

Data de recebimento: 05/04/2016
Data de aceitação: 30/05/2016

A DATA SOCIETY: INFORMATION, CONTROL, AUTOMATION AND THE LAW

André Martins Brandão¹

ABSTRACT:

The rapid evolution of communication and information technology in recent decades changed social relations. Human communication is remedied - begins to occur by other means. In this paradigm was born the so-called era of information or the information society, where information acquires essential good status, perhaps the more profitable and important of all - who controls information controls the world. Nowadays useful information can be extracted from data – content massively generated in networks every day, such as the internet. This raises a question: how to understand control relations having in sight the contemporary ocean of data? A deleuzian model can offer potential to understand the control relations in contemporary society. The evolution of information, control and automation technologies in recent decades and the emergence of the information era are trademarks of what can be referred as society of control, which has for objective the constant modulation of individuals for its purposes – a permanent production of subjectivity. The control mechanisms are continuously exercised over subjects, for the purpose of modulating their behavior and actions. As control is based in information, and information is extracted from data – the society of control can be referred as a society of data, a broader concept that can cover other mechanisms and technologies. Big data and analytics technologies can be examples of this functioning. These technologies were created to search for correlations in the ocean of data, in order to extract useful information, generating impacts in almost every social field, including the Law. In this context, another question appears: how big data technologies may transform legal practice in the contemporary data society? On one hand, there are possibilities for the use of big data technologies in order to assist in the legal decision-making process, as well as the reverse is possible. It is necessary to deeply understand the structure and functions of big data technologies, in order to enable its use to help human decision-making. On the other hand, these technologies may create new forms of discrimination, and also increasingly threaten privacy and even the personal identity construction process. There are paradoxes in these technologies that may be less problematic with the construction of a big data ethics.

Keywords: Data society; information; control; automation; big data; Law.

¹ Doutorando em Filosofia do Direito pela Pontifícia Universidade Católica de São Paulo - PUC/SP. Bolsista CAPES.. Pesquisador do Grupo de Estudos sobre Direito, Análise, Informação e Sistemas (GEDAIS) da PUC/SP. Pesquisa principalmente sobre a interseção entre filosofia, tecnologia e direito. andrebrandao@gmail.com

UMA SOCIEDADE DE DADOS: INFORMAÇÃO, CONTROLE, AUTOMAÇÃO E O DIREITO

RESUMO

A rápida evolução das tecnologias de comunicação e informação nas recentes décadas alteraram as relações sociais. Comunicações humanas são remediadas – começam a ocorrer por outros meios. Nesse paradigma nasceu a chamada era ou sociedade da informação, onde a informação adquire status de bem essencial, talvez o mais importante e lucrativo de todos – quem controla a informação controla o mundo. Hoje em dia informação útil pode ser extraída de dados – conteúdo gerado massivamente em redes todos os dias – como a internet. Isso levante uma questão: como entender as relações de controle no oceano de dados contemporâneo? Um modelo deleuziano pode oferecer potencial para entender as relações de controle da sociedade contemporânea. A evolução das tecnologias de informação, controle e automação nas recentes décadas e a emergência de uma era da informação são marcas do que pode ser referido como sociedade de controle, a qual tem por objetivo a modulação constante de indivíduos para os seus propósitos – uma permanente produção de subjetividade. Os mecanismos de controle são continuamente exercidos sobre os sujeitos, com o fim de modular os seus comportamentos e ações. Como o controle é baseado em informação, e a informação por sua vez extraída de dados, a sociedade de controle pode ser também referida como sociedade de dados, nome mais abrangente para abarcar outros mecanismos e tecnologias. As tecnologias de big data e analytics podem ser exemplos desse funcionamento. Tais tecnologias foram criadas para procurar por correlações no oceano de dados, em ordem de extrair informação útil, gerando impactos em quase todos os campos sociais, incluindo o direito. Nesse contexto, surge outra questão: como as tecnologias de big data podem transformar a prática jurídica na atual sociedade de dados? Por um lado, existem possibilidades para o uso de tecnologias de big data com fulcro de auxiliar no processo de tomada de decisão, assim como o reverso é possível. É necessário entender profundamente a estrutura e o funcionamento das tecnologias de big data, em ordem de possibilitar seu uso no processo humano de tomada de decisão. Por outro lado, tais tecnologias podem criar novas formas de discriminação, e também ameaçar a privacidade e até o processo de construção de identidade pessoal. Existem paradoxos nesses tecnologias, que podem ser menos problemáticos como construção de uma ética de big data.

Palavras-chave: Sociedade de dados; informação; controle; automação; big data; direito.

Introduction

Network theorists, such as Manuel Castells, often speak about an information society and the emancipatory possibilities of the new communication methods. Gilles Deleuze (2008, p.224-225), on other hand, affirms that “the conception of a control mechanism, giving the position of any element within an open environment at any given instant (whether animal in a reserve or human in a corporation, as with an electronic collar), is not necessarily one of science fiction.” The author remembers that “Felix Guattari has imagined a city where one would be able to leave one's apartment, one's street, one's neighborhood, thanks to one's (dividual) electronic card that raises a given barrier;” What happens is that electronic collar could also “easily be rejected on a given day or between certain hours; what counts is not the barrier but the computer that tracks each person's position--licit or illicit--and effects a universal modulation.”

From these writings, it can be seen that a deleuzian model offers potential to understand the complex relations in contemporary societies, which have information as their most valuable good, mainly extracted from data. The first chapter of the present paper will follow the deleuzian model as a way to understand the modes of control of the contemporary world.

The rapid evolution of Technologies of Information and Communication (TIC) played a major role in the construction of this new society, almost edging science fiction. Maybe Gibson had imagined something like it in *Neuromancer*, but Clarke e Asimov certainly didn't. This evolution modifies human relations. The velocity of communication or access to information, the new habits formed by social networks and information control – times are changing, Dylan would say.

In the last ten years the possibility to generate content increased to levels never seen before. How to navigate efficiently in this ocean of data? Big data technologies present an answer to this question, created to seek for correlations in the data stream, with the goal to extract useful information.

Based on that, the second chapter of the present conjecture poses the following question: how big data technologies can modify legal practice? It has for hypothesis that the use of big data technologies generates impacts on many areas touched by the legal phenomena, such as privacy, discrimination, fundamental right to information and even automation of (legal) decision-making, among others. This paper aims to construct a conjecture about possible transformations in legal practice from these impacts.

1. Information Era: for whom?

In the last few decades information technologies have experienced an intense evolution. Telephones became mobile, being carried wherever a person goes – the number of mobile phones in certain countries is bigger than the number its habitants. The internet, a network firstly formed by documents connected among themselves by hyperlinks, and nowadays full of other software and apps, such as web crawlers, enables instantaneous sharing and access of data all over the globe – modifying personal interrelations. Information is the nucleus of these technologies. Manuel Castells affirms:

“A technological revolution centered in information technologies is accelerated remodeling the material basis of society. Economies all over the world started to maintain interdependency, presenting a new form of relation between economy, state and society in a system of variable geometry.” (CASTELLS, 1999, p. 21)

The new paradigm has for principal source of energy the information technology, as the past ones had from the steam engine to electricity (CASTELLS, 1999, p. 50). In Castell's model (1999, p. 25) the informational paradigm was constituted in the seventies, mainly in the USA, from the interaction of a specific american segment interacting with the global economy, creating a new mode of production, management, communication and life.

The information era is born for Castells (1999, p. 31), and inside it the information society, a society based on capitalist and informational paradigms, where productivity sources are found in technologies of knowledge generation, information processing and symbol communication. Information society, in Castells's (1999, p. 35) model, follows industrial society, that had as motor the production and circulation of goods. Every society depends on information and knowledge, but in this new paradigm the technologies of information themselves are developed, having the objective of knowledge accumulation and greater complexity in informational processing.

Castells (1999, p. 38) warns that it is not a forced cultural homogenization to speak of a global informational society. It deals with fundamental traces among sociotechnical systems from these cultures, as there is reference to an industrial global society, which had as its basis production and circulation of merchandise.

In this informational scenery, for Castells (1999, p. 36), there is a strong connection between culture and productive forces, mainly because it is constructed on knowledge and information technologies. Therefore, new forms of human interaction and social transformation must appear. With this hope, Castells demonstrates his faith in the possibility of human emancipation in the informational times through social transformation. The present paper, on other hand, proposes the observation of current times through the deleuzian model, as a way to try to understand the contemporary control relations, based on information and data.

1.1. From a disciplinary to a control society.

In the deleuzian model contemporary society is established on control. The evolution of information technology in the last few decades and the appearance of the informational era are trademarks of what can be called control society, which has for objective the constant modulation of individuals – the production of subjectivities. The control society is temporally preceded by what Michel Foucault called disciplinary society. The investigation of disciplines is important to clarify the functioning of control.

Michel Foucault thought that closed spaces shape people and produce subjects – a subjectification process. The life of an individual consists in going from a space to another, from house to school, from school to work, sometimes to hospital, to barracks or even prison. Confined spaces model individuals, each one with its own rules and peculiar confinement – family, school, hospital, work and prison, among others. Foucault calls this model disciplinary society, instituted in the centuries XVIII e XIX, and that reached its peak in the beginning of the XX century. Discipline act on bodies and their actions, being constantly exercised from vigilance structures and essential for the constitution of the industrial capitalism.

Foucault located the disciplinary societies in the eighteenth and nineteenth centuries; they reach their height at the outset of the twentieth. They initiate the organization of vast spaces of enclosure. The individual never ceases passing from one closed environment to another, each having its own laws: first the family; then the school ("you are no longer in your family");

then the barracks ("you are no longer at school"); then the factory; from time to time the hospital; possibly the prison, the preeminent instance of the enclosed environment (...). Foucault has brilliantly analyzed the ideal project of these environments of enclosure, particularly visible within the factory: to concentrate; to distribute in space; to order in time; to compose a productive force within the dimension of space-time whose effect will be greater than the sum of its component forces. (Deleuze, 2008, p. 219)

The disciplinary societies illustrated by Foucault (2009, p. 187) utilize discipline and normalization to increase its efficiency and productivity. They do that through closed spaces – like school or prison. The disciplinary model, through technologies of control and identity constitution, allows the construction of more docile and able subjects, working in their bodies and actions. Foucault sees in this practice the process of *subjetificação* in the industrial society, where the model is the factory, in which the objective is the sum of individual forces to reach a greater productive force.

“Esses métodos que permitem o controle minucioso das operações do corpo, que realizam a sujeição constante de suas forças e lhes impõem uma relação de docilidade-utilidade, são o que podemos chamar as “disciplinas”. Muitos processos disciplinares existiam há muito tempo: nos conventos, nos exércitos, nas oficinas também. Mas as disciplinas se tornaram no decorrer dos séculos XVII e XVIII fórmulas gerais de dominação. (...) O momento histórico das disciplinas é o momento em que nasce uma arte do corpo humano, que visa não unicamente o aumento de suas habilidades, nem tampouco aprofundar sua sujeição, mas a formação de uma relação que no mesmo mecanismo o torna tanto mais obediente quanto é mais útil, e inversamente. Forma-se então uma política das coerções que são um trabalho sobre o corpo, uma manipulação calculada de seus elementos, de seus gestos, de seus comportamentos. (...) A disciplina fabrica assim corpos submissos e exercitados, corpos “dóceis”.” (Foucault, 1999, p. 118)

Disciplinary power, acting in closed spaces and having the objective to constitute subjectivities, is mainly performed by three technologies, according to Foucault (1999, p. 148-149): first hierarchical vigilance: “graças às técnicas de vigilância, a ‘física’ do poder, o domínio sobre o corpo se efetua segundo as leis da ótica e da mecânica, segundo um jogo de espaços, de linhas, de feixes de graus, e sem recurso (...) ao excesso, a força”. In social institutions, like school, students are under constant surveillance to continue the practices of discipline. The fear of being caught doing something considered wrong oppresses subjects, which are in constant surveillance, making their behavior normalized according to the power relations. The second aspect is called normalizing sanction. Those who behave in a deviant fashion become subjects of sanctions aiming normalization from physical and psychological punishments. The third and last is the exam. The quantitative thinking in modern society transforms subjects in numbers, in a quantifiable object that will have its story analyzed by documents collected through life.

Nowadays disciplines are in crisis. Deleuze affirms (2008, p. 219) that all means of confinement are in crises. Daily news talk about reforms of schools, industries, prisons as an attempt to solve it – but they are doomed with the news forces that come. “They are the control societies substituting disciplinary societies” The disciplinary societies did not cease to exist – control overlaps the old societies. It is not a pure and simple substitution. Controls takes place, but disciplines keep on acting.

But in their turn the disciplines underwent a crisis to the benefit of new forces that were gradually instituted and which accelerated after World War II: a disciplinary society was what we already no longer were, what we had ceased to be. (...)We are in a generalized crisis in relation to all the environments of enclosure--prison, hospital, factory, school, family. The family is an "interior," in crisis like all other interiors--scholarly, professional, etc. The administrations in charge never cease announcing supposedly necessary reforms: to reform schools, to reform industries, hospitals, the armed forces, prisons. But everyone knows that these institutions are finished, whatever the length of their expiration periods. It's only a matter of administering their last rites and of keeping people employed until the installation of the new forces knocking at the door. These are the societies of control, which are in the process of replacing disciplinary societies."Control" is the name Burroughs proposes as a term for the new monster, one that Foucault recognizes as our immediate future. (DELEUZE, 2008, p. 219-220).

Deleuze finds the name control in the work of the *beat* writer William S. Burroughs (1914-1997). He was one of the main writers of the American counterculture, involved in the beatnik movement with Allen Ginsberg and Jack Kerouac, among others. Burroughs is constantly concerned with control mechanisms in his work. Addiction appears as a possibility of a control model, for money, drugs, sex or power. In a paper published in the late 70's – called *Limits of Control* – Burroughs describes his understanding of control, by drawing its limits.

The American author sees the basic limits for control: it needs time to be exercised, because it also need opposition or acceptance – there is no such thing as a complete control. According to Burroughs (1978, p. 38) one can control a dog or a worker, but if full control is established, the controlled subject becomes nothing more than an automaton, a robot – there is no life there. It is not necessary to control an automaton; it is possible to use it. Control mechanisms try to exercise it in the most complete way, but if they are extremely successful, there is nothing left to control – only machines to be used.

When there is no more opposition, control becomes a meaningless proposition. It is highly questionable whether a human organism could survive complete control. There would be nothing there. No persons there. Life is will (motivation) and the workers would no longer be alive, perhaps literally. The concept of suggestion as a complete technique presupposes that control is partial and not complete. You do not have to give suggestions to your tape recorder nor subject it to pain and coercion or persuasion. (BURROUGHS, 1978, p. 38)

Burroughs (1978, p. 39) believes that he who tries to exercise control do it because of self-preservation, seeking to protect certain advantage position. Control practices are needed to maintain that position. Control is successful when it succeed at maintain equilibrium, avoiding situations that would need application of full force.

The name of the new monster – control -, found in Burroughs, starts being used by Deleuze to define contemporary society. The new technologies of information are essential to that process. The information society, imagined by Castells, can be seen from Deleuze's model as the society of control.

In the discipline logic the means of confinement are independent variables – each one works by their own rules. Discipline is based on closed architectural spaces – house, school, barracks, hospital, prison. House rules are not valid in school, as school rules are not the same

as work ones, for example. In each space the individual experiments a new beginning. He returns to zero. There is a common language to closed spaces, but it works analogically, do not allowing connection between spaces, and favoring independent variation. (DELEUZE, 2008, p. 220).

In control society the control means are inseparable variations. They work continuously without interruption – the environments connect with each other. While individuals under discipline did not cease to begin, those under control, given its continuity, do not finish a thing. Formal education becomes permanent – school invades work, the latter invades the first, and both the house, for example. Language in modes of control is numeric, and not analogical, what enables that connection. Different from closed spaces, which function as molds in the disciplinary society, the control mechanisms act as modulations, which enable scenery of metastability, given the possibility of continuous self-deformations and changes from one point to another (DELEUZE, 2008, p. 221).

In control society the factory stops being the model for work, which starts to be the company. A company is a soul or a gas – spreads through all spaces of life, in other days compartmentalized.

the factory was a body that contained its internal forces at the level of equilibrium, the highest possible in terms of production, the lowest possible in terms of wages; but in a society of control, the corporation has replaced the factory, and the corporation is a spirit, a gas. Of course the factory was already familiar with the system of bonuses, but the corporation works more deeply to impose a modulation of each salary, in states of perpetual metastability that operate through challenges, contests, and highly comic group sessions. (...).The factory constituted individuals as a single body to the double advantage of the boss who surveyed each element within the mass and the unions who mobilized a mass resistance; but the corporation constantly presents the brashest rivalry as a healthy form of emulation, an excellent motivational force that opposes individuals against one another and runs through each, dividing each within.. (DELEUZE, 2008, p. 225).

In disciplinary societies, the signature and registration number function as two poles: the first identifies individuality and the second the position inside a mass. An individuation and a massification occur – a unique mass for power exercise is constituted and at the same time the mold for the individual. In control societies those two poles are substituted by the cypher, which marks the access or rejection to information in a numerical way. The individuals start to be dividuals, and the masses samples. Money can distinguish societies, according to Deleuze:

Perhaps it is money that expresses the distinction between the two societies best, since discipline always referred back to minted money that locks gold as numerical standard, while control relates to floating rates of exchange, modulated according to a rate established by a set of standard currencies. (DELEUZE, 2008, p. 226).

Each form of society corresponds to certain types of machine they operate, not because those machines determine society, but because they show the social forms capable of generating and using them. Deleuze (2008, p. 223) affirms that sovereignty societies, previous to disciplinary ones, correspond to simple machines, like sheaves, levers and clocks. Disciplinary societies correspond to energetic machines, which possessed the active danger of sabotage and the passive of entropy. Control societies operate through informatics machines

and computers – information technologies, having the danger of piracy and virus introduction, and the passive of interference.

It is an evolution both in technology and capitalism, which stops being of concentration and production, the coverage model, and starts to turn to services and stocks – the model of the market's constant flow. Mass media and marketing start being tools of social control, and the gigantic flow of information is in the nucleus of the strategies. This possibility of speed modified the way of power exercise from discipline to control, according to Deleuze (2008, p. 224): “Control is short-term and of rapid rates of turnover, but also continuous and without limit, while discipline was of long duration, infinite and discontinuous. Man is no longer man enclosed, but man in debt.”

Deleuze describes an image in which the mechanisms of control are continuously exercised on subjects, aiming the modulation of their behaviors and actions. The narrative resembles a dystopian future imagined by Orwell or Huxley, but it is a raw presentation of how data societies function, based on information, control and automation technologies.

The conception of a control mechanism, giving the position of any element within an open environment at any given instant (whether animal in a reserve or human in a corporation, as with an electronic collar), is not necessarily one of science fiction. Felix Guattari has imagined a city where one would be able to leave one's apartment, one's street, one's neighborhood, thanks to one's (dividual) electronic card that raises a given barrier; but the card could just as easily be rejected on a given day or between certain hours; what counts is not the barrier but the computer that tracks each person's position--licit or illicit--and effects a universal modulation. (DELEUZE, 2008, p.224-225).

Deleuze (2008, p.225) highlights the importance of the study of control mechanisms nowadays, aiming to show how they overlap disciplines. This kind of study is aimed by the present paper: “The socio-technological study of the mechanisms of control, grasped at their inception, would have to be categorical and to describe what is already in the process of substitution for the disciplinary sites of enclosure, whose crisis is everywhere proclaimed.”

2. Big data technologies

The fast evolution of information and communication technologies in the last ten years increased the possibility to generate content in a way never seen before. It is estimated that from the invention of writing until 2006 humanity accumulated about 180 Exabytes (EB) of data. An Exabyte is a unity of measurement of data equivalent to a quintillion of bytes (10^{16}). It is followed, in a descending order, by petabytes, terabytes, gigabytes, megabytes, kilobytes, bytes and finally bits. This number has grown ten times between 2006 and 2011, reaching 1600 Exabytes. (FLORIDI, 2012)

The growth of the volume of data is stimulated by social network applications, which permit to any subject the creation of large set of content almost without costs. Smartphones accelerate this process – subject dispose of a nomad port, capable of generate content from anywhere. The internet of things (*IOT*) is also a factor that must be considered, multiplying the autonomous artificial sources of data (seen as subjects, in this conjecture). There is the expectation to extract valuable information from these sources (ERMOLAYEV V. et al, 2014, p. 4).

“As a result, there is much talk about Big Data.” (FLORIDI, 2012, p. 436) The name Big Data is defined by the ASE International Conference (*Academy of Science and Engineering, de Harvard*) as: *“(...) large, diverse, complex, longitudinal, and/or distributed*

data sets generated from instruments, sensors, Internet transactions, email, video, click streams, and/or all other digital sources available today and in the future.” (2014, p. 01)

The exponential growth of the processing power by multiple nucleuses, the growth in speed in networks and parallel computing, among others, allow today the exploration of big data. The paper about *Map-Reduce* technology by J. Dean e S. Ghemawat (2008) was important for many developments in this area. Despite the *Map-Reduce* being already known at the time, the combination with *Apache Hadoop* allowed processing of large sets of data in a cluster of nodes. What makes big data an important tool today is the possibility to deal with large sets of data.

Never before was necessary to deal with such an amount of data – it is an epistemological problem: how to deal with this complexity? Another problem is ethical: how to use those data? (FLORIDI, 2012).

The approximation with the epistemological problem shows that big data technologies lead with complexity recognizing patterns or correlations. It works with the possibility to find patterns and use them, for the wanted goals – that enter in the ethics field. The neuralgic point of big data is pattern recognition or correlations in complexities of data, generating useful and relevant knowledge. (FLORIDI, 2012).

Many professionals of the information area utilize big data aiming collection and analysis of data. *“Fundamentally, big data analytics is a workflow that distills terabytes of low-value data down to, in some cases, a single bit of high-value data. The goal is to see the big picture from the minutia of our digital lives.”* (FISHER et al. 2012, p. 53)

Some theorists even speak of a society of data or economy of data, and not society of information and economy of information *“It’s all data now: Data Economy and Data Society. This is a confession that we are no longer in control of the knowledge contained in the data our systems collect.”* (GRELLER, 2012) In this way, the use of this huge amount of data to extract information and enable control is a key factor in society today – that is why society of data seems a good nomenclature to describe what happens in these days.

Big data, in a certain way, is a name utilized to describe *“large volumes of unstructured and structured content—usually in amounts measured in terabytes or petabytes—that enterprises want to harness and analyze.”* Traditional technologies of data managing, of relational character, have difficulties to deal with this large insertion of data, becoming difficult to manipulate them. It is necessary new technologies that can deal with those difficulties. (ERMOLAYEV V. et al, p. 32)

The analysis of massive quantities of data requires different tools than traditional science, mainly for the need of velocity in processing. Inductive logic appears as a way to fast analyze the great amount of data in a big data system, having for main characteristic its potential for pattern recognition and predictability.

2.1. Conjectures about legal practice in the world of big data.

The world has been through intense transformation. The accelerated development of information and communications technologies in the last few decades, as well as automation and control technologies, has played a major part in that journey. Information is one of the most valuable goods today, enabling exercise of a still unmeasured control. Big data technologies have the ambition to a more fast and efficient navigation on the ocean of data (which only tends to grow) – enabling new control modulations.

The market and professional practice are also changing in many fields. The stock markets stopped being places where a euphoric human chaos of buying and selling reigned, and became controlled by informatics systems. The analysts work on their chairs (mainly with graphics), administrating their accounts using the computer. The derivative market (in which

speed is essential) increasingly uses artificial intelligence systems to perform operations in an unmatched speed for a human being. In the real estate market the internet is already an essential place to do negotiations. Also in the movable products market – search for goods in the internet already became a normal task. Therefore, it is asked: which possible transformations have occurred (or might occur) in legal practice?

The hypothesis is that in current data society, of control and information, the use of big data technologies generate impacts to many areas touched by the legal phenomenon, as privacy, discrimination, right to information, democracy and even automation of (legal), among other. This work seeks to build a conjecture about possible transformations in legal practice from these impacts.

It is important to clarify that this work does not see the new technological tools only as a good thing – a way to solve old problems of the legal system. They are a new way to work, to deal with problems, but not a *Deus ex machina*. As all new methods they possess advantages and disadvantages, and this will be taken into account in this work.

2.1.1. Big data as a tool in decision-making.

The predictive potential of big data technologies, using pattern recognition in large sets of data, shows itself as a possible legal tool in a society based on information, control and automation. Moses and Chan (2014) affirm that analytical methods, mainly in USA, start to colonize administration of justice and legal practice. The authors suggest that “*with sufficiently large datasets and the right analytics and machine learning techniques, we will have simple answers to traditionally difficult questions.*” (MOSES; CHAN, 2014, p. 643) So, patterns found by the analysis can help in human decision-making.

In the paper *Quantitative legal prediction—or—how i learned to stop worrying and start preparing for the data-driven future of the legal services industry* Daniel Martin Katz (2013) affirms that quantitative legal prediction is closer to legal practice, and will define great part of the innovation in the legal services industry, mainly supported by the access to enormous sets of semi-structured legal information.

One of the tasks of the human being specialized in law is to predict the possible outcome of legal disputes, generally supporting his strategy from that prediction. Legal training and experience are key factors to enable this task. Experienced and well trained jurists are capable of predict with more efficiency the outcome of cases. These professionals are expensive and have limitation, as all human beings. Katz (2013, p. 928) defends that this flaw can be remedied by prediction technologies, which “*are designed to remedy or supplement the shortcomings of human reasoners.*” It is not a complete automation of decision, a substitution of human beings, but a combination between man and machine, aiming to increase efficiency in decision-making processes: “*the age of quantitative legal prediction is about a mixture of humans and machines working together to outperform either working in isolation. The equation is simple: Humans + Machines > Humans or Machines.*” (KATZ, 2013, p. 929)

These kinds of technology already exist. Prediction technics are already used in decision-making in public and private sectors. Lex Machina² is an example. It was created in 2010 aiming to predict the result of legal cases involving intellectual property, using a big data software. In Virginia statistical data about the possibility of criminal relapse is used to fundament probation decision, using informatics systems. (MOSES; CHAN, 2014). Some

² lexmachina.com

American jurisdictions use data analysis tools to help in decision about contractual bail, founded in objective risk measure's (MURPHY, 2013). A bayesian model was developed to predict results of settlements in *class actions* involving *securities fraud*. (McSHANE et al, 2012.)

The use of those tools certainly was due to advancements in computational power (to gather and analyze big sets of data) in the last few decades. But legal data analysis was born with the use of statistics, and after with the use of machine learning as a computational tool for decision-making – “the empirical turn”. It is a change in paradigm in the form of legal practice. Legal decisions are often founded in legal knowledge, in the experience and a certain intuition of the one who decides. An empiric approach, as propose by Lee Loevinger (1949) jurimetrics, aim to construct those decisions using mathematical and statistical models, seeking to fundament them in statistical evidence. *Machine learning*, in turn, as others statistical technics, aims to infer certain observation data through time. The classical statistics work with hypothesis build from theories or other researches, testing those hypotheses aiming to find statistical inferences. Unlike it, *machine learning* does not work with given hypothesis, “it analyses ‘training’ data and, through the use of an algorithm, identifies the ‘best’ hypothesis linking input data to outputs. The ‘training’ data is simply the examples fed into the algorithm, from which it ‘learns’ potential predictive relationships.” (Moses e Chan, 2014, p. 647). Machine learning technology is driven by inductive algorithms, seeking the statistical prediction of certain results.

In the 70's and 80's the first technological tools for decision-making tried to mimic the traditional legal reasoning processes. Called *expert systems*, they had the objective to use programs to model the way that a legal actor would deal with a case, including reason their conclusions. (TYREE, 1989). It was an attempt of isomorphism of the human jurist – the creation of a human mirror. Moses and Chan (2014) argument that those systems fail to mimic correctly the reasoning of a specialist. The author affirm that “not only are legal rules often contradictory, circular, ambiguous or deliberately vague or contestable, but they rely on social context and human interpretation and cannot be applied directly to raw facts.” (MOSES; CHAN, 2014, p. 657) Systems that aim to help in human decision making would be more successful than those that seek to replace human beings.

Closer to Loevinger's jurimetrics and *Machine Learning* (differentiating from the last one mainly because of the amount of data), big data technologies use empirical technics, mainly inductive methods: “big data analytics speaks the language of probability, enhancing decision-making by estimating the likelihood that particular facts are or will be true” (MOSES; CHAN, 2014, p. 663). Unlike *experts systems*, that tried to mimic a kind of deductive human reasoning.

Moses and Chan (2014, p. 653-656) present a tridimensional model to evaluate big data tools in the legal area: a technical, a social one and a normative dimension. The technical dimension evaluates functionality and affectivity – if big data tools are capable of improving legal decision-making; the social dimension evaluates the factors that may impact in the use of these tools; the normative dimension tries to understand if these tools are adequate to morals. The authors call the first criterion “‘effectiveness’, the second ‘acceptability’ and the third ‘appropriateness’”.

The first dimension relates to precision or exactness. The information provided must be accurate. Studies like Ruger (2004) show that possibility. Machine learning predictions in this case, using the *classification tree* method, related to individual votes of the Supreme Court Justices in future cases was compared to predictions of law professors and lawyers. The program won for 75% to 59%. Even though this result, it is good to remember that statistical errors are common, as do correct results. Generated information can be useful or useless to the

given situation. Any prediction is just a probabilistic attempt, never a complete certainty. (KATZ, 2013)

Moses and Chan (2014, p. 666) also remembers that human preconceptions are introduced in data analysis – “*It is a human who identifies and selects the data to be analyzed and chooses the algorithm to be employed. Some of the time, a human also selects the attributes and variables that are treated as relevant.*” It is not a totally impartial and objective rationality. This is the main reason that those who seek for help in decision-making technologies cannot believe that data, numbers or generate information speak for themselves. They are constructed in the basis of some preconceptions (mainly the programmer and designer), and from the moment they go through the human process of interpretation and comprehension gain new senses from each actor’s situational locus. Sense is not a given, but always a construct.

As for the dimension of acceptability, Moses and Chan (2014) consider hard to predict the level of acceptance in the legal profession. It depends on costs, benefits and technology capacity. The legal field is usually reluctant to new technologies, preferring human reasoning. A technology that wanted to replace human beings, like expert systems, would hardly be accepted. Technologies like big data analysis, which aims to help human beings in the decision-making process, seems to have more chance of acceptance.

The normative dimension deals mainly with compatibility to legal norms, transparency and *accountability*. Moses and Chan (2014) affirm that certain uses of big data may be contrary to legal norms. Classical legal reasoning takes into account normative factors in decision-making, while inductive big data logic works with identification of correlation of factors with certain results. This correlation does not need to be normative or even have a causal relevance. This is an open door to discriminatory decisions, which can be grounded in problematic or historically prohibitive criteria.

Transparency and *accountability* are also important. Actors in legal areas usually are not well prepared to deal with technological tools. The lack of understanding about the functioning of a tool may lead to difficulties in understanding the results. Moreover, the difficulty to deal with inductive processes can jeopardize the level of accountability (this was the computer’s fault!)

There are possibilities that big data technologies might help with legal decision-making, as the opposite is possible. It is necessary knowledge of the functioning and structure of the tools, in order to enable its use inside the limits proposed, and help human decision-making.

In a general perspective, parts of the functions of all white collars careers were automatized, in finances, medicine or law. Katz (2013) affirms that in a near future the technological evolution will occur with a combination between humans and machines, aiming to accomplish more efficient tasks, that none could do alone.

2.1.2. The problem of privacy, big data and the user-product.

The problem between privacy and new technologies is not exclusive of the current times. When Warren e Brandeis wrote their paper - *The Right to Privacy (1890)*, the worries were with new technologies at the time, as cameras and newspapers, which had invaded the sacred place of domestic life. In current times this problem is bigger. Technology is more accessible – there are more than 2.4 million internet users in the world – certainly a bigger set than Brandeis and Warren imagined when they wrote about protection of privacy.

Scandals about privacy violation erupted in the last few years in all spaces: public and private, familiar and social. People have never been so observed. In the paper *big data in small hands*, Hartzog and Selinger (2013, p. 251) argument that the social benefits of big data

technologies must be reconciled with the risk that they present to privacy: *“concerns arise regarding profiling, tracking, discrimination, exclusion, government surveillance, and loss of control.”*

In the article *Abandon all Hope?* Graham Greenleaf (2014, p. 640) deals with some of the factors that difficult preservation of privacy: environmental factors, *“these include the progressive elimination of processing costs (Moore’s law); the progressive elimination of data storage costs; and the explosion of online transactions.”* The great generation and collection of personal data, that involves *“digitization of new/more personal data; visual data collection; voluntary disclosure; big data aggregations; de-anonymisation of transactions; re-identification techniques; government IDs; and commercial personal data collection.”*; intensification of the use and processing of personal data, by many tools designed with that objective *“including by more powerful processing of personal data (analytics, search and ranking, etc); through commercial interconnection (eg, advertising syndication, internet of things); and by state interconnection.”*; the reduction of safety in personal data and the increase of its retention, *“including by permanent retention of personal data (both on our own devices, and on third-party devices); by the endemic failure of security systems; and the risks of massive systems failure and data unavailability.”*; the crescent transference and divulgation of personal data, that occur generally by *“state surveillance (Snowden revelations); foreign state surveillance (Snowden again); international data mobility (cloud services, etc); malicious hacking (markets in stolen personal data, botnets); and unintended data breaches.”*

There are risks in the use of predictive technologies, according to Hartzog and Selinger (2013, p. 252), *“the relegation of decisions about an individual’s life to automated processes based on algorithms and artificial intelligence raises concerns about discrimination, self-determination, and the narrowing of choice.”*

In a world of information, control and automation, with massive data systems, it is harder to save individuals privacy, and the own individual, which became a product in the internet market – following the maxim: when the service is free, the user is the product. As Greenleaf remembers (2014, p. 641), *“the aggregation of information about individuals for the purpose of on selling that information for other marketing activities has come to predominate.”*

Hartzog and Selinger (2013) reaffirm this concern in other way: big data technologies benefit organizations and not individuals, because the latter are the products – after all, if one is not paying for a service, she is not a consumer, but the product. In online interactions personal information are exchanged for free services. From there, organizations start to know all individuals preferences, offering products that relate to it, among others possibilities. Organizations are not prepared to share the wealth created by personal information with individuals.

Crişan, Zbucnea and Moraru (2014) affirm that online data safety is no longer possible. Every individual’s online activity is submitted to monetization – in the model that he is no longer a consumer, but the product. The authors defend that companies and individuals should be more responsible with sensible personal information, knowing the impossibility of online safety.

On the other hand, privacy right systems continue to be reinforced. In Europe, for example, the Regulation 12 and the Convention 108 have the function to fight menaces to privacy, including those outside the continent, mainly through *“more explicit data minimization, the ‘right to be forgotten’, data portability, ‘privacy by default’, stronger extraterritoriality, local representative requirements, and fines proportional to business size.”*(GREENLEAF, 2014, p. 642) Attempts like that. According to Greenleaf (2014), are often unsuccessful, mainly because of the ability of American companies, legally constituted or not, to purchase, process and use personal data from around the world without restrictions

“Privacy standards in other countries do not matter much if personal data can be liberated to the US ‘Safe Harbor’.” (GREENLEAF, 2014, p. 642)

The world is in permanent tension: territorialization and deterritorialization, control and liberty. The reigning model in today’s information economy is the user-product. One must remember that there are still places where this model does not apply, but this does not mean that it isn’t the dominant model – the exception confirms the rule. Greenleaf (2014, p. 641-642) advise that the dominant model is not necessarily permanent, but for that a second internet bubble must burst, or *“a concerted effort by the rest of the world to reject privacy invasive business practices. Neither is impossible, nor likely to occur rapidly.”*

2.1.3. The paradox and big data ethics

On the paper *Three paradoxes of big data* Richards and King (2013) deal with three aporetic points about benefits and prejudices of big data technologies. The first paradox deals with transparency. Big data technologies promise to use the collected data to know the world better, but, at the same time, promote invisible collections of data, using tools involved in mystery. The author ask: *“If big data spells the end of privacy, then why is the big data revolution occurring mostly in secret?”* (RICHARDS; KING, 2013, p. 42)

The second is the identity paradox. Richards and King (2013) affirm that at the same time it intends to identify, big data also threatens identity. In the majority of liberal democracies people have to right to choose about their own identity – each person can chose what wishes to be (at least under the enchanted veil of the rule of law). By combining all kinds of data, it is possible that identity choices become an imposition. “I am” becomes “you are”, and “I like” becomes “you like”.

The last one is the power paradox. It is told that big data technologies were designed to benefit all. In this model there are those who will have more benefits, and those who will have less. Richards and King (2013) affirm that those who will have more benefits will be institutions that control the new tools over individuals, who have their data constantly mined, analyzed and grouped. The penumbral area of technology, without clear legal limitations and opaque technical limits, makes individuals and institutions move themselves in shifting grounds. Many individuals remain under the institutional discourse of technology as a *Deus ex machina*, of free services for good will and safety in online environments. Organizations continue to use personal information under the risk of scandals. The result of this relation is *“an uneasy, uncertain state of affairs that is not healthy for anyone and leaves individual rights eroded and our democracy diminished.”* (RICHARDS; KING, 2013, p. 45)

Richards and King (2014) affirm that big data technologies increase the power of organizations, mainly in the occident, a fact that requires the development of a big data ethics. Human concepts like identity, privacy, freedom of choice, transparency and confidentiality must be balanced with the use of big data, under the risk they cease to exist. This is the mission of a big data ethics.

This leads to the development of a big data ethics as four principles that should govern the flow of data in the society based on information, control and automation. In first place, privacy must be recognized as a rule in the flow of information. In second place, it must be recognized that sensible personal data shared may stay private. In third, the necessity of transparency in big data practices. In fourth, the authors affirm the necessity of recognition that big data may compromise individual’s identity *“by allowing institutional surveillance to identify, categorize, modulate, and even determine who we are before we make up our own minds.”* (RICHARDS; KING, 2014, p. 396). Although the construction of this kind of ethics is necessary, what is seen is that the data flow continues to follow its path to control.

Conclusion

Firstly this paper had the objective to investigate a data driven society, founded in information, control and automation, mainly based on the works of Gilles Deleuze and Félix Guattari. On the Deleuzian model, contemporary society is founded on control. The evolution of information technology in recent decades and the emergence of the information era are trademarks of what can be referred as society of control, which has for objective the constant modulation of individuals for its purposes – a permanent production of subjectivity. Deleuze provides an overview where control mechanisms are continuously exercised over subjects, for the purpose of modulating their behavior and actions. The narrative resembles a dystopian future imagined by Orwell or Huxley, but it is a raw presentation of how data driven societies function, based on information, control and automation technologies.

Secondly the paper investigated an example of these technologies – big data, mainly concerning how big data technologies may transform legal practice in contemporary society. The fast evolution of information and communication technologies in the last ten years increased the possibility to generate content in a way never seen before. It is estimated that from the invention of writing until 2006 humanity accumulated about 180 Exabytes of data. This number grown ten times between 2006 and 2011, reaching 1600 Exabytes. How it is possible to navigate more efficiently in this data ocean recreated daily? Big data technologies were created to search for correlations in this ocean of data, in order to extract useful information, using inductive logic used to enable efficiency and speed, in opposition to classical deductive logic.

On one hand, there are possibilities for the use of big data technologies in order to assist in the legal decision-making process, as well as the reverse is possible. It is necessary to understand the structure and functions of big data technologies deeply, in order to enable its use to help human decision-making. Big data technologies also enable a new form of legal intervention – the creation of customized standard rules in areas that could benefit from these legal constructions, such as contract law. On the other hand, these technologies may create new forms of discrimination, and also increasingly threaten privacy and even the personal identity construction process. There are paradoxes in these technologies that may suffer less stress with the construction of a big data ethics. As that old English rock band warned: *welcome to the machine*.

The present paper has no pretension to be the correct mirror of nature. It is instigation to debate, to dissent and dialog. The positions shown here cannot be seen, and have no pretension to be, the natural truth. All theories are models of reality – work like a map. Reality is extremely complex; therefore a reduction is needed in order to understand it. A theory is a reduction of a complex reality, in order to understand it somehow. Just like a map, it permits to go to a destination. But a theory cannot be mistaken with reality, as a map cannot be mistaken with the city – if the map were the city, it would be the city, and not a map – and this counts for theories also.

Every theory, as every map, has the goal to situate a subject in a complex reality. Physics permits to understand the functioning of certain natural phenomena, as do chemistry; math enables to reach proved results; theory of law permits to situate a subject in a reality of social regulation.

In case a theory cannot reach its goal to guide subjects in a complex space – it must be thrown away, searching for a more adequate for that task. There is a problem in ontologizing theories, in other words, mistake them for reality, as this discard process becomes much more difficult.

References:

AKERKAR, Rajendra. *Big Data Computing*. London: CRC, 2014.

BOYD, D. and K. Crawford. *Critical questions for big data*. Information, Communication & Society 15(5): 662–679, 2012.

BURDON, M.; HARPUR, P. D. *Re-Conceptualising Privacy and Discrimination in an Age of Talent Analytics*. UNSW Law Journal, Sydney, Australia, v. 37, n.2, p.679-712, 2014.

BURROUGHS, William S. *The Limits of Control*. Semiotext(e): Schizo-Culture, v. III, n. 2, p. 38-42, 1978.

CARNAP, Rudolf. *Logical Foundations of Probability*. Chicago: University of Chicago Press, 1963.

CASTELLS, Manuel. *Era da informação: Economia, Sociedade e Cultura. A sociedade em rede, v. 1*. 2 ed. Tradução: Roneide Venancio Majer. São Paulo: Paz e Terra, 1999.

CRİŞAN, C., ZBUCHEA, A., MORARU S. *Big Data: The Beauty or the Beast*. Strategica: Management, Finance, and Ethics, Bucareste, n. 2, 2014.

DEAN, J.; S. GHEMAWAT. *MapReduce: Simplified data processing on large clusters*. Communications of the ACM 51(1): 107–113, 2008.

DELEUZE, Gilles. “*Post scriptum sobre as sociedades de controle*” In *Conversações 1972-1990*, p. 219 – 226, Tradução Peter Pál Pelbart, São Paulo: Editora 34, 2008.

DELEUZE, Gilles; GUATTARI, Felix. *Mille Plateaux: Capitalisme et schizophrénie 2*. Éditions de Minuit: Paris, 1980.

DESCARTES, Rene. *Discours de la méthode*. Paris: Mozambook, 2001.

DESCARTES, Rene. *Discours de la méthode*. Paris: Mozambook, 2001.

DONEDA, Danilo. *Da privacidade à proteção de dados pessoais*. Rio de Janeiro: Renovar, 2006.

DRAPER, R. N.; BOX, G. E. P. *Response Surfaces, Mixtures, and Ridge Analyses*. 2 ed. New Jersey: John Wiley & Sons, 2007.

DRUMMOND, Victor. *Internet, Privacidade e Dados Pessoais*. Rio de Janeiro: Lumen juris, 2003.

ERMOLAYEV V., AKERKAR R., TERZIYAN V, COCHEZ M. *Toward Evolving Knowledge Ecosystems for Big Data Understanding*. In: AKERKAR, Rajendra (Ed.). *Big Data Computing*. Sognal, Norway: CRC, 2014.

FAN, W.; BIFET, A.; YANG, Q.; YU, P. *Foreword*. In *Proc First Int Workshop on Big Data, Streams, and Heterogeneous Source Mining: Algorithms, Systems, Programming Models and Applications*. New York: ACM., 2012.

FERGUSON, A. G. *Big Data Distortions: Exploring the Limits of the ABA LEATPR Standards*. Oklahoma Law Review, Oklahoma, v. 831, n. 66, 2014.

FISHER, D.; DELINE R.; CZERWINSKI M.; DRUCKER S. *Interactions with big data analytics*. Interactions 19(3):50–59, 2012.

FLORIDI, Luciano. *Big data and Their Epistemological Challenge*. In: Philosophy and Technology. Vol. 25, issue 4. Springer Science, 2012.

FOUCAULT, Michel. *Vigiar e Punir: nascimento da prisão*. 20 ed. Tradução Raquel Ramallete. Petrópolis: Vozes, 1999.

_____, *Microfísica do Poder*. Organização e tradução: Roberto Machado. Rio de Janeiro: Graal, 2009.

GREENLEAF, Graham. *Abandon all Hope?* UNSW Law Journal, Sydney, Australia, v. 37, n.2, p.636-642, 2014.

GRELLER, W. *Reflections on the knowledge society*. <http://wgreller.wordpress.com/2010/11/03/big-data-isnt-big-knowledge-its-big-business/>, 2012. (acessado em 18 de julho de 2014).

HARTZOG, W.; SELINGER, E. *Big data in small hands*. Stanford Law Review, Califórnia, v. 66, n. 2, 2013.

HOFFMAN, S.; PODGURSKI, A. *Big Bad Data: Law, Public Health, and Biomedical Databases*. Journal of Law, Medicine and Ethics, Massachusetts, n. 34, 2012.

HUME, David. *An inquiry concerning human understanding*. New York: Oxford University Press, 2007.

KATZ, D. M. *Quantitative Legal Prediction – or – How I Learned to Stop Worrying and Start Preparing for the Data Driven Future of the Legal Services Industry*. Emory Law Journal, Atlanta, v. 62, n. 2, 2013.

LIN, J.; DYER, C. *Data-Intensive Text Processing with MapReduce*. Morgan & Claypool Synthesis Lectures on Human Language Technologies, 2012.

LOEVINGER, Lee. *Jurimetrics: The Next Step Forward*, Minnesota Law Review, Minnesota, v. 455, n. 33, 1949.

MALLE, Jean-Pierre. *Big Data: farewell to Cartesian thinking?* <http://www.paristechreview.com/2013/03/15/big-data-cartesian-thinking/>, 2013.

MANYIKA, J.; CHUI, M.; BROWN, B.; et al. *Big data: The next frontier for innovation, competition, and productivity*. McKinsey Global Institute, 2011.

McSHANE B. B.; WATSON. *Predicting Securities Fraud Settlements and Amounts: A Hierarchical Bayesian Model of Federal Securities Class Action Lawsuits*. Journal of Empirical Legal Studies, Pennsylvania, v. 9, n.1, p. 482-505, 2012.

MOSES, L. B.; CHAN, J. *Using Big Data For Legal And Law Enforcement Decisions: Testing The New Tools*. UNSW Law Journal, Sydney, Australia, v. 37, n.2, p.643-678, 2014.

MUPHY, Tonia Hap. *Mandating Use of Predictive Coding in Electronic Discovery: An Ill-Advised Judicial Intrusion*. American Business Law Journal, Chicago, v. 50, n. 3, p. 609-657, 2013.

PORAT, A.; STRAHILEVITZ L. S.; *Personalizing Default Rules and Disclosure with Big Data*. Michigan Law Review, Michigan, n. 112, 2014.

PUGLIESI, Marcio. *Teoria do Direito*. 2 ed. São Paulo: Saraiva, 2009.

REICHENBACH, Hans. *Erfahrung und Prognose*. Springer Fachmedien Wiesbaden, 1983.

RICHARDS, N. M.; JONATHAN H. *Three Paradoxes of Big Data*. Stanford Law Review, Califórnia, v. 41, n. 2, 2013.

_____, *Big Data Ethics*. Wake Forest Law Review, Winston-Salem, v. 89, n. 4, 2014.

RODOTÁ, Stefano. *A Vida na Sociedade da Vigilância: A privacidade hoje*. Rio de Janeiro: Renovar, 2008

RUGER, T. W.; KIM, P. T.; MARTIN, D. A.; QUINN, K. M. *The Supreme Court Forecasting Project: Legal and Political Science Approaches to Predicting Supreme Court Decisionmaking*. Columbia Law Review, New York, v. 1150, n. 104, 2004.

SCHREIBER, Anderson. *Direitos da Personalidade*. Ed. Atlas, São Paulo, 2011.

SCHROFF, Gautam. *The intelligent web: search, smart algorithms and big data*. Oxford: Oxford university press, 2013.

SKYRMS, Bryan. *Escolha e acaso*. São Paulo: Cultrix, 1966.

SOLOVE, Daniel J. *Privacy Self-Management and the consent dilemma*: Harvard Law Review, vol. 126. MA: Harvard University Press, 2013.

TYREE, Alan. *Expert Systems in Law*. Australia: Prentice Hall, 1989.

WARREN, Samuel; BRANDEIS, Louis. *The right to privacy*. Cambridge, MA: Harvard Law Review, Vol IV, N. 5, 1890.