Rethinking E.U. data protection in the Big Data world

By Alessandro Mantelero

In order to protect the rights of European citizens in a world characterized by technologies that enable to collect, analyse and share an unprecedented amount of data, the new E.U. proposal for a general data protection regulation re-defines the existing legal framework to address the challenges of the digital society.

Although the European proposal shifts the focus of the regulation from users' self-determination to security and risk assessment, the notice and consent are still fundamental aspects of the new framework. Nevertheless, the complexity of data processing and the power of modern analytics drastically limit the awareness of data subjects, their capability to evaluate the various consequences of their choices and their free and informed consent. Moreover, the "transformative" use of personal data makes often impossible to give the description of all the potential uses of the data at the time of their initial collection.

To give an answer to these issues, the paper suggests reconsidering the historical evolution of data protection and the fundamental issues of the related regulations. Starting from the first generation of data protection regulations to the new European proposal, the author suggests a revision of the "notice and consent" model focused on the opt-in. The paper proposes a different approach, based on the opt-out model and on a deeper preventive control by data protection authorities, which should be adopted when the data subject cannot be totally aware of the tools of analysis and their potential effects.

1. Introduction

In the last few years, the debate surrounding data protection and privacy has focused on the future wave of new regulations. Driven by the Web 2.0 environment and the economy of data, private companies and governments have become even more data-centric. However, the high demand for personal information, the complexity of the new tools of analysis and the increasing numbers of sources of data collection, have generated an environment in which the "data barons" [Mayer-Schönberger and Cukier, 2013] (i.e. big companies, government agencies, intermediaries) [Mantelero, 2014] have a control over digital information which is no longer counterbalanced by the user's self-determination. Nevertheless, all the ongoing proposals for a reform of data protection regulations, both in the U.S. and Europe, are still focused on the traditional main pillars of the so called "fourth generation" of data protection laws [Mayer-Schönberger, 1997; Bygrave, 2002], which are represented by the purpose specification an use limitation principles and the "notice and choice" model[1]. This kind of approach seems to be inadequate in the present Big Data context [Cate and Mayer-Schönberger, 2013] and in a digital world characterized by an asymmetric distribution of the control over information. It is also inadequate in a

[1] In the U.S. the traditional approach, based on different sectorial regulations, underestimated the role played by user's choice, adopting a marked-oriented approach; nevertheless the recent guidelines adopted by the U.S. administrations seems to adopt a different approach, reinforcing self-determination, although these new set of principles are still unimplemented. See The White House, 2012, 47-48. On the "notice and choice" regime in U.S., see also Richards and King, 2014.

digital economy where users accept not having an effective negotiation of their personal information, due to market concentration and social and technological lock-ins.[2]

For these reasons, it is necessary to re-consider the existing data protection legal framework and define new models, which better address the various issues of this new digital environment. Different proposals have been advanced by legal scholars, which focus on privacy by design [Cavoukian and Jonas, 2012; Cavoukian and Reed, 2013], contextual privacy [Nissenbaum, 2010], data uses [Cate and Mayer-Schönberger, 2012] and other combined solutions. Nevertheless, many of these proposals adopt a holistic approach to the problem. In contrast, this article suggests the adoption of different solutions for situations in which the role of the consent-based model is outdated and the contexts in which the traditional model based on opt-in can be preserved.

In doing so, the experience from the past should not be forgotten. In many cases, the first answer given by the legal system to new technological and social revolutions [Mayer-Schönberger and Cukier, 2013] is represented by the introduction of new ad hoc rules. Nevertheless, the lack of knowledge of past experiences makes it difficult to find adequate answers to the new questions that technology poses.

In the light of the above, this article reconsiders the history of data protection and its evolution from mainframe to Big Data, in order to give an answer to the contemporary problems of privacy and data protection. This is not a mere cultural interest in this historical perspective, since there are evidently a numbers of similarities between the context of the 50's-60's and the present. For this reason the analysis of that experience can offer elements to address the new challenges and to re-think the data protection framework.

2. The reasons of data protection and the first generations of regulations

Before considering the different reasons that induce the law to protect personal information, it should be noticed that European legal systems do not recognize the same broad notion of the right to privacy, which exists in U.S. case laws. At the same time, data protection laws in the European countries do not draw their origins from the European idea of privacy and its related case law.

With regard to the notion of right to privacy (and in brief), in the U.S. the right to privacy covers a broad area that goes from informational privacy to the right of self-determination in private life decisions [Henkin, 1974; Wacks, 1980a; Wacks, 1980b; Parent, 1983; Zimmerman, 1983; Murphy, 1996]. On the other hand, in European countries this right mainly focuses on the first aspect and is related to the activities of the media.

With regard to the origins of data protection in Europe, it is worth pointing out that the European data protection regulations, since their origins in the late 60's, have focused on the information regarding individuals, without distinguishing between their public or private nature [Costa and Poullet, 2012]. The right to privacy and data protection do not concern the same aspects, even if they are entangled and connected in many senses: there is only a partial overlapping, given that private facts are also referred to individuals, but at the same time a lot of personal information is publicly available and, for this reason, it does not fall into the field of the right to privacy. However, the legal issues related to the protection of personal information had a more recent recognition in law, both in the U.S. and Europe [Schwartz, 2013], dating from the 60's, whereas the primitive era of the right to privacy was at the end

[2] See below para. 4.

of 19th century when the penny press assumed a significant role in limiting the privacy of the people of upper classes [Schudson, 1978]. For these reasons, our analysis should start from the computer revolution of the late 50's and not one century before, when the first decision on informational privacy were adopted in Europe[3], independently from the U.S. legal doctrine and before the milestone article of Warren and Brandeis [Warren and Brandeis, 1890].

The first generations of data protection regulations were characterized by a national approach: regulations were adopted at different times and were different in the extension of the protection they provided (some regulations regarded only the information processed by government agencies, other…..) and the remedies they offered.

The notion of data protection was originally based on the idea of control over information, as demonstrated by the literature of that period [Westin, 1967; Solove, 2008]. At that time, the migration from dusty paper archives to computer memories was a Copernican revolution which, for the first time in history, permitted the aggregation of information about every citizen previously spread over different archives. For this reason, the first regulations represented the answers given by legislators to the rising concern of citizens about social control as the introduction of big mainframe computers gave governments and big companies the opportunity to collect and manage large amount of personal information [Bennett, 1992].

In that period, people were afraid of being visible like a gold fish in a glass bowl [Brenton, 1964; Packard, 1964; Miller, 1971]. In the mainframe era a concentration of information, which was massive for the time, was in the hands of few entities, which were able to support the investments required by the new mainframe equipment. This concentration was also induced by the centralized architecture of mainframes, with their single central processing unit and a main memory in which all the computational power was placed and made available to other specialized terminals, which were connected.

The solution given by the legal systems was the opportunity to have a sort of counter-control over the collected data. The purpose of the regulations was not to spread and democratize power over information but to increase the level of transparency about data processing and guarantee the right to access to information. Citizens felt they were monitored and the law gave them the opportunity to know who controlled them, which kind of data were collected and for which purposes.

Technically speaking, a fundamental element of these new regulations was the mandatory notification to independent authorities of the creation of every new database, necessary in order to know who had control over information. Another key component of the first legal frameworks is the rights to access, which allowed citizens to ask the data owners about the way in which the information was used and, consequently, about their exercise of power over information. Finally, the entire picture was completed by the creation of *ad hoc* public authorities, to guarantee the respect and enforcement of citizen's rights, control over the data owners and the reaction against abuses.

In this model there was no space for individual consent, due to the economic context of that period. The collection of information was mainly made by public entities for purposes related to public interests, so it was mandatory and there was no space of autonomy in terms of negotiation about personal information. At the same time, personal information did not have an economic value for the private

[3] See Trib. civ. Seine, 16 giugno 1858, in *D.P.*, 1858.3.62.

sector, as the data about clients and suppliers were only used for operational functions regarding the execution of the activities of the company.

Nevertheless, there was also another element that contributed to exclude the role of self-determination: the lack of knowledge, the extreme difficulty for ordinary people to understand how the mainframes worked. The computer mainframes were a sort of modern god, with sacral attendants, a selected number of technicians that was able to use this new equipment. In this scenario, it did not make sense to give citizens the chance to choose, since they were unable to understand the way in which the date was processed.

Finally, it is worth pointing out that all these aspects (concentration of information, centralised architecture, complexity of data processing) are now present again in the Big Data context, hence the practical relevance of this past experience, which will be more extensively considered in the following paragraphs.

3. The new generations of regulations and the economic value of personal information

The following period – during the 80's and 90's – was the era of distributed computers: now many people had the chance to buy a personal computer to collect and process information. The big mainframe computer "became" the small desktop personal computer, with a relatively low cost, and consequently the computational capacity was no longer an exclusive privilege of governments and big companies, but became accessible to many other entities and to consumers.

At that time, we also witnessed to another transformation represented by the advent of direct marketing. This was based on customer profiling and necessarily required an extensive data collection to apply data mining software and to suggest to any single consumer the commercial proposal that was most suitable for him or her. This was a new form of data processing driven by new purposes: information was no longer collected to support supply chains, logistics and orders, but to sell the best product to every single user. From this perspective, the data subject became the focus of the process and the information referred to him or her assumes an economic and business value, given its role in sales.

These changes of the technological and business frameworks created new requests from society to legislators: since personal data had become an economic and strategic asset of companies, citizens wanted to have the chance to negotiate their personal data and gain something in return.

Since the new generations of the European data protection laws put the personal information in the context of fundamental rights[4], the main goal of these regulations was to satisfy the economic interest to have a free flow of personal data to boost the economy at large and especially the digital economy. This is also affirmed by the Directive 95/46/EC, which represents at the same time the general

[4] See Council of Europe, Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data, opened for signature on 28 January 1981 and entered into force on 1st October 1985; OECD, Annex to the Recommendation of the Council of 23rd September 1980: Guidelines on the Protection of Privacy and Transborder Flows of Personal Data.

framework and the synthesis of this second wave of data protection laws[5]. Nevertheless the roots of data protection still remained in the ground of personality rights and, for this reason, the European approach is less market-oriented than in other legal systems and it recognizes the fundamental role of public authorities in protecting the individual against unwilled or unfair exploitations of his or her personal information for market purposes.

The theoretical model of fundamental rights, based on self-determination, and the rising data-driven economy highlighted the important role played by user consent in data processing [Schwartz, 2013; Tzanou 2013].[6] Consent is not only the expression of choice about the use of a personality right made by third parties, but is also the instrument to negotiate the economic value of personal information [Acquisti and Grossklags, 2005]. In this new data-driven economy users cannot be put aside and personal data cannot be exploited for business purposes without any involvement of data subjects: it is necessary that data subjects become part of the negotiation, since data are no longer used mainly by government agencies for public purposes, but also by private companies for monetary revenues [OECD, 2013].

Nevertheless, an effective self-determination about data processing, both in terms of protection and economic exploitation of personality rights, cannot be obtained without adequate and prior notice, which describes how the data are processed and the purposes of data processing. For these reasons, the new generation of data protection laws has added a new layer represented by the "notice and consent" model [Kuner, 2012] to the existing paradigm based on transparency and access.

Finally, it is important to notice that during the 80's and 90's the level of data analysis that was possible was still limited and users were able to understand the correlation between data collection and related purposes of data processing (e.g. profiling users to suggest the buying of goods similar to…). So, at that time, informed consent and self-determination were really synonyms; this would not be the case in the future Big Data era.

4. The future generation of regulations in a context characterized by Big Data and big players

The present Bid Data era is different from the previous period both in terms of economic and technological context, with direct consequences on the adequacy of the legal model adopted to protect personal information.

The new environment is mainly digital and is characterized by an increasing concentration of information in the hands of a few entities, both public and private. The role played by specific subjects in the generation of data flows is the principal reason for this concentration. Governments and big private companies collect huge amounts of data while performing their daily activities. This bulk of information represents a strategic and economically relevant asset, since the management of large

[5] The EU Directive 95/46/EC has this bivalent nature, since it was written on the basis of the existing national data protection laws, in order to harmonize them, but at the same time it also provided a new set of rules. See the recitals in the preamble to the Directive 95/46/EC.

[6] See Charter of Fundamental Rights of the European Union (2010/C 83/02), art. 8, OJEU, 30 March 2010, C83/389. See also Productores de Música de España (Promusicae) v Telefónica de España SAU, C-275/06, para. 63 s., online at http://curia.europa.eu/juris/liste.jsf?language=en&jur=C,T,F&num=C-275/06&td=ALL; Bundesverfassungsgericht, 15 Decmber 1983, Neue Juristische Wochenschrift, 1984, 419 ss..

databases enables these entities to assume the role of gatekeepers with regard to the information that can be extracted from the datasets, by limiting access to the data, perhaps to specific subjects only or to circumscribed parts of the entire collection, or by keeping it completely closed.

Not only governments and big private companies acquire this power, but also the intermediaries in information flows (e.g. search engines, Internet providers, credit report agencies, marketing companies), which do not generate information, but play a key role in circulating it.

There are also different cases in which information is accessible to the public, both in raw and processed form. This happens with regard to open data sets made available by government agencies, information held in public registries, data contained in reports, studies and other communications made by private companies and, finally, online user-generated contents, which represent a relevant and increasing portion of the information available online.

The concurrent effect of all these different sources only apparently diminishes the concentration of power over information, as access to information is not equivalent to knowledge [Gurstein, 2011]. A large amount of data creates knowledge if the holders have the adequate interpretation tools to select relevant information, to reorganize it, to place the data in a systematic context and if there are people with the skills to define the design of the research and give an interpretation to the results generated by Big Data analytics [Bollier, 2010; Boyd and Crawford, 2012].

Without these skills, data only produces confusion and less knowledge in the end, with information interpreted in an incomplete or biased way.

In the Big Data context [Mayer-Schonberger and Cukier, 2013; Kallinikos, 2012], the availability of data is not sufficient. It is also necessary to have the adequate human and computing resources to manage it. For this reason, control over information does not only regard limited access data, but can also concern open data, over which the information intermediaries create an added value by means of their instruments of analysis. Because only few entities are able to invest heavily in equipment and research, the dynamics described above enhance the concentration of power over information and are increased by the new expansion of Big Data and its global dimension.

Under many aspects, this new environment resembles the origins of data processing, when in the mainframe era technologies were held by a few entities and data processing was too complex to be understood by data subjects.

Could this suggest that in the future the scenario will change again in a sort of "distributed Big Data analytics", as it happened in the change from mainframe to personal computer? Probably not. The new "data barons" do not base their position only on expensive hardware and software, which may become cheaper in the future. Neither is their position based on the growing number of staff with specific skills and knowledge, able to give an interpretation to the results of data analytics. The fundamental element of the power of "data barons" is represented by the large databases they have. These data silos, considered the goldmine of the 21st century, do not have free access, as they represent the main or the side-effect of the activities realized by their owners, due to the role they play in creating, collecting or managing information.

For this reason, with regard to Big Data, it does not seem so easy to imagine the same process of "democratization" that happened concerning the computer equipment during the 80's: the access to the above mentioned large databases is not only protected by legal rights, but it is also strictly related to the peculiar positions held by the data holders in their market and to the presence of entry barriers.

Another aspect that characterizes and distinguishes this new form of concentration of control over information is given by the nature of the purposes of data collection: data processing is no longer focused on single users (profiling), but it increased by scale and it trying to investigate attitudes and behaviours of large groups and communities, up to entire countries. The consequence of this large scale approach is the return of the fears about social surveillance, which characterized the mainframe era.

Nevertheless, it is important to highlight that this new potentially extensive and pervasive social surveillance differs from the past, since the modern surveillance is no longer realized only by intelligence apparatus, which autonomously collects a huge amount of information through pervasive monitoring systems. It is the result of the interplay between private and public sectors, based on a collaborative model made possible by mandatory disclosure orders, which are issued by courts or administrative bodies, and extended to an undefined pool of voluntary or proactive collaborations from big companies [Council of Europe, 2008]. In this way, governments obtain information with the indirect "co-operation" of the users who probably would not have given the same information to public entities if requested. Service providers for example collect personal data on the base of private agreements (privacy policies) with the consent of the user and for specific purposes [Reidenberg, 2014], but governments exploit this practice by using mandatory orders to obtain the disclosure of this information. This dual mechanism hides from citizens the risk and the dimension of the social control that can be realised by monitoring social networks or other services and using Big Data analytics technologies [See European Parliament, 2013; European Parliament, Directorate General for Internal Policies, Policy Department C: Citizens' Rights and Constitutional Affairs, Civil Liberties, Justice and Home Affairs, 2013a; European Parliament, Directorate General for Internal Policies, Policy Department C: Citizens' Rights and Constitutional Affairs, Civil Liberties, Justice and Home Affairs, 2013b; DARPA, 2002; National].

In this scenario, the traditional data protection framework defined in the 90's goes to crisis [Cate, 2006; Cate and Mayer-Schönberger, 2012; Rubinstein, 2013], since the new technological and economic (i.e. market concentration, social and technological lock-ins) context undermined two of its fundamental pillars: the purpose specification and use limitation principles and the "notice and consent" model.

The purpose specification and use limitation principles have their roots in the first generations of data protection regulations, since they are strictly related to the intention of avoiding extensive data collections, which may imply risks in terms of social surveillance and control. With the advent of the new generation of data protection regulations – during the 80's and 90's –, these principles not only that represented a limit to data processing, but also became a key element of the "notice and choice" model, since they define the use of personal information made by data controllers that is an important information impacting the user's choice. Nevertheless, the advent of Big Data analytics makes it difficult to provide detailed information about the purposes of data processing and the expected outputs. Since Big Data analytics are designed to extract hidden or unpredictable inferences and correlations from datasets, the description of these purposes is becoming more and more "evanescent", as a consequence of the "transformative" [Tene and Polonetsky, 2012] use of Big Data, which makes it often impossible to explain all the possible uses of data at the time of its initial collection.

These critical aspects concerning the purpose specification limitation have a negative impact on the efficiency of the "notice and consent" model. First, the difficulty in defining the expected results of data processing induces introducing generic and vague statements in the notices about the purposes of data collection. Second, also in the hypothesis of the adoption of long and detailed notices, the complexity of data processing in the Big Data environment does not offer to users a real chance to understand it and to make their choice.

Moreover, this scenario is made worse by the economic, social and technological constraints, which definitively undermine the idea of self-determination with regard to personal information that represented the core principle of the generation of data protection regulations approved during the 80's and 90's. As mentioned before, we assisted to an increasing concentration of the informational assets, due to the multinational or global nature of some big players of the new economy, but also due to merger and acquisition processes, which created big companies both in the online and offline markets. In various cases, mainly with regard to online services, these large scale trends drastically limit the number of the companies that provide specific kind of services, which consequently have hundreds of millions of users. This dimension of the dominant players also produces social and technological lock-in effects that increase data concentration and represents further direct and indirect limitations to user's self-determination and choice.[7]

In the described scenario, characterized by complex data processing and concentration of control over information, the decision to maintain a model mainly focused on "notice and choice" represents a risk, since it is easy for companies to give notice and require the consent without an effective self-determination of users, given the above-mentioned reasons.

This leads us to reconsider the role of user's self-determination and to differentiate the situations in which users are not able to understand deeply the data processing and its purposes, or are not in the position to decide,[8] from the other different situations in which they can take an actual, free and aware decision. With regards to the first hypothesis, there it seems to be an analogy between the characters of data processing in the Big Data era and what it happened in the mainframe age: like at the beginnings of computer age, today, data is collected by a limited number of entities and users are not able to understand the purposes and methods of data processing. In these cases the focus cannot be maintained mainly on the user and his or her self-determination: the role of users should be reduced and conversely the role of independent authorities should be increased. Data protection authorities, rather than users, have the technological knowledge to evaluate the risks associated to data processing and can adopt legal remedies to tackle them. Furthermore, they are also in the best position to balance all the different interests of the various stakeholders with regard to extensive projects of data collection and data mining [Article 29 Working Party, 2012].

The suggestion is not to change the entire traditional model of data protection, but to reshape with regard to the Big Data context and the other contexts in which asymmetries in data negotiation

---

[7] A social lock-in effect exists in social networks and it is the consequence of the dominant position held by some big players, which intrinsically limits the user's possibility to recreate the same network elsewhere and, for this reasons, it also reduces user's propensity to change platform and limits their chances of not being profiled or tracked. There is also a technological lock-in, which is related to technological standards and data formats adopted by service providers and it limits the data portability and migration from one service to another.

[8] See also Article 7 (4), Proposal for a regulation of the European Parliament and the Council on the protection of individuals with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation), COM(2012) 11 final, Brussels, 25 January 2012 (hereinafter abbreviated as Proposal), online at http://ec.europa.eu/justice/data-protection/document/review2012/com_2012_11_en.pdf ("Consent shall not provide a legal basis for the processing, where there is a significant imbalance between the position of the data subject and the controller"). In 2013, the Committee on Civil Liberties, Justice and Home Affairs (LIBE) of the European Parliament has dropped the Article 7 (4) of the Proposal of the Commission, see Proposal for a regulation of the European Parliament and of the Council on the protection of individual with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation),(COM(2012)0011 – C7 0025/2012 – 2012/0011(COD)), compromise amendments on Articles 1-29 and on Articles 30-91 (hereinafter abbreviated as PGDPR-LIBE), online at http://www.europarl.europa.eu/meetdocs/2009_2014/documents/libe/dv/comp_am_art_01-29/comp_am_art_01-29en.pdf and http://www.europarl.europa.eu/meetdocs/2009_2014/documents/libe/dv/comp_am_art_30-91/comp_am_art_30-91en.pdf.

drastically reduce the role of self-determination.[9] In the remaining cases, the "notice and consent" model, as traditionally designed, can still be effective, although it needs to be reinforced by increasing transparency, service provider's accountability and data protection-oriented architectures.[10]

5. A subset of rules for Big Data and lock-in situations

The context described above and the related observations suggest defining specific rules for Big Data uses and the situations characterized by asymmetries in data negotiation. The necessity to distinguish this area seems not to be felt neither by the E.U. legislator, in the proposal for a new data protection regulation, nor by the U.S. administration, in the Consumer Privacy Bill of Rights [The White House, 2012]. Although the E.U. proposal provides various rules that can be useful, it still adopts a holistic approach, in which the consent is still "purpose-limited"[11] and based on the "notice and choice" and the opt-in model.

Conversely, legal scholars and companies propose a radical and general different approach, which focuses on the use of the data, on the likely risks (of benefits and harms) associated with the proposed use of the data and on accountability [Cate and Mayer-Schönberger, 2013]. Although this last approach has the undoubted merit to underline the crisis of the traditional model and to suggest a solution more suitable to address the issues of the existing and future context of data processing, nevertheless it offers a holistic solution, but this "one solution fits all" approach does not seem to be consistent with the different contexts.

With regard to Big Data context, the new issues should not be necessarily addressed by making a choice between a "consent based" model and a "corporate accountability" model. Although Big Data and lock-in effects drastically limit self-determination at the moment in which the data are collected, the fundamental right of any person to decide about his or her own information cannot be erased and users should have the right to be informed about data processing and not to take part of it.

In this sense, the model here suggested is the result of the past experiences: like in the first regulations, the decision about data processing cannot be left to users, but at the same time user's rights to oppose to data processing and not to have personal data collected – codified in data protection laws during the 90's – should be preserved.

The fundamental pillars of this model are the adoption of the "opt-out" scheme and the definition of a rigorous data protection assessment, which should be publicly available. With regard to the latter, the same approach that is used in the field of product security and liability (e.g. drugs authorization) should be extended to data processing: in presence of complex data processing systems or data collections influenced by lock-in effects, the risk and benefit assessment should not be done by users, but it should be made by companies, under the supervision of data protection authorities. Users should only decide to give their information or, when the data have already been collected, to exercise or not their right to opt-out.

The most critical element of this model is represented by the criteria which should be adopted in order to define the situations to which the model should be applied or, in other words, to define when the

[9] See below para 5.
[10] See Articles 13a, 23, 32a, 33, 33a, 34, 35, 39, PGDPR-LIBE.
[11] See Article 7 (4) Proposal.

complexity of data processing and the asymmetries between data gatherers and data subjects make inadequate the "notice and choice" model.

On this aspect, the last version of the EU proposal seems not to consider the situations characterized by asymmetries in data negotiation[12], as demonstrated by the erasure of the provision which stated that "consent shall not provide a legal basis for the processing, where there is a significant imbalance between the position of the data subject and the controller". If the provision had been preserved, the above-suggested approach based on prior assessment and opt-out would have represented a possible way to find the alternative legal basis for processing.

On the contrary, the EU Proposal considers the issues related to Big Data analytics and data protection risk analysis, although in an implicit way. In this sense, article 32a does not mention Big Data, but, in listing the cases in which the data protection assessment is required, introduces a new criterion that is represented by the "large scale" of data collection.[13] The EU Proposal requires a prior assessment when there is a large data collection for a long period ("more than 5000 data subjects during any consecutive 12-month period")[14] or there is a "large scale" data collection that involves special categories of information or of data subjects.

Although the Proposal adopts a solution based on risk assessment, this approach diverges from that adopted in this article, which is focused on the use of Big Data analytics and on the complexity of data processing rather than on the dimension of datasets and the nature of information or of data subjects.

The solution proposed by the European legislator is in line with the traditional approach focused on the elements of data processing (nature of the data, categories of data subjects, period of data processing) rather than on the adopted technologies and on the difficulty for the user to be aware of the implications of data processing. Nevertheless, given the predictive nature of Big Data and the consequent impossibility to define ex ante the purposes or the related data processing, it seems to be more adequate, in terms of data protection, to require a mandatory data protection impact assessment in any cases in which these analytics are applied to data sets containing personal information. In this sense, the focus is on the tools used to manage the information and the dimension of data collection only represents a necessary corollary.

Finally, with regard to the dimension of databases, it should also be noted that the border between Big Data archives and normal databases is difficult to define. The criterion of the threshold value of 5.000 data subjects, which has been adopted by the EU Proposal,[15] seems to be questionable: considering the nature of data or data subjects, it may be adequate or not. The use of Big data analytics whit regard to information concerning relatively small groups of subjects with peculiar characteristics (geographical distribution, occupation, similar preferences or behaviours) may have relevant predictive effects and also consequences in terms of discrimination or diversity in the approaches that can be adopted by data processors. For this reason, the a priori definition of a threshold value of data subjects is inadequate and it seems to be better to adopt the notion of "large scale", which is also provided by the EU Proposal. Although it is less definite, it induces a case by case analysis of the existing relationship between the information contained in a database and the total information available about the population to which the data relates.

---

[12] See Art. 7 (4) Proposal.
[13] See Art. 32a (2) LIBE. This parameter is added to the traditional criteria of the nature of the information processed (e.g. sensitive) and the nature of data subjects (e.g. children, employees).
[14] See Art. 32a (2) (a) LIBE.
[15] See Art. 32a (2) (a) LIBE.

In the suggested model, companies intend to adopt a strategy based on Big Data analytics should conduct a prior assessment of its impact on data protection, social surveillance and discrimination, in order to adopt all the adequate measures and standards to reduce it. This assessment, as in clinical trials, should be conducted by third parties and supervised by data protection authorities, which should also define the professional requirements of these third parties. Once the assessment is approved by data protection authorities, the process should be considered secure in terms of protection of personal information and risks of social surveillance or social discrimination and, for this reason, companies can enlist all users in the specific data processing, without any prior consent, but giving them a previous notice that mentions the results of the assessment[16] and providing them the opt-out option.

Obviously the entire system works only if the political and financial autonomy of data protection authorities, both from governments and corporations, is guaranteed. For this reason, it would be preferable if a model based on mandatory fees, paid by companies when they submit their requests of authorization to data protection authorities, could be adopted. This will give autonomous resources to authorities, related to their activities, without being influenced by the entities under their surveillance.

At the same time, in this model, independent authorities assume an important role in balancing all the different implications of data processing, not only in terms of data security but also in terms of social impact and ethical use of data. Conversely, a different assessment exclusively based on the adoption of security standards or corporate self-regulation would not have the same extent and independency. This does not mean that forms of standardization or co-regulation cannot be adopted [Calo, 2013]; nevertheless, the proposed reduction of the role of user's self-determination should have a necessary counterbalance in the active role of public and independent authorities acting in the interest of the whole society.

This model should offer clear and public procedures for assessment. These, undoubtedly, represent an economic burden for companies; nevertheless, in case of positive evaluation of data processing plans, these procedures allow companies to use data for complex and multiple purposes, without the inconvenience of acquiring a specific opt-in choice every time the data are used for new purposes. Companies should only inform users about any changes and give them the chance to opt-out.

From the user's point of view, the assessment conducted by the data protection authorities on one hand gives them a guarantee of an effective evaluation of the risks related to data processing and, on the other hand, the opt-out allows them to receive information about data processing and to decide if they do not want to be part of the data collection.

Finally, it might be noted that the suggested approach based on the opt-out model undermines the chances for the user to negotiate his or her consent and to earn an adequate revenue from data controllers. Nevertheless, the strength of this objection is reduced by the above-described limits to self-determination: in the majority of the cases the negotiation is reduced to the alternative "take it or leave it". From this perspective, in the considered cases, a prior assessment conducted by independent authorities and an opt-out model seem to offer more guarantees to users than an apparent, but inconsistent, self-determination based on "notice and choice" and on the opt-in model.

[16] The notice should also describe how to access to the impact assessment report. This report is a short version of the documentation related to the assessment and it does not contain corporate sensitive information, in order to balance trade secrets and publicity of the assessment. Nevertheless, in presence of litigations, courts or data protection authorities may have access to the complete documentation and may disclose it to the compliant.

Bibliography

1. Acquisti, A., and Grossklags, J. (2005). Privacy and rationality in individual decision making. Security & Privacy, IEEE, 3(1), 26–33.

2. Article 29 Working Party (2012), Letter from the Article 29 Working Party addressed to Google regarding the upcoming change in their privacy policy, online at http://ec.europa.eu/justice/data-protection/article-29/documentation/other-document/files/2012/20120202_letter_google_privacy_policy_en.pdf [accessed 28.02.2014].

3. Bennett C.J. (1992), Regulating Privacy: Data Protection and Public Policy in Europe and the United States, Cornell University Press.

4. Bollier D. (2010), The Promise and Peril of Big Data, The Aspen Institute, online at http://www.aspeninstitute.org/sites/default/files/content/docs/pubs/The_Promise_and_Peril_of_Big_Data.pdf [accessed 28.02.2014].

5. Boyd D., and Crawford, K. (2012), Critical Questions for Big Data: Provocations for a Cultural, Technological, and Scholarly, Information, Communication & Society, 15 (5), 662-679.

6. Breckenridge Carlyle A. (1970), The Right to Privacy, University of Nebraska Press.

7. Brenton M. (1964), The Privacy Invaders, Coward-McCann.

8. Bygrave L.A. (2002), Data Protection Law. Approaching Its Rationale, Logic and Limits, Kluwer Law International.

9. Calo R. (2013), Consumer Subject Review Boards: A Thought Experiment, Stan. L. Rev. Online, 66, 97-102.

10. Carr G. J. (1977), The Law of Electronic Surveillance, Clark Boardman Co., Ltd.

11. Cate F.H., and Mayer-Schönberger, V. (2012), Notice and Consent in a World of Big Data. Microsoft Global Privacy Summit. Summary report and outcomes, online at http://www.microsoft.com/en-au/download/details.aspx?id=35596 [accessed 28.02.2014].

12. Cate F.H., and Mayer-Schönberger, V. (2013), Data Use and Impact. Global Workshop, online at http://cacr.iu.edu/sites/cacr.iu.edu/files/Use_Workshop_Report.pdf [accessed 28.02.2014].

13. Cate H.F. (2006), The Failure of Fair Information Practice Principles, in Winn I., (ed.), Consumer Protection in the Age of the Information Economy, Ashgate, 343–345, online at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1156972 [accessed 28.02.2014].

14. Cavoukian A., and Jonas, J. (2012), Privacy by Design in the Age of Big Data, online at http://privacybydesign.ca/content/uploads/2012/06/pbd-big_data.pdf [accessed 28.02.2014].

15. Cavoukian A., and Reed, D. (2013), Big Privacy: Bridging Big Data and the Personal Data Ecosystem Through Privacy by Design, online at http://www.privacybydesign.ca/content/uploads/2013/12/pbd-big_privacy.pdf [accessed 28.02.2014].

16. Congressional Research Service (2008), CRS Report for Congress. Data Mining and Homeland Security: An Overview, online at www.fas.org/sgp/crs/homesec/RL31798.pdf [accessed 28.02.2014].

17. Costa L., and Poullet, Y. (2012), Privacy and the regulation of 2012, C. L. S. Rev., 28 (3), 254-262.

18. Council of Europe, Convention for the protection of individuals with regard to automatic processing of personal data, opened for signature on 28 January 1981 and entered into force on 1st October 1985, online at http://conventions.coe.int/Treaty/Commun/QueVoulezVous.asp?NT=108&CL=ENG [accessed 28.02.2014].

19. Council of Europe (2008), Guidelines for the cooperation between law enforcement and internet service providers against cybercrime, online at http://www.coe.int/t/informationsociety/documents/Guidelines_cooplaw_ISP_en.pdf [accessed 28.02.2014].

20. Creamer J. S. (1971), A Citizen's Guide to Legal Rights. NY, Holt, Rinehart & Winston.

21. DARPA (2002), Total Information Awareness Program (TIA). System Description Document (SDD), Version 1.1, online at http://epic.org/privacy/profiling/tia/tiasystemdescription.pdf [accessed 28.02.2014].

22. Brion Davis D. (1971), The Fear of Conspiracy Images of Un-American Subversion from the Revolution to the Present, Cornell University Press.

23. Donner, F. (1981), The Age of Surveillance the Aims and Methods of America's Intelligence System, Vintage-Ramdom House.

24. European Parliament, Directorate General for Internal Policies, Policy Department C: Citizens' Rights and Constitutional Affairs, Civil Liberties, Justice and Home Affairs (2013a), The US National Security Agency (NSA) surveillance programmes (PRISM) and Foreign Intelligence Surveillance Act (FISA) activities and their impact on EU citizens, online at http://info.publicintelligence.net/EU-NSA-Surveillance.pdf [accessed 28.02.2014].

25. European Parliament, Directorate General for Internal Policies, Policy Department C: Citizens' Rights and Constitutional Affairs, Civil Liberties, Justice and Home Affairs (2013b), National Programmes for Mass Surveillance of Personal Data in EU Member States and Their Compatibility with EU law, online at http://www.europarl.europa.eu/committees/it/libe/studiesdownload.html?languageDocument=EN&file=98290 [accessed 28.02.2014].

26. European Parliament (2013), Resolution of 4 July 2013 on the US National Security Agency surveillance programme, surveillance bodies in various Member States and their impact on EU citizens' privacy, online at http://www.europarl.europa.eu/sides/getDoc.do?pubRef=-//EP//TEXT+TA+P7-TA-2013-0322+0+DOC+XML+V0//EN [accessed 28.02.2014].

27. Gurstein, M. (2011), Open data: Empowering the empowered of effective data use for everyone?, First Monday, 16 (2), online at http://firstmonday.org/ojs/index.php/fm/article/view/3316/2764 [accessed 28.02.2014].

28. Henkin L. (1974), Privacy and Autonomy, Colum. L. Rev., 74, 1410-1433.

29. Hoffman D., and Halperin, M. H. (1977), Freedom vs. National Security Secrecy & Surveillance, Chelsea House.

30. Kallinikos, J. (2012), The Allure of Big Data, ParisTech Rev., November 16, online at http://www.paristechreview.com [accessed 28.02.2014].

31. Kuner C. (2012), The European Commission's Proposed Data Protection Regulation, Privacy & Sec. L. Rep., 11, 1-15.

32. Long E. (1966), The Intruders, Frederick A. Paeger.

33. Mantelero A. (2014), Social Control, Transparency, and Participation in the Big Data World, Journal of Internet Law, 23-29.

34. Mayer-Schönberger V. (1997), Generational development of data protection in Europe, in Agre P., and Rotenberg , M. (Eds.), Technology and privacy: the new landscape, The MIT Press, 1997, pp. 219-241.

35. Mayer-Schönberger V., and Cukier, K. (2013), Big Data. A Revolution That Will Transform How We Live, Work and Think, Jhon Murray

36. Miller A.R. (1971), The Assault on Privacy Computers, Data Banks, Dossiers, The University of Michigan Press.

37. Morgan R. (1980), Domestic Intelligence, University of Texas Press.

38. Murphy R.S. (1996), Property Rights in Personal Information: An Economic Defense of Privacy, Geo. L. J., 84, 2381-2573.

39. National Research Council (2008), Protecting Individual Privacy in the Struggle Against Terrorists: A Framework for Program Assessment, The National Academies Press.

40. Nissenbaum H. (2010), Privacy in Context. Technology, Policy, and the Integrity of Social Life, Stanford University Press.

41. OECD (2013), Exploring the economics of personal data: a survey of methodologies for measuring monetary value, OECD Digital Economy Papers, No. 220, OECD Publishing.

42. OECD, Annex to the Recommendation of the Council of 23rd September 1980: Guidelines on the protection of privacy and transborder flows of personal data, online at http://www.oecd.org/internet/ieconomy/oecdguidelinesontheprotectionofprivacyandtransborderflowsofpersonaldata.htm#preface [accessed 28.02.2014].

43. Packard V. (1964), The Naked Society, David McKay.

44. Parent W.A. (1983), A New Definition of Privacy for the Law, Law & Phil., 2 (3), 305-338.

45. Reidenberg J. (2014), The Data Surveillance State in the US and Europe, Wake Forest Law Review, forthcoming, online at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2349269#! [accessed 28.02.2014].

46. Richards N.M., and King, J.H. (2014), Big Data Ethics, Wake Forest Law Review forthcoming, online at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2384174 [accessed 28.02.2014].

47. Rubinstein I.S. (2013), Big Data: The End of Privacy or a New Beginning?, International Data Privacy Law, 3 (2), 74-87.

48. Schudson M. (1978), Discovering the News: A Social History of American Newspapers, Basic Books.

49. Schwartz P.M. (2013), The E.U.-US Privacy Collision: A Turn to Institutions and Procedures, Harv. L. Rev., 126, 1969-1992, online at http://www.harvardlawreview.org/media/pdf/vol126_schwartz.pdf [accessed 28.02.2014].

50. Solove J.D. (2008), Understanding Privacy, Harvard Univerity Press.

51. Tene O., and Polonetsky, J. (2012), Privacy in the Age of Big Data: A Time for Big Decisions, Stan. L. Rev. Online, 64, 63-69.

52. The White House (2012), A Consumer Data Privacy in a Networked World: A Framework for Protecting Privacy and Promoting Innovation in the Global Digital Economy, online at http://www.whitehouse.gov/sites/default/files/privacy-final.pdf [accessed 28.02.2014].

53. Theoharis, A.G. (1978), Spying on Americans. Philadelphia, Temple University Press.

54. Tzanou M. (2013), Data protection as a fundamental right next to privacy? 'Reconstructing' a not so new right, International Data Privacy Law, 3 (2), 88-99.

55. Wacks R. (1980a), The Poverty of 'Privacy', Law Quarterly Review, 96, 77-89.

56. Wacks R. (1980b), The protection of Privacy, Sweet & Maxwell.

57. Warren S.D., and Brandeis, L.D (1890), The Right to Privacy, Harv. L. Rev., 4, 193-220.

58. Westin A.F. (1967), Privacy and Freedom, Atheneum.

59. Wise D. (1976), The American Police State the Government Against the People, Ramdom House.

60. Zimmerman L.D. (1983), Requiem for a Heavyweight: A Farewell to Warren and Brandeis's Privacy Tort, Cornell L.Rev., 68, 296-367.