

Masters of Big Data: concentration of power over digital information

Alessandro Mantelero*

Politecnico di Torino

ABSTRACT

In our society, information has assumed a fundamental role in every decisional and social process. Governments and big private companies collect huge amounts of data that represent a strategic and economically relevant asset. The predictive relevance of big data management and the global dimension of this phenomenon has led us to reflect on the nature and the dynamics of a centralized power held by only a few subjects. The increasing power that derives from big data necessitates the adoption of adequate remedies to control and limit the information asymmetries and their consequences in terms of economic advantages and social control. From this perspective it is important to adopt adequate measures to control those who have this power, in order to limit possible abuse and illegitimate advantages, and, at the same time, to increase access to information, in order to spread informational power.

* Alessandro Mantelero is Assistant Professor (tenured position) of Commercial Law at the Politecnico di Torino and Faculty Fellow at the Nexa Center for Internet and Society.

Introduction

The current era is considered to be the information age, in which information has assumed a fundamental role in every decisional and social process. This definition does not reflect an unchanging framework since the nature of information changes over time, due to a different distribution, size and means for effective analysis. With regard to the distribution of information, not everyone has access to all sources of information and not everyone has the ability or the instruments to exploit them. This situation leads to the concentration of power over information in the hands of few subjects.

In terms of accessibility the principal reason why large amounts of data are available only in the hands of a limited number of subjects is mainly because of the role the latter play in the generation of the related data flows. Governments and big private companies, while performing their daily activities, collect huge amounts of data that represent a strategic and economically relevant asset. This asset is then managed in an exclusive way by denying third parties access to the whole set of raw data, by restricting access only to certain subjects or by releasing only selected data without giving access to raw data.

Not only the public or private entities which collect and manage data acquire a relevant informational power, but also the subjects that assume the role of intermediaries in information flows, which do not generate information, but play a key role in circulating it. This gives them access to large amounts of data. A clear example of this category are the Internet providers, which can manage, monitor and analyse a huge flow of data generated by third parties. The same role is assumed,

A. Mantelero © 2012

off-line, by companies collecting information from third parties in order to create large databases to deliver information services (e.g. credit rating services, market analysis). As in the first case, informational power is concentrated in the hands of a limited number of intermediaries, who also have the power to define who can access this information, often in exchange for economic benefit.

Finally there are cases in which the information is accessible, both in raw and processed form; some examples are open data (sets of data made available to the public by the government); private data made available under a legal obligation; studies and reports generated by many different entities; user-generated-content available online. Even if all these sources of information give the impression of a significant amount of data available, the excess of information that is generated can produce the effect of an overdose of information that has the opposite result to induce a decrease in knowledge. Information is not equivalent to knowledge, a large amount of data creates knowledge if those who manage it have the adequate interpretative keys to select relevant information, to reorganize it and to place the data in a systematic context. Without these skills data only produce confusion and less knowledge in the end, due to incomplete or biased information. This situation enhances the key role of experts in different fields, of opinion leaders and of people who offer their knowledge to organize and de-codify the overload of data thus giving non experts a conceptual map to guide them through the flow of information. Consequently a new category of person emerges: who holds informational power, related to their role of cultural intermediaries. They can use this power in an independent way but they can also be conditioned or oriented towards interpretation

which could be useful to third parties by reason of common and not declared interests.

This rapid overview shows that the ability to analyse data is fundamental in the big data context¹ and it is related to the role that the subject assumes in the data flow and the technical and human resources available for the analysis of the information. It is possible to dominate the world of big data using the human brain, or by means of the aggregation of experts in a think-tank, or using a computer brain and artificial intelligence which are able to process large amounts of data automatically and to extract new knowledge. In both case, if we exclude people who offer their knowledge to others for narcissism or altruism, the best human and computer brains work for a few entities who are able to invest heavily in equipment and research.

Finally we should consider a further psychological element that increases the power owned by the subjects who manage big data: recent studies have demonstrated that the large amounts of data available on-line induce people not to

¹ See The Aspen Institute, “The Promise and Peril of Big Data”, David Bollier Rapporteur (Washington, 2010), available at <http://www.aspeninstitute.org/publications/promise-peril-big-data> [Accessed January 4, 2012]. See also: OECD, Joint WPISP-WPIE Roundtable, “The Economics of Personal Data and Privacy: 30 Years after the OECD Privacy Guidelines” (Paris, December 1, 2010) available at http://www.oecd.org/document/22/0,3746,en_2649_34255_46565782_1_1_1_1,00.html [Accessed January 4, 2012]; McKinsey Global Institute, “Big data: The next frontier for innovation, competition, and productivity” (June 2011) available at http://www.mckinsey.com/mgi/publications/big_data/ [Accessed January 4, 2012]; Boyd-Crawford, “Six Provocations for Big Data”, available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1926431 [Accessed January 4, 2012].

memorize the information they have revived, but the location in which the information is available². In this way the relevance assumed by the position of the information rather than the information itself increases the power of the gatekeeper of the data.

Collective knowledge is developing in the digital interconnected world and a relevant part of our cultural background is migrating to the on-line context, but at the same time the consequence is the need to use the services offered by those who manage the information or assume the role of intermediary in the information retrieval process.

The power of the masters of data

The concentration of power over information is not new or a recent phenomenon, but the new expansion of big data, and its global dimension, leads us to reflect on the nature and the dynamics of a centralized power held by only a few subjects.

The predictive ability is the fundamental element that distinguishes this new kind of power from the ordinary power to acquire a large number of data or profile millions of people³. The size of datasets makes it possible to draw inferences about unknown facts from statistical occurrence and correlation, with results that are relevant in socio-political, strategical and commercial terms. Even if this method of

2 See B. Sparrow, J. Liu and D.M. Wegner, “Google Effects on Memory: Cognitive Consequences of Having Information at Our Fingertips” (July 14, 2011), available at www.sciencemag.org [Accessed February 15, 2012].

3 See The Aspen Institute, *The Promise and Peril of Big Data*, David Bollier *Rapporteur* (Washington, 2010), p. 12, see fn. 1.

analysis, focused on trends, is different from the more scientific and rigorous approach based on surveys of representative samples and even if its results could perhaps be biased, the same method is able to bring unexpected results. Examination of the data flows in the evolving datasets shows trends and information without the need of prior working hypothesis, changing the traditional paradigm of social analysis in which the design of the study sample represents the first step which is then followed by the analysis of raw data. Traditional sampling method is limited because it tends to infer or imagine the relevant aspects to be examined, so unexpected tendencies may not be considered. On the contrary with the big data approach new trends are self-explanatory, though it does involve a degree of statistical error. In order to obtain significant results this inductive method requires the largest amount of information possible, as it is not possible to select a representative sample in an analysis that does not start out with a precise target of study. Unlike the traditional method the study of big data is focused on tendencies which emerges from collected information.

Despite the weakness of this approach, more focused on correlation than on statistical evidence, it is useful to predict and perceive the birth and evolution of macro-trends, that can be later analysed in a more traditional statistical way in order to identify their causes⁴. These results concerning the emerging tendencies, even without knowing the causes that generated them, give the owner of big data a

4 See C. Anderson, "The End of Theory, Will the Data Deluge Make the Scientific Method Obsolete" (March 26, 2008), available at http://www.wired.com/science/discoveries/magazine/16-07/pb_theory [Accessed January 16, 2012].

competitive advantage in the business world and increase the possibility of social control by states and groups of power.

Limited access to this information also has another consequence in terms of increasing their economic value. In a society based on information, that requires large amounts of data in order to take decisions, having large datasets and the instruments to extract predictive knowledge is a fundamental asset which can be exploited with a big payback. From this perspective it emerges that the availability of the data is not sufficient, but it is also necessary to have the adequate computing resources to manage it. Without the latter more data available will not generate more knowledge, but only more confusion. The overflow of information without the adequate instruments to organize it, creates nothing but noise and, in the end, less knowledge. For this reason power over information, in the big data era, is not only related to data characterized by a limited access, but it can also concern open data, over which the information intermediaries create an added value by means of their instruments of analysis.

Finally there is another element characterizing the power of the masters of big data, that of being geo-localized in terms of the nationality of the owners of the datasets and of the place in which these big data farms are placed. In many cases these two geographic elements do not coincide, as, especially when the master of data is a private company, the owner of the datasets is an Western company, while the data is stored in data-centres in countries where the operating cost are lower, such as in Asia or in emerging economies.

Geo-location has several relevant consequences, not only with regard to the

applicable law and the level of data protection, but also in terms of political influence exerted by local governments. In many cases national laws give judicial and administrative bodies power to access databases which can compromise data protection, trade secrets and public interests (when the data stored is related to governmental activities).

There is also another aspect that is relevant from an E.U. perspective, related to the geographical concentration of power over information: IT (information technology) is the sector which is most involved in big data management and the big players in this market are U.S. companies, therefore there is an asymmetric distribution of power over data between the U.S. and the E.U. This situation implies evident risks in terms of service continuity, competition, lock-in, and moreover can represent a possible threat to E.U. interests due to the influence of the U.S. legislative and political context.

For these reasons the recent E.U. draft of new data protection regulation has adopted the strategy to protect personal data of European citizens wherever they are, even outside E.U. borders⁵. Since the E.U. does not have an IT sector strongly

5 See Article 3(2) of the Proposal for a Regulation 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, available at http://ec.europa.eu/justice/data-protection/document/review2012/com_2012_11_en.pdf [Accessed February 13, 2012]. See also V. Reding, “Your data, your rights: Safeguarding your privacy in a connected world”, SPEECH/11/183 (Brussels, March 16, 2011), available at <http://europa.eu/rapid/pressReleasesAction.do?reference=SPEECH/11/183&format=PDF&aged=0&language=EN&guiLanguage=it> [Accessed

involved in big data management and it can not exert political influence over foreign companies, European legislators try to use legal instrument in order to limit the risks of unauthorized access and fraudulent use of data. The intent of E.U. regulators is to influence development and business strategies of the masters of data in order to compel third countries to reach a compromise on the guarantee offered with regard to data protection in general. This is an ambitious goal, based on the consciousness of the importance of international data flows and of the effects of the introduction of legal barriers. The E.U. acted in the same way when the general Directive on data protection was introduced in 1995 and imposed its model on other countries. After more than fifteen years the framework has changed a little. On one hand there is greater consciousness of the importance of data protection among people all around the world, but on the other hand the U.S. seems not to be interested in an uncritical acceptance of the E.U. model and aims to assume a leadership role in data protection regulation⁶. In this context the theme of big data and its related power can be one of the key elements in the global conflict between legal models on data protection, involving not only the U.E. and the U.S., but also other countries like India and

January 4, 2012], and V. Reding, “Privacy matters – Why the EU needs new personal data protection rules”, SPEECH/10/700 (Brussels, November 30, 2010), available at <http://europa.eu/rapid/pressReleasesAction.do?reference=SPEECH/10/700> [Accessed January 4, 2012].

6 See The Department of Commerce Internet Policy Task Force, “Commercial Data Privacy and Innovation in the Internet Economy: A Dynamic Policy Framework” (December 2010), available at <http://www.commerce.gov/sites/default/files/documents/2010/december/iptf-privacy-green-paper.pdf> [Accessed January 7, 2012].

China where various activities concerning IT are outsourced.

The possible limits to the power of the masters of the data: control, access, competition

The increasing power that derives from big data necessitates the adoption of adequate remedies to control and limit the information asymmetries and their consequences in terms of economic advantages and social control.

The draconian idea of indiscriminately deleting all information collected when a given period of time has elapsed does not seem the best solution. This obligation to erase data is sometimes considered a form of the more ancient *droit d'oubli* (right to be forgotten), but is not correct, because this right prevents the media revisiting past events which are now long forgotten within the social group. In this way the law ensures that some episodes of the individual's life are not recalled, permitting the individual to affirm their right of self-determination in order to make a new life after being involved in a scandal or crime⁷. Although the right to be forgotten is related to the privacy of individuals, it concerns events made public that are not of a confidential nature; for this reason the right to be forgotten can not be invoked in relation to private affairs that have not been the subject of previous disclosure. This legal protection does not apply equally to all the information concerning an individual's past life, but only to the information that has been the subject of media attention. The difference between this right and the right to delete the data is evident

⁷ See *Melvin v. Reid* [1931] 112 Cal. App. 285, 297 Pac. 91, e *Sidis v. F-R Publishing Corporation* [1940] 113 F.2d 806 (2d Cir.).

and related to the different context in which the later is exercised. The erasure of information in order not to make it available after a certain period of time is not related to the idea of oblivion and the need to remove the memory of past negative events, because it does not distinguish between what is properly protected by the droit d'oubli and what it is only characterized by having been collected a long time ago.

A generalized mandatory erasure of data not only constitutes a misunderstanding of the original nature of right to be forgotten, but is also technically difficult and onerous to implement and, finally, will drastically limit the collective memory that the Internet represents. With regards to the large databases on which the phenomenon of big data is based, this remedy is also inefficient due to the difficulty of verifying the systematic and selective erasure of large amounts of data. It should also be considered that the rules concerning personal data protection only impact with a limited area of the entire amount of information managed by the masters of big data, because in many cases their power derives from the management of databases that do not refer to identified or identifiable subjects⁸, even if they have a predictive relevance.

For these reasons we need to identify different solutions in order to limit the power of the masters of big data and obtain a better allocation of informational power. To achieve this goal it is important to adopt adequate measures to control those who have this power, in order to limit possible abuse and illegitimate

⁸ See Article 2 (a) of the Directive 1995/46 on the protection of individuals with regard to the processing of personal data and on the free movement of such data [1995] OJ L281/31.

advantages, and at the same time to increase access to the information, in order to increase the number of subjects able to create and manage large amounts of data and to spread the informational power currently in the hands of a few bodies.

The need to control these great aggregations of data is also related to their political and strategic relevance and should lead the creation of specific international independent authorities and the introduction of a mandatory notification of the creation of a big database, as happened at the beginning of the computer age when there was a similar concentration of power in the hands of a few subjects due to the high cost of the first mainframes.

These authorities will be able to control the invasive attitude of governmental power with regard to large databases and the power of the owner of big data, but can also have an important role in the definition of specific standards for data security.

This approach may be considered in contrast with the spontaneous evolution of informative and business processes and also be considered as an undue interference of the State in the economic system, but in the information society, bearing in mind the value assumed by the data, a concentration of large amounts of data should be controlled, as happens with all the relevant and strategic assets.

This will be a long and tortuous journey, as it is based on international cooperation, nevertheless it is important to start it as soon as possible, using the existing international bodies, rather than introducing new regulations when it is too late. At the same time, any solutions should be graduated in an appropriate manner, avoiding the involvement of every kind of data-farm built somewhere in the world, but considering only the data-farms with an absolutely remarkable dimension

or a considerable importance because of the data collected (e.g. police or military databases).

Another important effect of these policies concerns the transparency of the information society, in terms of the capacity to know those who have great informational power due to big data management. This knowledge is fundamental to act on the other central aspects, namely the access to data and the data sharing, in order to limit the power of the masters of the big data and give society the opportunity to have access to that fundamental resource which information is. From this prospective, a key role is played by open data. Opening public databases to citizens and giving them raw data not only reduces the power of the owners of information, in terms of the exclusive access to the data, but also limits their advantage in terms of technical and cultural analysis skills. The access to data does not mean that everyone will immediately have new knowledge and predictive capacity, because, as mentioned above, technical equipment is necessary. However, the availability of the data permits citizens to put together their economic and cultural resources, even without a business oriented action, in order to constitute groups dedicated to the analysis and processing of the raw data. In this way alternative and autonomous centres focussed on big data management could be created. They will take a different and longer amount of time to achieve the same results that the masters of big data are able to obtain in a short time, thanks to their resources, but for some research with civic relevance the speed of obtaining the result is less important than the result itself (e.g. fight against corruption by analysing the relationship between people and institutions).

The final critical issue regarding big data concerns the geopolitical distribution of informational power, which represents an emerging problem for Europe. From this perspective the comparison with the U.S. assumes relevant importance, because in the U.S. public sector not only is there a structural homogeneity still unknown to the E.U., but also many resources for modernizing the management of information through the use of cloud computing technologies have been invested, creating a limited number of big databases. Considering the U.S. private sector, even if big European companies are able to collect and analyse a large amount of data, the excellence of the U.S. companies in some key areas of the information and communication technology sector (search engines, cloud computing services, user-generated-content platforms, social networks) puts this nation in a better position to control the world's informational flows, since the largest stream of data is present in the areas most closely related to ICT.

From a geo-political and industrial perspective, this situation may be a weakness for the E.U., in terms of the loss of control over the data of its citizens due to the need to entrust the management of strategic information to foreign entities. In order to reduce this risk, the E.U. has strengthened the protection of personal data by applying the European data protection law when goods or services are offered to “data subjects in the Union”⁹, or when their behaviour is monitored. At the same time, European industry is being urged by the European Commission to assume a

9 Article 3 (2) (a) of the Proposal for a Regulation 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, see fn. 5.

more important role in the implementation of new computer architectures, such as cloud computing¹⁰.

At this moment in time, stimulating competition in the development of new technologies in the ICT sector and strengthening the regulatory framework for the protection of information are the only possible responses in order to limit the potentially negative effects due to the use of services provided by foreign companies, giving the control of the (big) data concerning European citizens to companies that may potentially be influenced by legal and political foreign interests.

¹⁰ See N. Kroes, “The Digital Agenda: Europe's key driver of growth and innovation”, SPEECH/11/629 (Brussels, October 4, 2011), p. 4, available at <http://europa.eu/rapid/pressReleasesAction.do?reference=SPEECH/11/629&format=HTML&aged=1&language=EN&guiLanguage=en> [Accessed February 13, 2012].