 2015) and selecting from innumerable applications (Mayer-Schönberger & Cukier, 2013: p. 94), measure themselves during their everyday physical activities.

Thus, countless online activities, a standard feature of everyday life, involve the production and the processing (Tene & Polonetsky, 2013: p. 255) of an unprecedented volume of personal data (Committee on Commerce, Science, and Transportation, 2013). Although it is doubtful whether someone’s recorded heart rate constitutes personal data, many, or perhaps most of the kinds of information described above as examples are, actually, personal data under the General Data Protection Regulation of the EU. This is because in the age of Big Data, the collection of a huge volume of data enables firms to draw numerous conclusions that relate to one person and makes it possible to identify a natural person. Provided that an item of information collected by a company relates to a natural person, who can be identified, directly or indirectly, this information is personal data (CJEU, 2003: p. 27; A29DPWP, 2007, 2008). In other words, the criterion that has to be met, and which “makes” the data personal is not actual
identification, but the capacity to identify, directly or indirectly, one person 
(Tene, 2008: p. 16). To sum up, if there is a capacity to identify the individual, to 
whom the “recorded heart rate” mentioned above relates, the data are personal 
and in particular health data (Panagopoulou-Koutnatzi, 2015b) and are fully r-
egulated by the GDPR.

After having collected masses, sometimes, of personal data, which users pro-
duce “just by existing” (Powles & Hodson, 2017; Gray, 2015, Brabazon & Red-
head, 2014), many firms behave as “owners” (Prins, 2006: pp. 223-224; Almunia,
2012) of this information (Cohen, 2000: p. 1375), by exchanging it (O’Neil, 2016:
151; Prins, 2006: p. 228; Hoofnagle, 2003; Michaels, 2008) or by further 
processing it (Crawford & Schultz, 2014). In this case, some scholars even talk 
about theft of humanistic property (Mann, 2000), this theft having been perp e-
trated by private enterprises, while others argue that natural persons should re-
ceive fair compensation for the collection, processing, exchange and use of their 
personal data (Litman, 2000), since there should be no free lunch when it comes 
to invading privacy (Laudon, 1996: p. 103).

Given the above practices, which show at least an important loss of the user’s 
control over her personal data, this paper examines the validity and law fulness of 
the data subject’s consent to the processing of their personal data, studies the 
inability to anonymize such data and also, provides an overview of specific “pri-
ivate networks of knowledge”, which any digital company is able to build (own 
and control) in violation of the fundamental right to non-discrimination.

2. The Subject’s Consent to Data Processing

One of the fundamental principles of data protection law in Europe and beyond 
is respect for personal autonomy (Bottis, 2014: p. 148). Legal provisions on per-
sonal data safeguard constitutionally-protected rights to informational 
Hence, it has been consistently supported by authors that the fundamental (A-
rticle 8(1-2) of CFREU; Article 16(1) of TFEU) right to the protection of personal 
data refers to control by the subject over the processing of her data (Oostveen & 
Irion, 2016; Rengel, 2014). The key tool for a legal control of personal data is the 
subject’s consent to the processing (Tene & Polonetsky, 2013: pp. 260-263; So-

The European lawmaker recently regulated the protection of natural persons 
with regard to the processing of personal data and the free movement of such 
data (GDPR), and in this Regulation, took into account these aspects of control 
(Recitals (7) and (68) of GDPR) and legislated that the previous subject’s consent 
shall be a necessary prerequisite for the lawfulness of data processing (Article 
6(1)(a) of the GDPR). In particular, under the GDPR, the collection and 
processing (Article 4(2) of the GDPR) of personal data shall be lawful if the data 
subject has given consent to the processing of his or her personal data (Recital 
(4) and (42) of the GDPR) for one or more specific purposes (Article 6(1)(a), 
Recital (32) of the GDPR). Moreover, “consent” of the data subject means 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 (Recitals (42), (43), Articles 7(4), 4(11) of the GDPR).

One would assume, therefore, that a “single mouse-click” on any privacy policy’s box, by which users may give their consent, should not be considered to fulfill the criterion of “freely given, specific, informed and unambiguous” indication of the data subject’s wishes by which the individual has to signify agreement to the processing. Quite the opposite is true: under Recital 32 of the GDPR, consent can also be given by “ticking a box, when visiting an internet website” (the repealed Directive 95/46/EC makes no mention of the capacity to give consent simply by ticking a box).

Thus, the data subject’s consent to the collection and processing of her personal data may be validly and lawfully given by a single “mouse-click” on the box of a webpage, the terms of use and the privacy policy of which—almost—nobody reads (Turow, Hoofnagle, Mulligan, Good, & Grossklags, 2006: p. 724; Pingo & Narayan, 2016: p. 4; Gindin, 2009; Chesterman, 2017). Given that, as documented, in most cases the users “generously click” on any box that may “pop-up” (Vranaki, 2016: p. 29), private enterprises legally (and with individual’s “freely given, specific, informed and unambiguous” wishes) process (e.g. collect, record, organize, structure, store, adapt, alter, retrieve, consult, use, disclose, disseminate, make available, combine, restrict, erase or destroy) personal data.

3. Anonymizing Data: A Failure?

In several cases, after having collected personal data, firms anonymize them. This means that “effective” measures are taken and data are further processed in a manner which renders the re-identification of the individual impossible (Hon, Millard, & Walden, 2011; Stalla-Bourdillon & Knight, 2017). Anonymization constitutes further processing (A29DPWP, 2014) and always comes after the collection of data. Hence, given the legislated validity of consent that users have already given often by a single “mouse-click”, companies may legally anonymize their collection of personal data. Anonymized (ex-personal) data can be freely used e.g. shared with third parties, sold etc as the rules of data protection do not apply to “personal data rendered anonymous in such a manner that the data subject is not or no longer identifiable” (Recital (26) of the GDPR).

But in the age of Big Data, there is probably no safe way to render personal data truly anonymous (Scholz, 2017: p. 35; Schneier, 2015: pp. 50-53). Even after “anonymization”, the data subject remains technologically identifiable (Ohm, 2010: p. 1701; Sweeney, 2000; Golle, 2006; Gymrek, McGuire, Golan, Halperin, & Erlich, 2013; Bohannon, 2013; Narayanan & Shmatikov, 2008). The inability to anonymize personal data in a Big Data environment is due to the collection and correlation of a huge volume of data from multiple sources. The result is the possibility to draw “countless conclusions” about an individual, who may be
identified, directly or indirectly (Tene & Polonetsky, 2013: p. 257; Cunha, 2012: p. 270). In other words, anonymization can only be achieved in “Small Data” environments, given that the volume and the variety of data processed in the world of Big Data, facilitate and encourage (re)identification of any individual (Mayer-Schönberger & Cukier, 2013: p. 154).

We see, therefore, the anonymization of personal data, in a Big Data environment, portrayed as a failure. The same technology, which reassured us that we could not be identified, and so our personal data could be used for some noble purposes as, for example, medical research, now betrays us. A huge data set is almost magically, and reassuringly, turned anonymous, and then, adding a piece of information or two, it is turned back, some point later in time, to full identification (De Hert & Papaconstantinou, 2016: p. 184). If this is the case, where is our consent in this situation? A “single click” consent to this processing is from the outset pointless. The very specific purpose of the processing for which the individual has to give her initial consent has often, at the time of “mouse-click”, not even been decided yet by the firm who is the controller (Mayer-Schönberger & Cukier, 2013: pp. 152-153; Giannakaki, 2014: pp. 263-264). Thus, when users in fact ignore the final purpose (Steppe, 2017: p. 777; A29DPWP, 2008) for which consent is given (Bitton, 2014: p. 13), it is fair to support that they have lost control over their data (Solove, 2013: p. 1902). If no genuine consent can be given and if anonymization is indeed practically impossible, then there is no control at all (Carolan, 2016; Danagher, 2012). But this loss of control contrasts strongly with the goals and principles of the constitutional, in Europe, right to the protection of personal data. It defeats the raison d’être of all previous European legislation on data protection all the way since 1995.

4. Knowledge and the Fundamental Right to Non-Discrimination

Although the right to the protection of personal data is fundamental, probably not many people are aware of this right and much fewer have been documented to exercise powers which this right gives them (O’Brien, 2012; Hill, 2012). That people fail to exercise their rights or do not care about their personal data does not mean that this “apathy” should be “applauded” (Tene & Polonetsky, 2013: p. 263). A very important reason why it should be required that individuals demonstrate greater interest in their data protection, is that control over the processing of personal data enables the data controller to know (Mayer-Schönberger & Cukier, 2013: pp. 50-61; Cohen, 2000: p. 402).

In fact, in the Big Data environment control over the processing of personal data enables any firm to build its own “private networks of knowledge” (Powles & Hodson, 2017). These networks can lead, or perhaps has already led, to the accumulation of power, a power to an unprecedented extent and nature, resting in “private hands”. This power may undermine the fundamental right to equality and non-discrimination (Article 21 of CFREU). As early as in 1993, Gandy
spoke of a digital environment, where databases profiled consumers and sorted them into groups, each of which was given different opportunities (Gandy, 1993). Some years later, other scholars (Gilliom, 2001; Haggerty & Ericson, 2006) built on Gandy’s theory and explained the manners in which new, at the time, tools and datasets were used by governments and private companies alike, so as to sort people and discriminate against them. Today, as these authors argue, private enterprises focus on human beings and study users’ behaviors or movements or desires, so as to “mathematically” predict people’s trustworthiness and calculate each person’s potential as a worker, a criminal or a consumer. “Private algorithms”, which process users’ data, are seen as “weapons of math destruction” that threaten democracy and the universal value of equality (O’Neil, 2016: pp. 2-3, p. 151).

Today’s “free” Internet is paid for mainly by advertising, for the needs of which tons of personal data are collected (Richards & King, 2016: pp. 10-13). Processing of these data with the help of cookies enables firms to identify the user and detect her online or even offline activities (Lam & Larose, 2017; Snyder, 2011). Thereafter, the user’s data are used by private parties, to profile (Article 4(4) of the GDPR) people, to create “target groups”, to which personalized ads may target the correct consumers (Förster & Weish, 2017: p. 19). In the Big Data environment, profiling or sorting consumers into groups may indeed be extremely effective. But the line between a legal sorting and profiling in favor of private interests and an unlawful, as contrary to the principle of equal treatment, discrimination based on personal data collected is blurry (Gandy, 2010; Article 21(1) of CFREU). It is also alarmingly disappearing, as users are being discriminated against on grounds of their personal data, not only during advertising, but in general, while private companies provide any services or just operate, by analyzing the users’ data and “training their machines” (Mantlero, 2016: pp. 239-240; Crawford & Schultz, 2014: pp. 94-95, p. 98; Veale & Binns, 2017; Hastie, Tibshirani, & Friedman, 2009).

Given the correlations that Big Data allows and encourages, any private company that knows, for example, a user’s gender, or her origin or her native language, may discriminate against her (Boyd & Crawford, 2011; Panagopoulos-Koutnatzi, 2017). This can happen by sorting or profiling, not only on the grounds of this information, but also on the grounds of other multiple personal data (Tene & Polonetsky, 2013: p. 240), which the private party may find by combining a huge volume of data, such as the exact address, where the user lives, or even the information that a consumer suffers from diabetes or that she is the mother of three minors (O’Neil, 2016: pp. 3-5, pp. 130-134, p. 151; Rubinstein, 2013: p. 76). Hence, a private company can use these data to create a system that will sort people into lists, put the most promising candidates on top, and “pick” the latter to fill the vacant posts in the company (O’Neil, 2016: pp. 3-5, pp. 130-134).

To sum up, sorting or profiling by “private algorithms”, in favor of private interests and at the expense of people’s fundamental right to equality and
non-discrimination, analyzing and correlating data so as to project the “perfect ad” (See A29DPWP, 2013: p. 46) or promote the “appropriate good” at the “appropriate price” (Turow & McGuigan, 2014; EDPS, 2015: p. 19) or predict criminal behaviors (Chander, 2017: p. 1026) or “evaluate” the accused before sentencing courts (State v. Loomis, 2016), all these actions place almost insurmountable barriers in regulating the processing of personal data (Crawford & Schultz, 2014: p. 106). Knowledge and power seem to be accumulated in the hands of private entities in violation of people’s fundamental rights. Firms may or do dictate “privacy policies and terms of processing of data”, in conjunction with the continuous ticking of boxes with users’ eyes closed (Manovich, 2011). This reality calls for solutions that will enable people to regain control over their personal data—over themselves (Mitrou, 2009).

5. Conclusion

By processing personal data, several economically and socially useful purposes have been achieved (Manovich, 2011; Prins, 2006: pp. 226-230; Knoppers & Thorogood, 2017). The processing of Big Data is even more promising. At the same time, however, the lawfulness of mass-processing of personal data in the Big Data environment is being questioned by many scholars. Although it is very important to examine this lawfulness in each emerging program or software, during the use of which consent is “grabbed by a mouse-click”, it is much more important to understand the real conditions of this personal data processing, which many of us experience every day—or almost all of us experience many times each and every day.

The mass collection of personal data in an environment in which people do not meaningfully participate, in a setting of possibly opaque and discriminatory procedures (to predict, for example, people’s behavior in general via the use of an algorithm, and then apply this prediction to a particular person), should concern all of us deeply. This is especially so, when people cannot know the purpose or even, ignore the very fact of processing and, hence, never give their consent in any meaningful way. The “consent fallacy” (i.e. the inability of the individual-websurfer to form and express free, conscious and informed choices, Mitrou, 2009: p. 466; Mitrou, 2017: p. 77) is accentuated at the highest possible degree. The processing of massive amounts of personal data, in combination with the accumulation of knowledge and power in “private networks” in violation of fundamental right to non-discrimination calls for a new progressive approach to legal provisions that protect personal data, and also, for the development of new technology inserting privacy protection in the very design of information systems dealing with Big Data (De Hert & Papaconstantinou, 2016).

The European legislator with the General Data Protection Regulation made a significant effort to protect people’s rights on their personal data. Simultaneously, firms constantly devise and/or use new technologies of data-processing. This brings back to the discussion table some older academic opinions (Samuelson, 2000) that view commodification of personal data as a potential way, or even the
only way, to regain control (Malgieri, 2016). Such an approach, hotly debated, falls outside the purposes of this paper but will be discussed in our future work.

References


Michaels, J. D. (2008). All the President’s Spies: Private-Public Intelligence Partnerships in the War on Terror. *California Law Review*, 96, 901-966. [https://scholarship.law.berkeley.edu/cgi/viewcontent.cgi?referer=https://www.google.nl/ampredir=1&amp;article=1169&amp;context=californialawreview](https://scholarship.law.berkeley.edu/cgi/viewcontent.cgi?referer=https://www.google.nl/ampredir=1&amp;article=1169&amp;context=californialawreview)


