Big Data: the Beauty or the Beast

Camelia CRIȘAN[1],
Alexandra ZBUCHEA[2]
Steliana MORARU[3]

Abstract. *Big Data is a phenomenon that has been made possible by the IT and the social media revolutions - where content is created or generated by users and their interaction, at the same time with the exponential increase of the data storage capacity, according to the Moore's law. It has been a long-time dream of social scientists to investigate an issue of importance for large groups of people where n - the number of the investigated subjects - is not determined by some statistical complex formula, but rather by mentioning n=all. This would allow for better results, with wider applicability in the attempts to understand the society, its trends, ideas and how they propagate, as well as the capacity of taking more efficient decisions that concern purchase, education, health and politics. But what are the costs? Our paper aims at looking at means and ways through which Big Data is being generated, to provide examples of Big Data ownership and consequences derived from this, and to illustrate the use of Big Data for improving the life of the society's members. We define the Big Data, how it is generated, processed and the degrees of responsibility in maneuvering such precious resource. At the same time, our focus is on the backside of accumulating large amounts of personal information. We evaluate how and if major companies are handling Big Data properly - from disclosing information about gathering such data, processing it and using it to their own profit, with the informed consent of the subjects. In our research we discuss potential implications from the perspective of redefining what personal and private still means when individual data becomes a commodity.*

Keywords: *Big Data, data analysis, privacy, business ethics.*

[1] Lecturer, Ph.D., College of Communication and Public Relations, National University for Political Studies and Public Administration, camelia.crisan@comunicare.ro.

[2] Associate Professor, Ph.D., College of Management, National University for Political Studies and Public Administration, alexandra.zbuchea@facultateademanagement.ro.

[3] Ph.D. Candidate, College of Communication and Public Relations, National University for Political Studies and Public Administration, steliana.moraru@gmail.com.

Big Data is everywhere. Big Data is upon us. We live in the age of „big data". The era of Big Data has begun (Boyd & Crawford, 2014; Tene & Polonetsky, 2013). Or should we say: „Welcome to Big Data. Welcome to the end of computing as we have known it for 70 years" (Needham, 2013). All the above points of view are conveying one thing - we are witnessing a revolution in the way information is being gathered, stored and processed. From each Internet login, from each app usage, from each shopping item bought online, from the sensors of our engines millions and millions of pieces of information are being generated every minute. Such data does not only need huge capacity to be stored, but what researchers have recently found is that processing it, brings about patterns and correlations that are affecting large amounts of people or can increase the innovation potential of companies. For instance, in 2009 Google was able to track the expansion of pig flu epidemic by following searches for flu related topics. It did this two weeks before the US Center for Disease Control (Loukides, 2011). In another example, a computer scientist, Oren Etzioni, aggregating open data offered by airline companies, has set up a web search engine allowing future passengers to buy plane tickets at the best timing, for the best price (Mayer-Schonberger & Cukier, 2013). Even if it is about big or small results, big data analyses have allowed people make better, more informed decisions and as a result, their lives changed for better.

How big is Big Data?

‚Sangameswar (2013) says that Big Data refers to data of massive scale and complexity. If one unit of data is measured by a byte, the data stored in the world reached as of 2012, about 2.5 exabytes of data and that number is doubling every 40 months or so (McAffee & Brynjolfsson, 2012). This means 10^{18} bytes, while the largest measurement unit for data storage goes up to yottabytes, which is 10^{24} bytes. That is 10 followed by 24 zeros. Mayer-Schonberger and Cukier (2013) give a significant example related to amassing data in astronomy. It's about the Sloan Digital Sky Survey and its telescope in New Mexico, which has been collecting more data in a few weeks than it has been collected in the entire history of astronomy - 140 terabytes of information by 2010. But not all data gathered can be considered Big Data.

Boyd and Crawford (2014, p. 663) define Big Data as: "a cultural, technological and scholarly phenomenon that rests on the interplay of: technology, analysis and mythology". In their view Big Data rests on computer power, its analysis determine patterns which generate knowledge and insights that one could not have foreseen previously „with an aura of truth, objectivity and accuracy". Sangameswar (2013) defines more clearly the type of information that are part of Big Data: traditional enterprise data (customer information, web store transactions, etc.), machine generated and sensor data, weblogs, equipment logs and social data, including customer feedback streams, micro-blogging. McAffee and Brynjolfsson (2012, p. 63) say that Big Data has three types of characteristics: volume (which has been detailed above), variety (messages, updates, and images posted to social networks; readings from sensors; GPS signals from cell phones, etc.) and velocity (information is generated in real-time or nearly real-time which allows a company to be much more faster than its competitors).

As a result of these definitions, we understand that not any data gathered and analyzed by companies could be labeled as Big Data, but this title applies to all those cases where mass information is generated from a variety of sources at high rate. An important feature of Big Data is its messiness. It means, according to Mayer-Schonberger and Cukier (2013, p. 39) that „more trumps better", and thus a research where potentially the number of respondents equals the total of the researched population (n=all) should decrease for this reason its obsession for exactitude. The more information we add, the higher the potential for errors within data, as well as consistency for data formatting and combining various types of data. However, as the two authors put it, quoting Hopkins and Evelson (2011) "sometimes 2+2 equal 3.9 and that is good enough". This does not mean that the data is incorrect, just that when we whiteness for instance 1000 tweets per second it makes more sense to show tolerance for error rather than aim for clockwork precision. This reality is then transferred to the way data is analyzed.

Davenport and Patil (2012) claim that a new job - data scientist - is the sexiest job of the 21st century. They start their argument from presenting the case of a PhD graduate from Stanford which brought LinkedIn to the success it is today, just because his data analysis showed that people could develop their networks easier if they follow machine based algorithms in finding people they could be matched with, based on the information they have shared in their profile. A similar example of using data analysis is shared by Mayer-Schonberger and Cukier (2013) when they describe the success Amazon.com had, when replacing the comments and recommendation of professional reviewers to items customers may be interested it, based on books purchased from the same domain by other people who checked out certain item. It was all based on how the „traces" left by different users have been processed and analyzed to understand a pattern. The important issue here is that instead of trying to understand the Why, what is the Cause which determined purchases based on people preferences rather than expert recommendations, the companies were satisfied with identifying the pattern, and were not looking for the explanation of the pattern. In sociological analysis this equals with finding a correlation between two phenomena / variables. It means that the change in one goes along with the change in the other, but it is not necessarily determining it. Such types of results based on data processing are allowing companies to extract added value and innovate. It appears not only that Big Data is omnipresent, but also using and processing it is a highly economically viable option.

The detractors of "Big Data conquers all" position express, in our view, 3 main areas of concern: quality of data analysis, compensation for personal data usage, protection of privacy and intimacy. Big Data does not necessary mean better data or scientifically sound data, which could lead to scientifically sound research and thus quality of knowledge (Boyd & Crawford, 2014). Companies storing people's data should made them aware that such data may be used for economic purposes and, as a result, pay them in return for using their data (Buck, Horbel, Kessler & Germelmann, 2014). To the same extent, people should be made aware or educated to become more careful that the free usage of some apps in return to their personal data needed to install them may be a bad bargain for them. In terms of privacy and intimacy, we will refer to this by large in the next chapter, however, it is worth mentioning here that realizing at some point that your personal data is available to potentially anyone paying a good price to sell you something, or that a Big Brother can follow your every move tend to cast a shadow of fear and adversity towards companies for which we, sometimes non intentionally, allow access to our private online life.

The backside of Big Data from the individual and society perspective - breaching the right to intimacy

As mentioned above, the benefits of using Big Data are largely recognized, both at society level / macro-level policies (Bollier, 2010; Chen, Chiang & Storey, 2012; Gehrke, 2012; Lohr, 2012; Whitepaper, 2012), as well as at business-level strategies (Bollier, 2010; Lohr, 2012; McAffee & Brynjolfsson, 2012; Russom, 2011). Using Big Data has also downsides. For instance, it could be deceptive and could lead to false findings, either deliberately or unconsciously (Lohr, 2012; Yetiskin, 2014). Interpretation of Big Data is sensitive in several ways to biases (Bollier, 2010). In this context not just honesty in dealing with and analyzing Big Data is important, but also qualified work force is necessary. There is an increased demand for specialized analysts (Lohr, 2012; McAffee & Brynjolfsson, 2012), as well as for a new managerial approach (McAffee & Brynjolfsson, 2012; Yetiskin, 2014).

In business context, one of the most debated Big Data related issues is the privacy of consumers. Laurila et al. (2012) consider that "protecting privacy of individuals behind the data is obviously the key reason for access and usage limitations of Big Data". Respecting the right to privacy of the consumers and stakeholders is not just a matter of ethics but also a matter of good business. Companies have to consider not just the legislation, but also the requirements of the wider public to benefit of privacy and respect in their relationships with businesses in order to be trusted and preferred to their competition.

Agreeing that the main three characteristics of big data are volume, velocity and diversity (McAfee & Brynjolfsson, 2012; Whitepaper, 2012; Russom, 2011), we would like to add to these, a relevant forth one: personal character. Big Data is intimately related with individuals, comprising in many instances sensitive personal and financial information. Therefore, privacy issues are extremely important to consider when acquiring, storing, processing, analyzing and using Big Data. This is proved by the interest of governments to regulate this field, as well as numerous public scandals and consumer taking of stand in this respect.

Researchers investigating the regulation of big data in various countries tend to agree that the European Union has the amplest legislative system but none is comprehensive (Asay, 2013; Guo, 2012; MacDermott & Smith, 2013). The main difference between the European and American approaches is that the first imposes tight governmental regulations, while the other lets the industry self-regulate and gives customers more liberty to decide. Data Protection in the EU involves high standards; only legitimate data collection is allowed under tight security. It sets criteria to be considered. There is an agreement between the EU and the US to comply with the European standards, but just a few American companies have accepted the terms (Guo, 2012). In the US, consumers have the choice to control their information and protect their privacy. The weak aspect is that people do not generally read privacy notices or they do not understand them (Asay, 2013). This, corroborated with some debatable aspects of the legislation in the US, could generate privacy breaches or other problems in handling Big Data (Asay, 2013).

Big Data management has to consider several complex privacy-related aspects. The main points of reference would be: the scale and the aims of the collector; the media used to gather information.

Big Data could be collected by businesses of all sorts, by regional / national collectors and public organizations (open data). Especially in the last case data sets are shared for the benefit of a wider public, while personal information is protected. Still privacy breach could occur (Gehrke, 2012).

Research of Big Data, even in academic context, involves sensitive issues, especially privacy-related ones (Laurila et al., 2014). Three main aspects are to be considered: data security, data anonymization and the respect of privacy by the researchers. If the data is specifically collected for the research, consent from the participants / subjects of the investigation also has to be secured. In order to share the results of the research, privacy during the entire research flow is a must.

Special privacy issues could emerge also when using Big Data for the benefit of larger communities. Many discussions are related, for instance, with the use of Big Data in health care in the US (Bollier, 2010; Groves et al., 2013). Privacy is a key-factor in the process of sharing vital data, as well as to investigate the data corpus available. Other relevant aspects opposing the use of integrated Big Data in healthcare are related with the interests of various actors, and with various ethical aspects (Boillier, 2010).

Big Data is collected using various channels: offline/administrative, online and mobile. The Internet and mobile phones are increasingly more challenging due to their dynamics. The Internet is the one that changed the way information is dealt with and generated the use and the research on big data. More recently, the development of smartphones offers new type of information and databases (mobile big data), as well as new challenges. This refers to the need to manage large-scale information generated by the use of smartphones, including online and application use. This type of data allows "understanding real-life phenomena, including individual traits, as well as human mobility, communication, and interaction patterns" (Laurila et al., 2012)

The first sensitive issue is to decide what information to collect. It is not just a matter of management – of having significant information, but also of ethics – of justifying the storage of that specific information. An additional ethical and legal aspect is to obtain the approval of each individual to store, manipulate and use that information. In many cases the information is not used only by the organization that obtained it, but also by its associates. The transmission of data to third-bodies is also highly sensitive (Asay, 2013). Not just consumers do not have control over this information, but companies themselves loose the control.

One of the privacy breaches is the identity theft. It can occur in many forms, as it widely means the unauthorized use of information (MacDermott & Smith, 2013). 10% of the US online consumers were victim of an identity theft (MacDermott & Smith, 2013). Some artists and hacktivists draw the attention on the perils associated with Big Data wrongful handling, in order to make people aware of the sensitivity of the information they share online (Yetiskin, 2014) and how social reality can be manipulated. A severe privacy issue is the phishing phenomenon, since it involves in many cases disclosure and subsequent use of financial information.

A frequent aspect involving ethical aspects related with privacy violation is the use of cookies[4]. Many of the popular websites use cookies to track their visitors, some of them permanently not only while on their website. At least to a certain degree people know and accept this if they are interested in those websites. Nevertheless most of them are uneasy with the idea of being tracked by advertising (Bollier, 2010).

Even if privacy seems to be a hot topic in the context of Big Data and the Internet / mobile environment, the tendency overall, both considering businesses and governments is the growing control of consumers, as well as of citizens (Yetiskin, 2014). Organizations and individuals are caught between ethics and business/politics. Some delicate situations may arise. Sometimes companies take the ethical stand, but for whose benefit? For instance, Facebook protected its users from intrusion and loss of privacy against their employers, but, in fact, the company protected itself from future damage and lack of trust (MacDemott, 2013).

The trust of consumers in companies is an issue of prime importance. Big Data could be affected by the lack of trust. For instance consumers and individuals could provide partial or false information, so data will be from the beginning corrupted (Gehrke, 2012). Trust in online transaction would also influence the online shopping behavior. Other issues to be considered in this context are online (perceived) security system, information scanning, recommendation / review system, credibility, and virtual experience (Fang & Li, 2014). The shopping and searching behavior influences the data collected, as well as Big Data influences consumer behavior.

[4] A cookie, also known as an HTTP cookie, web cookie, or browser cookie, is a small piece of data sent from a website and stored in a user's web browser while the user is browsing that website. Every time the user loads the website, the browser sends the cookie back to the server to notify the website of the user's previous activity. Cookies were designed to be a reliable mechanism for websites to remember statefull information (such as items in a shopping cart) or to record the user's browsing activity (including clicking particular buttons, logging in, or recording which pages were visited by the user as far back as months or years ago).

Collecting Big Data - an Orwellian reality or an individual fully assumed risk?

Tracking and analyzing users' profiles and personal history across different online environments are among the main activities taken into consideration when we tackle the subject Big Data. In fact, nowadays every human activity, from eating habits, sports activities, relationships, hobbies, political options, relationships, holidays, work life, or payments to medical records can be online monitored, registered, traced and put into the service of an advertising campaign, a political debate or a fundraising initiative, just to name a few. The "market opportunity" of browsing through the digital footprints of the online consumer can be and is already translated for each type of institutions into insights, predictions and activities trends. In the view of this online magnifying glass, the question that both the consumers and the companies should ask is where privacy and confidentiality stand? This "Big Brother" concerns are not new, but they are renewable, as the world becomes more connected, through phones, Internet, computers, networks and video cameras. Data can be transferred, sold, processed, stored and used in ways that only George Orwell or Aldous Huxley have imagined. Nowadays, even a new specialization has emerged–data brokers. They work for companies who gather, harvest and then redistribute highly personal data about persons to anyone willing to pay for it.

No more than 15 years ago (Google was created just the year before, and Facebook or Twitter did not exist yet), Scott McNealy, at that time CEO of Sun Microsystem[5], put it very briefly - consumer privacy issues are a "red herring" and "you have zero privacy anyway" in a meeting with journalists. His statement raised many problems and was followed by numerous critical points of view. For the purpose of our article, we quote Stephen Manes, editor at PC World (an??). He affirmed, "he (McNealy) is right on the facts, wrong on the attitude. It's undeniable that the existence of enormous data-bases on everything from our medical histories to whether we like beef jerky may make our lives an open book, thanks to the ability of computers to manipulate that information in every conceivable way. But I suspect even McNealy might have problems with somebody publishing his family's medical records on the Web, announcing his whereabouts to the

[5] Sun Microsystems, Inc. was a company that sold computers, computer components, computer software, and information technology services and that created the Java programming language and the Network File System.In 2010was acquired by Oracle.

world, or disseminating misinformation about his credit history. Instead of ‘getting over it,’ citizens need to demand clear rules on privacy, security, and confidentiality”.

14 years later, in 2013, the world encountered a new face of the problem, as Edward Snowden[6] leaked the information regarding National Security Agency’s (NSA)[7] program called PRISM[8]. This brought a new perspective regarding the value of privacy and confidentiality and many citizens become more aware of the possible implications data-mining could have. Nine in ten (88%) US consumers are at least “a little” concerned about the privacy of their personal data, new figures show (GfK, 2014).

According to Clemons et al. (2014), there are three directions when it comes to online privacy. The first one addresses the not allowed actions targeting someone’s personal space. Usually, these are sponsored actions such as spam, pop-up advertising, and online marketing. The second type implies a more serious threat about a person’s privacy, including identity theft and fraudulent activities. The third type is a more silent, but the most profound, the personal profiling developed by companies like Google or Facebook in order to obtain advertising benefits. The personal profiling includes all the types of information mentioned above, blended together in order to better understand who is the persons using their services. Although most of the companies claim that such measurements are mostly for helping customers to receive a personalized service, in reality the potential of personal data rises above the basic needs.

[6] Edward Joseph „Ed” Snowden is an American computer professional who leaked classified information from the National Security Agency, starting in June 2013.

[7] The National Security Agency (NSA) is a U.S. intelligence agency responsible for global monitoring, collection, decoding, translation and analysis of information and data for foreign intelligence and counterintelligence purposes - a discipline known as Signals intelligence. NSA is also charged with protection of U.S. government communications and information systems against penetration and network warfare

[8] PRISM is a clandestine mass electronic surveillance data mining program launched in 2007 by the National Security Agency (NSA), with participation from an unknown date by the British equivalent agency, GCHQ. PRISM is a government code name for a data-collection effort known officially by the SIGAD US-984XN. The Prism program collects stored Internet communications based on demands made to Internet companies such as Google Inc. under Section 702 of the FISA Amendments Act of 2008 to turn over any data that match court-approved search terms.

As mentioned earlier, people have started to be more preoccupied about their online presence. As recently as 2013, Pew Research, in a study regarding anonymity, privacy, and security online, revealed that 64% of people seeking online privacy clear their cookies and browser history, while 41% have disabled cookies. The same study found that 86% of the Internet users have tried to be anonymous online and they have taken at least one step to mask their behavior or avoid being tracked, and 55% have taken measures to hide from specific persons or organizations. To encourage free online navigation, many of the available browsers have an anonymous browsing mode option. Moreover, there are a series of applications that enable anti-tracking software, in order to erase the browsing history and other data. For example, AdBlock Plus, one of the most popular browser extension for blocking banner ads, pop-up ads, rollover ads, preventing visiting known malware-hosting domains, and disabling third-party tracking cookies and scripts, has been downloaded by 300,000,000 people (according to their own statistics). Some add-ons, for example Lightbeam for Firefox, allow users to visualize the first and third-parties sites s/he interacts online. Other reports highlight the increasing use of ad blocking add-ons for browsers and even Google trends reports (2013) showed that this type of software has an annual growth of 43%. For example, IAB report from 2012 regarding the consumer and online privacy stated that 45% of the respondents used clean-up programs and 30% used ad-blocking software. A 2013 report from PageFair[9] estimated that the average adblocking rate on 220 monitored website was 22.7%. According to their estimation, the adblock rate will continue to increase, reaching a 100% level in 2018.

Regardless though, for every action taken in order to protect the users' privacy, there is a counterpart that reminds them about the facilities they may gain from offering personal information. The online industry, and not only, has made a purpose from gathering as much as possible data, in order to offer a personalized experience to each consumer. From advertisers, to governments and nongovernmental organizations, each of them looks for opportunities to have access to users' data, in order to mine it and to be able to perform their activities even better.

[9] PageFair is a free service that allows websites owners to measure how many of their visitors block ads, and attempt to recover the lost revenue. For the mentioned report, they have been collecting anonymous data on adblocking behavior from their clients in 2012.

Research methodology

We propose in this paper an exploratory research in order to map the relation between the consumer and selected companies in what concerns the use of their data and the terms and conditions they agree upon when they start using their services and products. Our aim is to compare the type of information requested by two international companies and two Romanian companies from their users through the „Terms and Conditions" - contractual relations. Our methodological approach consisted in analyzing the online documents publicly posted on each company's website and conducting content analyses of the terms and conditions specified. The analysis unit has been the theme and the text we have studied are the provisions from Terms and Conditions. The main themes we have identified are: what data is being asked from users, how the data is created and used, and if the data can be used by third parties. All these themes have been split in sub-categories, for a more detailed analysis.

The coding process for each sub-category envisages giving points, incrementally, for each type of action required by the terms and conditions in relation to how much they are invading the intimacy and personal cyber space of the users.

As a result, the coding process for the analyzed text has been the following:

A. Data provided by users consisted in the following sub-categories co: account requirement, restrictions related to creating an account, type of information being displayed, options to restrict the company's access to personal data.

a. Account requirement: 1 - users need an account, 0 - users don't need an account.

b. Restrictions related to creating an account: 1 - restrictions are in place, 0 - restrictions are not in place.

c. Type of information being displayed: each type of information displayed received one point.

d. Options to restrict the company's access to personal data: 1 point for each limitation in service delivery the company is putting in place once the user decides to restrict the company's access to the personal data.

B: How data is created and used consisted in the following sub-categories: content property (creating, sharing, uploading, submitting, storing, sending, receiving), data storage, ceasing services.

a. Content property (creating, sharing, uploading, submitting, storing, sending, receiving): each type of activity performed with the raw data receives 1 point; each activity that is performed with the secondary data (processed information of the users) receives 0.5 points.

b. Data storage: storage of data indefinitely - 2 points, storage of data for definite amount of time based on self regulation - 1 point, storage of data according to law - 0 points.

c. Ceasing services: arbitrary cease of services - 2 points, cease of service due to misconduct of user - 1 point, cease of service based on contractual terms - 0 points;

C. Third party data usage consisted in the following sub-categories: use of data for the company's purposes; sharing data and using data for other purposes, tracking, monitoring and personal information analysis; data transfer in other countries; transparency on law enforcement requests.

a. Use of data for the company's purposes: for each purpose the personal data is used - 1 point;

b. Sharing data and using data for other purposes, tracking, monitoring and personal information analysis: all data is shared based on the acceptance of T&C - 2 points, data is shared based on legal regulation - 1 point, no data is shared with third parties - 0 points.

c. Data transfer in other countries: data transferred without restrictions once the T&C accepted - 1 point, data transferred based on legislation - 0 points.

d. Transparency on law enforcement requests: law enforcement requests made public - 0 point; law enforcement requests not made public: 1 points.

The criteria for selecting the four companies were their impact upon consumers in terms of daily use and the potential of generating data (over 1 million users, highly rate of daily content creation and sharing, number of monthly visits) and their turnover (over 500,000 Euro). Besides that, we took into consideration their location, including two Romanian-based companies. These companies are Google Inc., Facebook Inc., which are American based companies, and Dante International (owner of Emag.ro), and Orange, which are Romanian based companies.

Founded in 1998, Google Inc. has set itself the mission to organize the world's information and to make it universally accessible and useful. According to the company's financial reports, Google Inc. turnover in 2013 was USD 57.86 billion, the highest since the company was established, and currently, they process over 40,000 search queries every second on average, meaning more than 3.5 billion searches per day and 1.2 trillion searches per data, year worldwide. Taking into consideration the information offered by in4mation insights[10], a company specialized among other in Big Data, Google processes more than petabytes[11] a day. As Kulathuramaiyer and Balke (2006) stated, in the light of constant growth, Google is not really a competitor anymore, but already the environment.

The second company included in our research is Facebook Inc. Started as a student membership website, Facebook has surpassed at the beginning of 2014 1.23 billion monthly active users[12], 945 million mobile users, and 757 million daily users.

Emag.ro is one of the largest Romanian online stores, owned by Dante International. It started in 2001 as an online platform selling stationery and calculus systems. Currently, the online magazine offers products from a broad series of categories, from electronic equipment to cosmetics, toys, movies and fast moving consumer goods (FMCG). In their financial documents, the company reported 4 million users/month in 2013, and a turnover of 187 million Euros.

[10] in4mation insights, located in Needham, MA, was founded in 2006 by Mark Garratt and Steve Cohen. Their vision is to evolve the field of analytics and marketing research beyond the standard methods by providing the marketplace with highly innovative solutions and predictive tools.

[11] A petabyte is 1,048,576 gigabytes

[12] According to Facebook, active user is defined as an user who has logged into Facebook at least once in the previous 30 days.

Orange Romania is the France Telecom brand that offers worldwide mobile communications services, Internet and television, having 183 million clients worldwide. In Romania, it has 10,382,481 clients (as of October 2013) and a turnover of over 917,000 Euro. For this paper, we have analyzed the terms and conditions stated in the contract for the postpaid voice services.

Table 1. Comparison of Terms and Conditions from Google Inc, Facebook Inc, Emag.ro, and Orange Romania

Terms and conditions	**Google Inc.**	**Facebook Inc.**	**Emag.ro**	**Orange Romania (voice postpaid contract)**
Data provided by users				
Account requirement	1	1	1	1
Restrictions from creating an account	1	1	0	1
Public information displayed	3	3	0	0
Options to restrict the company's access to your personal data	1	1	1	1
How data is created and used: content property (creating, sharing, uploading, submitting, storing, sending, receiving); data storage, ceasing services				
Content property (creating, sharing, uploading, submitting, storing, sending, receiving)	9.5	4.5	9.5	1
Data storage	1	1	2	0
Ceasing services	2	1	1	-
Third party data usage: use of data for the company's purposes; sharing data and using data for other purposes, tracking, monitoring and personal information analysis; data transfer in other countries; transparency on law enforcement requests.				
Use of data for the company's purposes	7	13	8	7
Data transfer in other countries	0	1	1	1

Sharing data and using data for other purposes	2	2	2	2
Tracking, monitoring and personal information analyze	2	2	2	0
Transparency on law enforcement requests	0	0	1	1
Total	**30.5**	**31.5**	**27.5**	**14**

Results and discussions: there are close scores obtained by the companies which have been analyzed, apart from Orange, for which we have taken into account only one service. The highest scores are being recorded in those areas where the information is not only recorded but also processed and then either sent to other partners or sold for commercial purposes. The other high score is obtained in the area where personal information of users is treated as a commodity - where companies ask for this commodity in exchange to providing a certain service. In the rush to collect data, to share it, to analyze it, to process it, to mine it, in order to offer tailored services and to take advantage of every innovation, companies and consumers find themselves in middle of a strong debate regarding the privacy and confidentiality. Also, there are not big differences between the Romanian company and the US based ones - a sign that although the UE has tough regulation as regards personal data, they are either not applicable in Romania yet or the Romanian company is just doing things its way. The only big difference is in the sub-category how much of the personal information is displayed, where both Google and Facebook have higher scores than the Romanian companies. It's most probably something which relates with the type of business, rather than the care for the privacy rights in the case of the Romanian companies analyzed.

Fully securing our online data is no longer possible, and our online activities are subject to monetization, development and research. Consumers deserve to benefit from high standards of commitment from the companies they trust their information with. This means that both companies, and customers should act in a more responsible way confronted with personal and sensitive information. In the users case, many of them tend to be unaware of the potential dangers of over-sharing information on different online environments and the ways other persons, not companies, might take advantage of that specific information. The best example in this case is sharing information about ones location or holiday's location, leaving an open space for potential thieves. Also, research (Asay, 2013) and different experiments show that many users do no read the terms and conditions and the privacy policies[13].

Besides this, for the persons who read these documents, the language and the particularities of certain terms (e.g. data storage and legislation) might not be easy to understand. This leads us to the responsibility of the company. Under the façade of tailored services, we could see that the companies' practices go further. They can track many of our online activities, be it on their website or on others. Everything is measured and analyzed, making possible for third parties to benefit or could lead to a discriminatory profiling based on age, race, ethnicity etc.

In essence, our paper raises a few interesting issues to explore further: personal information and online actions are becoming commodities. At the same time, trade and revenues are generated by the primary and secondary processing of personal data. The actions of the companies are in a grey area, due to the fact that the information requested by Terms and Conditions is voluntary provided.

To put everything in balance is easy, but finding the right way to focus on the responsible way of collecting more data, because this is what future reveals to us, is a real challenge. The debate goes now to the ethical sphere, where the battle between acceptable and not acceptable, and the context and the purpose will play a bigger role in defining the ethical framework, more than legislation.

[13] Two situations are popular among the examples given to sustain these affirmations. In 2004, PC Pitstop, a company active in the technology field, put a clause in its end-user license agreement, offering $1,000 to the first person who emailed the company at a certain address. Only after five months and 3,000 sales, someone wrote the company asking for the sum of money. Another recent example, from 2010, refers to Gamestation, a computer game retail, which wanted to play a joke for April fools day, mentioned in their Terms and conditions that the users would sell them their souls. They added the „immortal soul clause" to the contract signed before making any online purchase, stating that customers grant the company the right to claim their soul. In that day, 7500 online agreements were signed.

What we see from our research is a need to invest more in the digital education of the consumer, to help him / her better understand his/her choices, the possible consequences of his/her online activities and the impact these could have upon shaping the legislation.

References

Asay, C.D. (2013). Consumer Information Privacy and the Problem(s) of Third-Party Disclosures. *Northwestern Journal of Technology and Intellectual Property*, 11(5), 321-330.

Bollier, D. (2010). The Promise and Peril of Big Data. Washington DC: Aspen Institute.

Boyd, D. and K. Crawford. (2012). Critical Questions for Big Data. Provocations for a Cultural, Technological and Scholarly Phenomenon. *Information, Communication & Society*, 15(5), 662-279.

Buck, C., Horbel, C., Kessler, T., and Germelmann, C. (2014). Mobile Consumer Apps: Big Data Brother is Watching You. *Marketing Review St. Gallen*, 1, 27-34.

Brown, J. (2010). Gamestation EULA collects 7500 souls from unsuspecting customers. Retrieved from www.geek.com,

Cardozo, N., Cohn, C., Higgins, P., Opsahl, K., and Reitman, R. (2014). The Electronic Frontier Foundation's Fourth Annual Report on Online Service Providers' Privacy and Transparency Practices Regarding Government Access to User Data. Retrieved from https://www.eff.org/files/2014/05/19/who-has-your-back-2014-govt-data-requests.pdf.

Chen, H., Chiang, R.H.L., and Storey, V.C. (2012). Business Intelligence and Analytics: From Big Data to Big Impact. *MIS Quarterly*, 36 (4), 1165-1188.

Clemons, E.K., Josh, W., and Fujie, J. (2014). Investigations into Consumers' Preferences Concerning Privacy: An Initial Step Towards the Development of Modern and Consistent Privacy Protections Around the Globe. *Proceedings of the 2014 47th Hawaii International Conference on Systems Sciences*, Waikoloa, Hawaii.

Dante International SRL (owner Emag.ro) (2014). Terms and Conditions and Privacy Policy. Retrieved from http://www.emag.ro/info/termeni-si-conditii.

Davenport, T.H., and D.J. Patil. (2012). Data Scientist - The Sexiest Job of the 21st Century. *Harvard Business Review* (October), 70-78.

Dowling, D.C. Jr. (2009). White and Case report –International Data protection and privacy law. Retrieved from http://www.whitecase.com/files/publication/367982f8-6dc9-478e-ab2f-5fdf2d96f84a/presentation/publicationattachment/30c48c85-a6c4-4c37-84bd-6a4851f87a77/article_intldataprotectionandprivacylaw_v5.pdf

EU (2014). European Directive on Data Protection. Retrieved from http://ec.europa.eu/justice/data-protection/.

Facebook (2014). Retrieved from https://www.facebook.com/legal/terms

Facebook (2014). Financial reports. Retrieved from http://investor.fb.com/results.cfm

Fang, Z., and Li, P. (2014). The Mechanism of "Big Data" Impact on Consumer Behavior. *American Journal of Industrial and Business Management*, 4, 45-50.

Gehrke, J. (2012). Quo vadis, data privacy? *Annals of the New York Academy of Sciences*, 1260, 45-54.

Google (2013). Trend Report. 2013. Adblocking and anti-tracking search queries Retrieved from http://www.google.com/trends/explore#date=1%2F2013%2012m&cmpt=q.

Google (2014). Financial report. 2014. Retrieved from https://www.google.com/finance?fstype=bi&cid=694653.

Google (2014). Terms and Conditions and Privacy Policy. Retrieved from http://www.google.com/intl/en/policies/terms/.

Groves, P., Kayyali, B., Knott, D., and Van Kuiken, S. (2013). The "Big Data" Revolution in Healthcare. Accelerating Value and Innovation. Center for US Health System Reform.

Herther, N.K. (2014). Global Efforts to redefine Privacy in the Age of Big Data. *Information Today*, 31(6), 33-36

IAB UK Report. (2012). Consumers and Online Privacy. Retrieved from http://www.iabuk.net/research/library/consumers-and-online-privacy-2012.

Kulathuramaiyer, N., and Balke, W.-T. (2006). Restricting the View and Connecting the Dots, Dangers of a Web Search Engine Monopoly. *Journal of Universal Computer Science*, 12(12), 1731-1740.

Laurila, J.K., et al. (2012). The Mobile Data Challenge: Big Data for Mobile Computing Research. Proceedings of the Workshop on the Nokia Mobile Data Challenge, in Conjunction with the 10th International Conference on Pervasive Computing.

Lohr, S. (2012). The Age of Big Data. *New York Times*. Retrieved from www.nytimes.com.

Loukides, M. (2011). What is Data Science. Sebastopol, CA: O'Reilly Media.

MacDermott, S., and Smith, J.R. (2013). The Future of Privacy: A Consumer-Oriented Approach to Managing Personal Data Online. *Tuntherbird International Business Review*, 55 (1), 3-12.

Mailat, C. (2014). Cel mai bun an pentru eMAG a fost încheiat cu o pierdere de aproape 7 mil. Lei. Retrieved from http://www.capital.ro/cel-mai-bun-an-pentru-emag-a-fost-incheiat-cu-o-pierdere-de-aproape-7-mil-lei-cum-explica-iulian-stanciu-rezultatul-negativ.html

McAfee, A., and Brynjolfsson, E. (2012). Big Data: The Management Revolution. *Harvard Business Review*, 59-66.

Needham, J. (2013). Disruptive Possibilities. How Big Data Changes Everything. Sebastopol, CA: O'Reilly Media.

Oprea, M. (2014). Legea Big Brother, declarată neconstituțională de CCR. Retrieved from http://www.avocatnet.ro/content/articles?id=37865.

Orange (2014). Terms and Conditions contract (postpaid services). Retrieved from https://www.orange.ro/produse-si-servicii/termeni-si-conditii.

Page Fair (2013). Page Fair Report. 2013. The Rise of Adblocking. Retrieved from http://downloads.pagefair.com/reports/the_rise_of_adblocking.pdf.

Pavolotsky, J. (2013). Privacy in the Age of Big Data. *Business Lawyer Journal*, 69(1), 217-225.

PewInternet (2013). PewInternet Report. 2013. Anonymity, Privacy, and Security Online. Retrieved from http://www.pewinternet.org/2013/09/05/anonymity-privacy-and-security-online-2/

Protalinksi, J. (2014). Facebook passes 1.23 billion monthly active users, 945 million mobile users, and 757 million daily users, January. Retrieved from http://thenextweb.com/facebook/2014/01/29/facebook-passes-1-23-billion-monthly-active-users-945-million-mobile-users-757-million-daily-users/.

Richards, N.M., and King, J.H. (2013). Three paradoxes of Big Data. *Standford Law Review Online*, 66(41), 41-46.

Richards, N.M., and King, J.H. (2014). Big Data Ethics. *Wake Forest Law Review*, 49, 393-432

Russom, P. (2011). Big Data Analytics. TDWI Research.

Sangameswar, S. (2013). Big Data - An Introduction. Kindle Books, version 1.1.

Schonberger, V.M., and K. Cukier. (2014). Big Data. New York: First Mariner Books.

Sprenger, P. (1999). Sun Over Privacy: Get it over it. Retrieved from http://archive.wired.com/politics/law/news/1999/01/17538.

Stanton, D. (2014). GfK survey on data privacy and trust. Retrieved from http://www.gfk.com/Documents/GfK-Privacy-Survey.pdf.

Siegel, A. (2013). When simplicity is the solution. Retrieved from online.wsj.com.

Tene, O., and Polonetsky, J. (2013). Big Data for All: Privacy and User Control in the Age of Analytics, Nw. J. Tech. & Intell. Prop. 239. Retrieved from http://scholarlycommons.law.northwestern.edu/njtip/vol11/iss5/1, viewed on 10.09.2014.

Vasilache, A. (2013). Numărul Clientilor Orange România a crescut cu peste 10,38 milioane de utilizatori (SIM-uri valabile), 20% dintre aceștia fiind utilizatori de smartphone. Retrieved from http://economie.hotnews.ro/stiri-telecom-15870802-numarul-clientilor-orange-romania-crescut-peste-10-38-milioane-utilizatori-sim-uri-valabile-20-dintre-acestia-fiind-utilizatori-smartphone.htm.

Watson, H.D. (2014). Addressing the Privacy Issues of Big Data. *Business Intelligence Journal*, 9(2), 4-7.

Whitepaper (2012). Challenges and Opportunities with Big Data. A Community White Paper Developed by Leading Researcher across the US. Retrieved from http://www.cra.org/ccc/files/docs/init/bigdatawhitepaper.pdf.

World of Statistics (2013). Statistics and Science-A Report of the London Workshop on the Future of the Statistical Sciences. 2013. Retrieved from http://www.worldofstatistics.org/wos/pdfs/Statistics&Science-TheLondonWorkshopReport.pdf.

Yetiskin, E. (2014). Economediatic Data. An Introduction to Critical Big Data Studies. Retrieved from https://www.academia.edu/8122558/Ecomediatic_Data__An_Introduction_to_Criticial_Big_Data_Studies.