http://dx.doi.org/10.18623/rvd.v17i37.1795

THE DIGITAL ENVIRONMENT AND THE RIGHT TO PRIVACY IN RELATION TO *BIG DATA*

Émilien Vilas Boas Reis[1]
Escola Superior Dom Helder Câmara (ESDHC)

Bruno Torquato de Oliveira Naves[2]
Escola Superior Dom Helder Câmara (ESDHC)

ABSTRACT

We currently live with a large volume of data – *Big Data* –, which represents a revolution in the means of commercialization, advertising, competition, and even in the very delimitation of the private space. In digital environments, privacy must be considered under the possibility of much more serious consequences in the event of a breach. Thus, this article proposes to analyze the legal regulation of the digital environment, especially regarding the right to privacy. Bibliographic research (theoretical-qualitative) has been carried out to this end, with argumentative conclusions that allowed the perception of the right to privacy with greater fluidity in the face of personal data, bringing it closer to legal principles. The regulation carried out by the General Law for the Protection of Personal Data, particularly ambivalence of sensitive personal data, has been analyzed.

Keywords: *Big Data*; digital environment; General Law for the Protection of Personal Data; privacy; sensitive data.

1 Post-doctorate in Philosophy from the Universidade do Porto (UP). PhD in Philosophy from the Pontifícia Universidade Católica do Rio Grande do Sul (PUC-RS). Master in Philosophy from PPUC-RS. Graduated in Philosophy from the Universidade Federal de Minas Gerais (UFMG). Adjunct professor at ESDHC at undergraduate and graduate level (Master's and Doctorate). ORCID: http://orcid.org/0000-0003-0729-522X / email: mboasr@yahoo.com.br

2 Doctor and Master in Law from the Pontifícia Universidade Católica de Minas Gerais (PUC-MINAS). Professor in the Master's and Doctorate courses in Environmental Law and Sustainable Development at ESDHC. Coordinator of the specialization course in Urban and Environmental Law at PUC-MINAS Virtual. Professor at PUC-MINAS. Researcher at the Centro de Estudos em Bioética (CEBID). ORCID: https://orcid.org/0000-0001-8329-9420 / email: brunotorquato@hotmail.com

O MEIO AMBIENTE DIGITAL E O DIREITO À PRIVACIDADE DIANTE DO BIG DATA

RESUMO

Convive-se, hoje, com um grande volume de dados – Big Data –, o que representa uma revolução nos moldes de comercialização, propaganda, concorrência e, até mesmo, na própria delimitação do espaço privado. A privacidade no meio ambiente digital deve ser considerada sob a possibilidade de consequências muito mais gravosas, no caso de violação. Assim, este artigo propõe-se a analisar a regulação jurídica do meio ambiente digital, sobretudo frente ao direito à privacidade. Para tanto, realizou-se pesquisa bibliográfica (teórico-qualitativa), com conclusões argumentativas que permitiram perceber o direito à privacidade com maior fluidez diante dos dados pessoais, aproximando-o dos princípios jurídicos. Analisou-se a regulamentação realizada pela Lei Geral de Proteção de Dados Pessoais, em especial a ambivalência dos dados pessoais sensíveis.

***Palavras-chave**: Big Data; dados sensíveis; Lei Geral de Proteção de Dados Pessoais; meio ambiente digital; privacidade.*

INTRODUCTION

In addition to any Platonic conception, we currently live between two worlds: the concrete and the digital. The individual is born, built, and establishes relationships in both worlds. In the digital world, however, one is immortal.

Digital data is historically persistent. It seems that digital data will remain as long as humanity exists. There is even talk of a testament to digital assets to address the legal destination of voluntary events on social media such as *Facebook*, *Instagram*, and *Twitter* accounts, which accumulate personal data and have intangible value.

The large volume of information digitally available is what we widely refer to as *Big Data*. A set of data from all categories and formats makes up a universe of multiple interests. But how prepared is the Law to mediate all these – sometimes opposed – interests? Several countries are now concerned about digital environments' legal regulation.

There is a worldwide movement concerning data traffic, with means to determine acceptable ways of handling, controlling, and sharing this information. In 2016, the European Union passed Regulation No. 679, known as the General Data Protection Regulation (GDPR), aimed at the unification of data policies across countries. Brazil has more recently joined this movement with the passing of the General Law for the Protection of Personal Data (Lei Geral de Proteção de Dados Pessoais – LGPD).

It is clear that data technology is challenging legal norms and pre-established views, which reflects in the transformation of the very conception of people's private life.

This article analyzes how privacy performs in a context in which information whose circulation is not always intentional is spreading. This is a theoretical-qualitative research undertaken through bibliographic sources that contributed to ethical and legal issues brought about by the relationship between *Big Data* and Artificial Intelligence (A.I.) to investigate the new conformations of the right to privacy.

1 CONSIDERATIONS ON *BIG DATA*

Humanity is currently experiencing a great revolution in terms of information and its uses. At no other time has so much data been produced, which is accumulating on the Internet.

> Google processes over 40,000 searches every second; totaling 3.5 billion searches a day. Every minute, Snapchat users share 527,760 photos, while YouTube users stream over than 4.1 million videos. And there are also old systems, such as emails, that continue to experience significant growth. 156 million messages are sent every minute.
>
> Another aspect that needs to be considered, however, is that companies and machines also generate huge amounts of data. According to a survey by Statista, the number of sensors should reach 12.86 billion by 2020.
>
> In light of all of this, it seems plausible that data volumes will continue to increase rapidly. In a report by the International Data Corporation (IDC) called "Data Age 2025" ("Data 2025"), the amount of data created by 2025 is expected to reach a staggering 163 zettabytes. This is approximately ten times the amount of data available in 2017 (TAULLI, 2020).

After the invention of the computer, computational languages, the Internet, and social media, the world is faced with a new term used to illustrate the ability to deal with such an unprecedented accumulation of data. This expression is *Big Data*. The idea behind the term is that the endless information available and accumulated, and now available within the *online* universe, will be useful in several areas.

> Things really are speeding up. The amount of stored information grows four times faster than the world economy, while the processing power of computers grows nine times faster. Little wonder that people complain of information overload. Everyone is whiplashed by the changes (MAYER-SCHONBERGER; CUKIER, 2017).

One of the most intriguing examples of *Big Data*'s capacity was highlighted in the text *Detecting influenza epidemics using search engine query data*, in which Google researchers (GINSBERG et al., 2009) from Google's own search queries sought to track the advance of influenza (H1N1) in the United States.

Disease surveillance systems, such as the *Centers for Disease Control and Prevention* (CDC) and the *European Influenza Surveillance Scheme* (EISS), make control reports that do not come out with immediate results. To accelerate this process, the authors used the analysis of Google query data related to influenza for five years. The data collected comprised the years 2003 and 2008 and considered the 50 million most searched terms in the United States during the period of greatest advance of the disease.

Ginsberg et al. (2009) They took the most searched terms on Google and compared them with CDC's data on the spread of influenza in each of the nine regions divided by CDC, as well as with outpatient consultations. The authors were trying to create a forecasting model based on the

relationship between Google searches and other available data. By using their prediction model throughout the flu season between 2007 and 2008, the researchers were able to estimate the spread of the flu one to two weeks before the official CDC report. The realization was that search queries on Google could 'anticipate' and 'predict' the spread of flu before official agencies could do it and before an epidemic could spread. In short:

> Harnessing the collective intelligence of millions of users, Google web search logs can provide one of the most timely, broad-reaching influenza monitoring systems available today. Whereas traditional systems require 1–2 weeks to gather and process surveillance data, our estimates are current each day (GINSBERG et al., 2009, p. 1014).

Data (information) amounts have grown exponentially with each year. *Big Data* uses them to make predictions. The data can be read in: a) *bit* (short form of *Binary digit*), which expresses itself through 0 or 1, being used to measure the amount of data available within the network; b) *byte* (which means storage – e.g. a file is X *Megabytes*) (TAULLI, 2020). On the other hand, data originates from different sources:

- Web/social media (Facebook, Twitter, Instagram, YouTube)
- Biometric data (physical activity trackers, genetic tests)
- Point of sale systems (physical stores and e-commerce websites)
- Internet of things or IoT (ID tags and smart devices)
- Cloud systems (business applications such as Salesforce.com)
- Corporate databases and spreadsheets (TAULLI, 2020).

It is also possible to identify the data from your organization: (a) structured data, which is stored in databases or spreadsheets, and is small in volume (e.g., financial information, addresses, product information, telephone numbers); (b) unstructured data, which has no predefined formatting (e.g., images, videos, audio and text files, information from social media, etc.); (c) semi-structured data, which contains both structured and unstructured data (e.g. XML – *Extensible Markub Language*, JSON – *JavaScript Object Notation* –, which are application programming interfaces); and (d) temporal data, which can be structured, unstructured, and semi-structured and comprises interactive information (e.g., the gathering of information about a customer who visits a *website* or makes use of an application) (TAULLI, 2020).

In 2001, Doug Laney, of the Gartner consulting, firm wrote the piece *3D Data Management: Controlling Data Volume, Velocity, and Variety*, in which he suggested that data should be analyzed based on its volume,

speed, and variety. Regarding the volume, Laney had already manifested concern regarding the considerable increase in data, which meant that the velocities should also increase to allow for dealing with a large amount of data. Finally, the variety of data, which should also be considered due to the different types of information sources (LANEY, 2001).

Although he was a pioneer, data analysts said Laney's proposal was insufficient. Therefore, they proposed other elements that should be considered regarding the data.

> The following are some of the most common:
> • Veracity: refers to data deemed accurate. In this chapter, we will discuss some of the techniques used for assessing veracity.
> • Value: demonstrates data usefulness. It often relies on having a reputable source.
> • Variability: means that data will usually change over time. This is the case, for example, with social media content, which may change based on the general public's opinions regarding new developments and breaking news.
> • Visualization: deals with the use of visual resources – such as graphs – to better understand data (TAULLI, 2020).

Big Data may have three main characteristics: a) it allows for the analysis of large amounts of data; b) intends to organize the real world; and c) deals with correlation and not causality (MAYER-SCHONBERGER; CUKIER, 2017).

This is illustrated by Google researchers and seasonal flu's example. More than 50 million searched words were analyzed (incidence, location, time, etc.), and, from there, the researchers sought to organize the chaos of information, without worrying about the reason why the flu was spreading. Data analysis depends on the ability to collect, store, and process data. And while data can be passively collected, biases can be avoided.

The large amount of data analyzed is what makes *Big Data* different from what has taken place with information manipulation so far. Technological advances are important, as well as better algorithms, but it is the sheer amount of information that has been making *Big Data* more predictable with every day. For example, computers started beating humans in chess games, which has followed the same rules for centuries, because the system fed on more data over the years (MAYER-SCHONBERGER; CUKIER, 2017).

The chaos of the reality is covered by *Big Data*, as only 5% of digital data is structured. The large amount of data means that the 95% unstructured data are also apprehended:

> According to some estimates only 5 percent of all digital data is "structured" – that is, in a form that fits neatly into a traditional database. Without accepting messiness, the remaining 95 percent of unstructured data, such as web pages and videos, remain dark. By allowing for imprecision, we open a window into an untapped universe of insights (MAYER-SCHONBERGER; CUKIER, 2017).

The results achieved by *Big Data* are not related to reason and causality. The data is used to search for correlations and probabilities. We are moving towards the possibility of predicting events that have not yet occurred. The scientific method itself, which starts from unproven hypotheses, is at risk of being questioned.

> By letting us identify a really good proxy for a phenomenon, correlations help us capture the present and predict the future: if A often takes place together with B, we need to watch out for B to predict that A will happen. Using B as a proxy helps us capture what is probably taking place with A, even if we cannot measure or observe A directly. Importantly, it also helps us predict what may happen to A in the future. Of course, correlations cannot foretell the future, they can only predict it with a certain likelihood. But that ability is extremely valuable (MAYER-SCHONBERGER; CUKIER, 2017).

Big Data's scope also includes the current ability to quantify phenomena, as well as the fact that they are currently inserted in the *online* world. More than ever, Galileo's saying that "God writes the world in mathematical characters," proves to be correct. Books, geolocators, social relations, everything is being transformed into data. Using data collected from Twitter (social media), Cornell University researchers managed to analyze the behavior of individuals in different countries and cultures on a global scale. The authors used as data 509 million messages from 2.4 million people worldwide:

> Using Twitter.com's data access protocol, we collected up to 400 public messages from each user in the sample, excluding users with fewer than 25 messages. The resulting corpus contained about 2.4 million individuals from across the globe and 509 million messages authored between February 2008 and January 2010 (GOLDER; MAC, 2011, p. 1879-1880).

The researchers identified some common characteristics among those surveyed: people wake up in a good mood, but this feeling changes as the day goes by, a phenomenon related to the circadian rhythm and sleeping. They also found that people are happier on weekends, and that, because they wake up later on those days, peak positive affectivity occurs two hours later than during the week (GOLDER; MAC, 2011).

With *Big Data*, the world can be perceived as a series of information that must be worked on as data, and that is true even for human sciences.

Another important aspect to highlight is that the data is not always voluntarily provided. Much data is captured without its owners' consent through means unknown by them:

> Data can frequently be collected passively, without much effort or even awareness on the part of those being recorded. And because the cost of storage has fallen so much, it is easier to justify keeping data than discarding it. […] Websites log every click users make – sometimes even where the mouse-cursor moves – for analyzing and optimizing the content the sites present to visitors (MAYER-SCHONBERGER; CUKIER, 2017).

Current *websites* can record users' clicks, as well as store information based only on the movement of their *mouse*. In the *online* world, any user action may be subject to the ability to record data, and even if such data seems useless at first glance, within the boundless data storage and through relationships with other data, it becomes useful. "Data exhaust" is the name given by theorists to tracks left on the Internet:

> A term of art has emerged to describe the digital trail that people leave in their wake: "data exhaust." It refers to data that is shed as a byproduct of people's actions and movements in the world. For the Internet, it describes users' online interactions: where they click, how long they look at a page, where the mouse-cursor hovers, what they type, and more (MAYER-SCHONBERGER; CUKIER, 2017).

There is also the data collected from the *offline* world, such as when retailers, through cameras, track their customers in physical stores, capturing things as specific as to what their customers are looking at. With this information, store owners can manipulate data in favor of better *marketing* campaigns (MAYER-SCHONBERGER; CUKIER, 2017).

Data that was previously collected for a reason, when stored, can be reused for other future purposes. Large amounts of data are not simply discarded. One of the questions that can be asked with *Big Data* is: to the extent that the data is accurate about things, what might happen to human privacy and freedom?

Large private companies and governments are in possession of this big data:

> The Internet has made tracking easier, cheaper, and more useful. And clandestine three-letter government agencies are not the only ones spying on us. Amazon monitors our shopping preferences and Google our browsing habits, while Twitter knows what is on our minds. Facebook seems to catch all that information too, along

> with our social relationships. Mobile operators know not only whom we talk to, but who is nearby (MAYER-SCHONBERGER; CUKIER, 2017).

The accumulation of personal information leads to questions of 'how', 'why', and 'by whom' this data may be used. These are questions raised by *Big Data* that, perhaps, can no longer be avoided, given that there is a worldwide trend towards increasing data spending:

> The amount of money invested in data is enormous. According to the IDC (International Data Corporation), spendings on Big Data and analytics solutions are forecast to rise from $ 166 billion in 2018 to $ 260 billion by 2022. This represents an annual growth rate of 11.9%. The largest investors include banks, discrete and process manufacturers, professional services companies, and the Federal Government. They account for almost half of the total investment (TAULLI, 2020).

The data, in turn, depend on another fundamental area for its use: Artificial Intelligence.

2 CONSIDERATIONS ON ARTIFICIAL INTELLIGENCE (A.I.)

Despite being present in the popular imagination, the complete automation of robots, which will take the world for themselves and either destroy or enslave humanity, Artificial Intelligence still does not match this description. However, it is another important part of the understanding of data usage.

A.I. has a long story, having Alan Turing (1912–1954) as its precursor[3]. Current A.I.'s momentum, in turn, depended on the following factors: 1) explosive growth of *datasets* (data set): an element addressed on the previous topic; 2) infrastructure: which has been driven by Google in the past 15 years; 3) Graphics processing units (GPUs – Graphics Processing Units), chip technology created by the technology company NVIDIA (TAULLI, 2020).

However, there are other elements to consider when talking about A.I. A first-term is machine learning. In 1959, Arthur L. Samuel published the article *Some Studies in Machine Learning Using the Game of Checkers*. He, who worked at IBM, put in an article what he had already done in practical ways as a programmer for a checkers game for computers: the notion of how a machine, based on advanced statistics, would have the ability to "learn" something, making it perfect. In his words:

3 For a brief resumption of the historical path of A.I. to the present-day cf. TAULLI, 2020.

> Enough work has been done to verify the fact that a computer can be programmed so that it will learn to play a better game of checkers than can be played by the person who wrote the program. Furthermore, it can learn to do this in a remarkably short period of time (8 or 10 hours of machine-playing time) when given only the rules of the game, a sense of direction, and a redundant and incomplete list of parameters which are thought to have something to do with the game, but whose correct signs and relative weights are unknown and unspecified. The principles of machine learning verified by these experiments are, of course, applicable to many other situations (SAMUEL, 1983, p. 211).

The algorithm, which are procedures for doing something, is the language used by *machine learning*. Even though there are countless of them, they can be covered in 4 categories: 1) Supervised learning: deals with labeled data, for example, an identified photo; 2) Unsupervised learning: deals with unlabeled data, thus looking for patterns in data, for example, customer groups; 3) Reinforcement learning: a process that occurs through trial and error, making learning better over time. It has been instrumental in improving A.I.; 4) Semi-supervised learning: takes into account supervised and unsupervised language. In this case, the unsupervised language is transformed into supervised language through *deep learning* (TAULLI, 2020).

By the way, *deep learning* is an area of *machine learning*. It is a system that processes countless data to find certain patterns. *Deep learning* has been taken as the fundamental element for A.I. improvement. One of the sources of inspiration for *deep learning* is the thought that the brain is a machine (computer) and, therefore, thinking about the computer system in the light of brain functions would make perfect sense (TAULLI, 2020).

The basic structure for a *deep learning* model is the *Artificial Neural Network* (ANN):

> […] a function that includes units (which can also be called neurons, perceptrons, or nodes). Each unit will have a value and weight, which indicate its relative importance and then move into the hidden layer. The hidden layer uses a function, the result of which becomes the output (TAULLI, 2020).[4]

One of the techniques used to improve *deep learning* is retroprogramming, which consists of improving the neural network using the errors found in it, inserting new values into the network (TAULLI, 2020).

There are different types of neural networks, notably: (1) connected neural networks – the most basic network, where there are connections between neurons; (2) *recurrent neural networks* (RNN) – a network that, in addition to processing current entries, also processes previous entries, for

4 The hidden layer is the level of *deep learning* analysis.

example, automatic word search; 3) *convolutional neural networks* (CNN) – a network that takes into account different convolutions (variations) of data analysis, aggregating and combining different data, for example, face recognition; 4) *generative adversarial networks* (GAN) – a network that creates new types of output (e.g. videos and audios) from input data (TAULLI, 2020).

When it comes to A.I., the *Robotic Process Automation* (RPA), which is not related to physical robots but software robots, is another important item. *Bots* (diminutive of robot) reproduce repetitive work (TAULLI, 2020).

Thus, as we have been introduced to the notions of *Big Data* and A.I., our next item shall cover the ethical challenges involved.

3 ETHICAL AND LEGAL CHALLENGES CONCERNING *BIG DATA*, ALGORITHMS, AND ARTIFICIAL INTELLIGENCE

The contemporary world makes way to practical problems that force different reflections concerning discoveries. The combination of *Big Data*, Artificial Intelligence, and algorithms brings questions that are hard to answer. The first chapters purposely avoided emphasizing such problems, but they are now opportune.

The use of data, often without the authorization of those involved, the dependence on algorithms, the exclusion of human work along with the new technological processes, the end of certain careers, the questioning about the existence of a free will, the difficulty faced by the Law in dealing with new challenges, and the reduction of responsibility are some of the issues found.

An interesting point that current times bring to light is that ethical reflection should not occur with a focus on the devices themselves (*hardware*), but the use of data, programmed algorithms, as well as Artificial Intelligence, therefore, it should take place at the *software* level:

> It is not the hardware that causes ethical problems, it is what the hardware does with the software and the data that represents the source of our new difficulties. LoAD brings into focus the different moral dimensions of data. In doing so, it highlights the fact that, before concerning information, ethical problems such as privacy, anonymity, transparency, trust and responsibility concern data collection, curation, analysis and use, and hence they are better understood at that level (FLORIDI; TADDEO, 2016, p. 3).[5]

Luciano Florido, a prominent scholar on data ethics, claims that this branch of ethics dwells on moral problems related to data, including its

5 LoAD (*level of abstraction data-centric*)

generation, registration, processing, dissemination, etc. and on studies on algorithms, Artificial Intelligence, and possible conducts on the use of these techniques (FLORIDI; TADDEO, 2016).

Data ethics focuses on data collection and analysis, which includes the use of *Big Data* in areas as diverse as biomedical, social sciences, advertising, and others. Aspects such as data trust and transparency are also fundamental aspects:

> In this context, key issues concern possible re-identification of individuals through data-mining, –linking, –merging and re-using of large datasets, as well as risks for so-called 'group privacy', when the identification of types of individuals, independently of the de-identification of each of them, may lead to serious ethical problems, from group discrimination (e.g. ageism, ethnicism, sexism) to group-targeted forms of violence (FLORIDI; TADDEO, 2016, p. 3)

The ethics of algorithms is a sub-area that becomes fundamental due to the increased capacity of the algorithms and their increasing autonomy through Artificial Intelligence, as well as their ability to learn. It also includes the analysis of unwanted situations, such as discrimination. "In this case, some crucial challenges include moral responsibility and accountability by both designers and data scientists concerning unforeseen and undesired consequences, as well as missed opportunities" (FLORIDI; TADDEO, 2016, p. 3).

The practical ethical sub-area (professional deontology) deals with individual and corporate responsibility while dealing with data: "[…] Aimed at defining an ethical framework to shape professional codes on responsible innovation, development, and usage, which may ensure ethical practices fostering both the progress of data science and the protection of individual and collective rights" (FLORIDI; TADDEO, 2016, p. 3).

Data ethics encompasses data ethics itself, but also the ethics of algorithms and practical ethics, making it a transdisciplinary study.

4 DIGITAL ENVIRONMENT AND THE PREDICTIVE FUNCTION OF ALGORITHMS

In Environmental Law, among the didactic classifications of the environment, a new category can now be opened, pertaining to the digital environment, where new types of interpersonal relationships are locked, sometimes with different assumptions, from data either voluntarily, compulsorily, or illegally collected.

Celso Antônio Pacheco Fiorillo places the digital environment as a division of the cultural environmental environment:

> As a consequence, the cultural environment manifests itself in the 21st century in our country exactly in the face of a culture that undergoes several vehicles that reveal a new civilizing process specifically adapted to the information society, namely, a new way of living related to a culture of convergence in which radio, television, cinema, *video games*, the Internet, communications through landline and cellphone calls, etc., shape a new life that reveals a new facet of the cultural environment, that is: the digital environment (FIORILLO, 2012, p. 81).

The Internet integrates a significant part of the digital environment, but there are also *offline* data comprising important components of this space, regardless of the trend being the communication of all data through internal networks or the world wide web.

Due to the high amount of information available within the digital environment, human contact with this environment is done through algorithms, which are the instruments capable of "mining" this information or even establishing correspondences between them, and returning intelligible results based on the search criteria used. In short, their functions will be either descriptive or predictive, extracting representative patterns from the information.

The descriptive algorithmic function performs the data compilation and synthesis, whereas the predictive function aims to predict behaviors through data analysis.

An interesting example of a predictive function based on personal data is narrated by Charles Duhigg (2012), in "The Power of Habit."

Upon receiving coupons and product brochures for pregnant women at home, addressed to his teenage daughter, the angered father walked into a Target store in Minnesota to talk to their manager. As the girl was still in high school, his questioning was very reasonable, as it seemed that Target intended to influence her daughter to have a baby.

Embarrassed, the manager had to apologize.

Target is a large American chain, selling everything from supermarket products to department store items. The company wanted to figure out the habits of each individual buyer to target personalized advertisements. That is why they hired statistician Andrew Pole to analyze each of their customer's shopping preferences in 2002.

> If you use your Target credit card to buy a box of popsicles once a week, usually around half past six in the evening, and giant trash bags in the months of July and

> October, Target statisticians and computer programs determine that you have kids at home, tend to stop to buy food on the way home from work, and have a lawn that needs mowing in the summer, as well as trees that shed leaves in the fall. They will examine your other purchase patterns and notice that you buy breakfast cereal at times, but never milk – which means you must be buying milk somewhere else. That's why Target will send you discount coupons on skimmed milk and also on chocolate sprinkles, school supplies, garden furniture, rakes, and – as you're likely to want to relax after a long day at work – beer. The company will guess what you usually buy, and then try to convince you to buy it at Target. It can personalize ads and coupons sent out to each customer, although you may never realize that you received a different brochure from your neighbors in the mail (DUHIGG, 2012, p. 231).

New parents represent one of the largest niche markets. A 2010 survey by Target estimated that parents spend an average of $ 6,800 on baby products before their first birthday.

Based on buying habits statistics provided by future mothers – such as their name, spouse's name, and expected date of birth – Pole detected purchase patterns related to each stage of pregnancy and developed a prediction system. With a combination of products purchased, Pole could predict whether a woman was pregnant and what stage of pregnancy she was in, thus enabling the targeting of advertisements and coupons.

> Jenny Ward, a 23-year-old woman from Atlanta bought cocoa butter lotion, a bag big enough to serve as a diaper holder, zinc, magnesium, and a blue rug? There is an 87% chance she is pregnant, and that her delivery is scheduled for the end of August. Liz Alter, a 35-year-old woman from Brooklyn bought five packs of hand towels, a bottle of washing powder for "sensitive skin," baggy jeans, vitamins enriched with DHA, and lots of moisturizers? She has a 96% chance of being pregnant and is likely to give birth in early May. Caitlin Pike, a 39-year-old woman from San Francisco bought a $ 250 pram, but nothing else? She is probably shopping for a friend's baby shower. Besides that, her demographic data shows that she got divorced two years ago (DUHIGG, 2012, p. 236).

And that's how Target predicted that teenager was pregnant. The Minnesota store manager, after a few days, called the teenager's father to apologize again. Even more embarrassed, the father said he had a conversation with his daughter and found out she was indeed pregnant.

This case raises the question of information privacy and the changes that the market has undergone in recent years, with "tailor-made" product offers for each user.

5 RIGHT TO PRIVACY AND INTIMACY WITHIN DIGITAL ENVIRONMENTS

In modern times, reason brought raise to the need for precise concepts and classifications, but also made clear that the Law was bound to recognize individuality and the individual as its constructive agent.

Perceived as "that which allows us to define what is and what is not important to us" (TAYLOR, 1997, p. 47), identity enables the realization of the individual's potential according to their interests and convictions.

The plurality of man and the unfinished project of building his personality depending on notions of intimacy and privacy, which will reflect on autonomy as a determining element of human dignity.

Technological development is the product of this human autonomy in technique, but with it, the creators themselves may fall hostage to their work, given the risks to which they are exposed.

Thus, the individual's power of self-determination must also focus on stored data control and the process of decision making after this data is collected.

When it comes to data control, treatment, and sharing, the subjective right to prior and complete information on the procedure to be performed must be observed. This information makes up the autonomy itself, since its exercise requires, in addition to discernment, awareness of the situation, with the advantages and risks involved, especially when dealing with sensitive data.

There is data, such as those referring to health, that, upon collection, one must question whether even the holder is interested in knowing or not.

In fact, health data can only be collected, used, and preserved for the purposes of diagnosis and health care; medical, pharmacological, and other forms of scientific research, such as anthropological studies; forensic medicine and civil or criminal procedures, or other legal actions.

The issue of intimacy and privacy in data disclosure is extremely relevant for the analysis. As data is stored, its potential to reveal preferences, habits, thoughts, manifestations of human image, body, and health becomes well-known.

The rights to intimacy and privacy are guaranteed by the Brazilian Federal Constitution of 1988 through a provision in art. 5[6], item X. In the

6 "X – people's intimacy, private life, honor, and image are inviolable; being guaranteed the right to compensation for material or moral damage resulting from their violation."

Civil Code, private life in its double aspect of intimacy and privacy is protected by art. 21, which provides: "The private life of the natural person is inviolable, and the judge, at the request of the interested party, will enforce the means necessary to prevent or terminate an act contrary to this rule."

Intimacy is the sphere of projection of the individual in their inner relationship. The right to privacy is a larger personal circle, as it involves the individual's interpersonal relationships. While intimacy dwells within the person's most restricted compartment, with situations that one does not want to share, privacy portrays the individual's public, family, or social life, including the right to control the collection and use of personal data.

> The exercise of the personal right to private life is enforced in two main ways: a) a positive exercise, consisting in the free conduct of life itself, based on the sharing of ideas and thoughts, from public exposure, including availability, through legal affairs, images, objects, and personal manifestations; b) a negative exercise, revealed by the right to withdraw, to be kept secret, of not revealing personal information (NAVES; SÁ, 2017, p. 95).

The legal protection of intimacy and privacy should apply to every human being, whether born or unborn; with the possibility to extend beyond death, mainly in the digital environment.

A blurring of the old privacy frontiers, which are now more fluid and heterogeneous, can then be perceived. We are moving between expanding access to information and the need to restrict access to some data.

With that, one could question the legal nature formalizing the right to privacy. Traditional civil law classifies the right to privacy as a subjective right formalized through *legal rules* (DE CUPIS, 2004).

Contemporary Law Theory differentiates legal rules and principles, specified within the "legal norm" genre, and, regardless of its controversies, it has established that the legal principle is usually a more general norm, with more open characters, which will be complemented by the specific case. Thus,

> The principle always presents a different action from other legal norms since there are no preconditions for its application. Therefore, authors such as Klaus Günther understand that principles are norms whose conditions of application are not predetermined.
>
> Günther affirms that the *a posteriori* determination of the content is due to the fact the principle includes an *on-balance* ratio, in which all its conditions and application limits are comparatively evaluated within the specific case. Its characteristic fluidity leads to a basic need for legal application (NAVES; REIS, 2019, p. 156-157).

With the features imposed by *Big Data*, privacy seems to gain a more fluid character, with no conditions of application predetermined by the legislator, which leads it to the principled nature, densified by the factual and legal conditions for its content determination.

In 2018, Brazil passed Law No. 13,709 – General Law for the Protection of Personal Data (Lei Geral de Proteção de Dados Pessoais, LGPD) to provide on personal data privacy.

6 GENERAL LAW FOR THE PROTECTION OF PERSONAL DATA

The LGPD is relatively extensive, comprising ten chapters and sixty-five articles, aiming at the regulation of "the processing of personal data, also in digital media, by a natural person or legal entity under public or private law, to protect the fundamental rights of freedom and privacy and the free development of the natural person" (art. 1).

Personal data is information that identifies and characterizes the natural person and their manifestations, therefore, they are included under the rights of personality, covering the portrait-image, the attribute-image, the voice-image, and the expression of thought.

> The *portrait-image* refers to the holder's physiognomic characteristics, the representation of a person by their visual aspect; in short, it refers to their poster or photograph, both static – a painting – and dynamic – a film –, according to the protection provided by art. 5, item X, of the Constitution of the Republic. Differently, the *attribute-image* is the natural consectarian of life in society, consisting of a set of peculiar characteristics of a person's presentation and social identification. Thus, it concerns their social qualifications, which are repeated behaviors that allow for their identification. It does not concern the external image, providing a person's moral portrait. The *voice-image*, on the other hand, concerns the identification of a person through their sound timbre. Without a doubt, someone's personality is no less evident in their voice than in their physiognomic characteristics (FARIAS; NETTO, 2017, p. 409-410).

The Law established the 'personal data' genre and, in it, highlighted the 'sensitive data' species, which relate to more private and intimate aspects of the individual, such as information "on racial or ethnic origin, religious belief, political opinion, union or religious affiliation, philosophical or political nature, data related to health or sexual life, genetic or biometric, linked to any natural person" (art. 5, II).

The legislation has a separate section concerning sensitive data because it demands greater security and confidentiality in use, as they reflect the opinions and convictions of the holder, as well as their physical and mental health status, and political participation.

With the Law, it is expected that all medical data will be shortly available through a computerized system and shared among some health care entities.

Given its repercussion beyond the holder, Reis and Oliveira (2019) explain the bioethical principles concerning genetic data outlined in *Human Genome Editing: science, ethics, and governance*[7], which should guide the use of information regarding the genome, especially the promotion of well-being, transparency, due care, responsible science, respect for people, equity, and transnational cooperation. Some of these bioethical principles were incorporated into the LGPD, especially the teleological constraints on data use, consent, and individual and collective responsibility.

As already noted, the processing of personal data requires the consent of the owner or their legal guardian, with specific reference to the intended use. LGPD, however, does not require consent in some cases. Therein lies the issue, given the wide-ranging hypotheses, as is the case with paragraph e, which establishes the unnecessary consent when there is an intention to "protect the life or physical safety of the holder or a third party."

So, broadly speaking, third-party protection may justify the breaking of privacy related to health data. But wouldn't that open the door to a potentially dangerous relativization of privacy? What configures this "third party protection" regarding life and physical safety? Does this precept allow for the disclosure of medical secrets?

There are no answers to these questions, but only the reiteration of the fluidity of privacy and the ambivalence between the right to know for some and the right to not want others to know, for others.

Regardless of being an individual issue, health repercussions could be generational in cases of hereditarily transmitted data.

On the other hand, the public interest justification can prove fallacious and authoritarian. To speak of the prevalence of public interests over private interests is to admit the existence of a universalizable interest, which is, at the very least, difficult in contemporary plural societies.

There are two major risks if, on the other hand, the argument of the

7 Report released in 2017, written by the *National Academies of Sciences, Engineering, and Medicine* of the United States.

majority's interest is chosen: the risk of the tyranny of the majority, which does not even respect the fundamental rights to stand out from the rest; as well as the risk of considering that the whole is not comprised by parts.

Now, if there is this public interest as a predominant value in society, it is not a State interest, but a set of private interests, which build up within the historical inconstancy of the arguments.

Daniel Sarmento arguments on the impropriety of the "supremacy of the public interest":

> Therefore, the picture that is outlined before our eyes is much more that of convergence between public and private interests than that of collision. Such a situation, again, is not the exception, but the rule. In the vast majority of cases, the community benefits from the effective protection of the interests of its members. Especially because the public interest, in fact, is composed of the private interests of members of society, which is why, as a rule, it is impossible to dissociate public interests from private ones (SARMENTO, 2007, p. 83-84).

Another relevant point of the LGPD concerns sensitive data sharing, which may be allowed under certain circumstances.

At a glance, it seems that sharing sensitive personal health data with a view to obtaining an economic advantage is not allowed. However, art. 11 § 4 makes so many exceptions to this prohibition that it seems to allow more than oppose:

> § 4 – Communication or shared use between controllers of sensitive personal data related to health is forbidden with the objective of obtaining an economic advantage, **except in cases concerning the provision of health services, pharmaceutical assistance, and health care assistance**, provided that the § 5 of this article, including auxiliary services for diagnosis and therapy, for the benefit of the data subjects' interests, and to allow:
> I – data portability upon the holder's request;
> II – the financial and administrative transactions resulting from the use and provision of the services referred to in this paragraph.

Such sharing, however, is prohibited for operators of health insurance plans and other economic activities unrelated to health services (LGPD, article 11, § 5).

CONCLUSION

Within the digital environment, data constitutes the identity itself, transfigured in *bytes* to enable storage in a single language.

Big Data prevents human beings from being able to directly mapping and crossing data, which leads humanity to the inevitable path of Artificial Intelligence.

That is why the Law must establish regulations and procedures for the storage, control, disclosure, and sharing of data, especially regarding natural persons and legal entities.

To this end, rules such as the Brazilian LGPD have been formulated, with the objective of preserving freedom of expression and communication, but with guarantees to the rights to privacy, honor, image, and free competition.

The right to privacy, as a manifestation of the individual space that one chooses to share restrictively, gains new features under the influence of the digital environment, while, at the same time it finds new possibilities of manifestation, it opens space for new violations. In the case of the world wide web, data which has been freely released does not return to the private circle, given that data control may be perceived as property, over which complete control is intended.

The subject may even gain more dramatic connotations in the face of sensitive data, which reveal generational and medical information, religious beliefs, and political opinions. Although the rule of confidentiality prevails over such data – meaning that its disclosure requires the holder's consent – there will be situations in which its disclosure may become useful to preserving a third party's health and the boundaries of the law will have to be projected or retracted under factual circumstances.

As a result, the classic configuration of the regulations concerning privacy seems to fall apart as legal rules, giving way to principled fluidity – privacy principle –, as the legislator cannot describe their application conditions as they would in a standard rule or factual support.

REFERENCES

BRASIL. *Código Civil*. Lei n. 10.406. 10 jan. 2002. Institui o Código Civil. Available at: http://www.planalto.gov.br/ccivil_03/leis/2002/L10406compilada.htm. Access on: Mar 1, 2020.

BRASIL. Constituição (1988). *Constituição da República Federativa do Brasil*. Available at: http://www.planalto.gov.br/ccivil_03/constituicao/constituicaocompilado.htm. Access on: Mar 1, 2020.

BRASIL. Lei n. 13.709, de 14 de agosto de 2018. *Lei Geral de Proteção de Dados Pessoais* (LGPD). Available at: http://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/L13709.htm. Access on: Mar 1, 2020.

DE CUPIS, A. *Os direitos da personalidade*. Campinas: Romana Jurídica, 2004.

DUHIGG, C. *O poder do hábito*: por que fazemos o que fazemos na vida e nos negócios. Rio de Janeiro: Objetiva, 2012.

FARIAS, C. C.; NETTO, F. B.; ROSENVALD, N. *Novo tratado de responsabilidade civil*. 2. ed. São Paulo: Saraiva, 2017.

FIORILLO, C. A. P. *Curso de direito ambiental brasileiro*. 13. ed. rev., atual. e ampl. São Paulo: Saraiva, 2012.

FLORIDI, L.; TADDEO, M. What is data ethics? *Philosophical Transactions of the Royal Society a Mathematical, Physical and Engineering Sciences*, v. 374, p. 1-5, dec. 2016. Available at: https://royalsocietypublishing.org/doi/pdf/10.1098/rsta.2016.0360. Access on: Jan 30, 2020.

GINSBERG, J. et al. Detecting influenza epidemics using search engine query data. *Nature*, v. 457, p. 1012-1014, feb. 2009. Available at: https://www.nature.com/articles/nature07634. Access on: Jan 30, 2020.

GOLDER, S. A; MACY, M. W. Diurnal and seasonal mood vary with work, sleep, and daylength across diverse cultures. *Science*, v. 333, p. 1878-1881, sep. 2011. Available at: https://science.sciencemag.org/content/333/6051/1878.full. Access on: Jan 31, 2020.

LANEY, D. *3D data management*: controlling data volume, velocity, and variety. Available at: https://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf. Access on: Jan 31, 2020.

MAYER-SCHONBERGER, V.; CUKIER, K. *Big data*: the essential guide to work, life and learning in the age of insight. New York: Hougnton Mifflin Harcourt, 2013.

NAVES, B. T. O.; REIS, É. V. B. *Bioética ambiental*: premissas para o diálogo entre a ética, a bioética, o biodireito e o Direito Ambiental. 2. ed., rev. e aum. Rio de Janeiro: Lumen Juris, 2019.

NAVES, B. T. O.; SÁ, M. F. F. *Direitos da personalidade*. Belo Horizonte: Arraes, 2017.

REIS, É. V. B.; OLIVEIRA, B. T. CRISPR-CAS9, biossegurança e bioética: uma análise jusfilosófica-ambiental da engenharia genética. *Veredas do Direito*, Belo Horizonte, v. 16, n. 34, p. 123-152, jan./abr. 2019.

SAMUEL, A. L. Some studies in machine learning using the game of checkers. *IBM Journal*, v. 3, p. 210-229, jul. 1959. Available at: https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=5392560. Access on: Feb 1, 2020.

SARMENTO, D. Interesses públicos vs. Interesses privados na perspectiva da teoria e da filosofia constitucional. *In:* SARMENTO, D. (Org.). *Interesses públicos versus interesses privados*: desconstruindo o princípio da supremacia do interesse público. Rio de Janeiro: Lumen Juris, 2007. p. 23-116.

TAYLOR, C. *As fontes do self*: a construção da identidade moderna. São Paulo: Loyola, 1997.

TAULLI, T. *Introdução à inteligência artificial*: uma abordagem não técnica. São Paulo: Novatec, 2020.

Article received on: 03/04/2020.
Article accepted on: 05/06/2020.

How to mention this article (ABNT):

REIS, E. V. B.; NAVES, B. T. O. The digital environment and the right to privacy in relation to big data. *Veredas do Direito*, Belo Horizonte, v. 17, n. 37, p.135-156, jan.-abr. 2020. Available at: http://www.domhelder.edu.br/revista/index.php/veredas/article/view/1795. Access on: Month day, year.