EAI Endorsed Transactions
on Scalable Information Systems
Research Article EAI.EU

Effects of data protection laws on data brokerage businesses

G. Birckan[1, 2, *], M. L. Dutra[2], D. D. J. de Macedo[2], and A. F. G. Viera[2]

[1] Federal Police, Brazil
[2] Federal University of Santa Catarina, Florianópolis, Brazil

Abstract

For years now, players known as data brokers have operated on an unregulated market, by building businesses from the aggregation and processing of data, away from public scrutiny. Due to the genesis of laws that ended up exposing their model, along with debates on data privacy, a new generation of data protection acts, much stricter, are being enacted worldwide. The shift away from the data commoditization paradigm might, at first, indicate the extinction of such businesses. In this paper, we present related concepts and a brief history that led to that current scenario, along with some collaborative-oriented proposals for the adequacy of the data brokerage industry to a transparent perspective.

Keywords: digital identity, internet privacy, personal data, data brokers.

Received on 17 June 2020, accepted on 20 July 2020, published on 23 July 2020

doi: 10.4108/eai.13-7-2018.165673

1. Introduction

This text is an extension of the paper published on the DIONE2020 Conference [1]. Big Data, Cloud Computing, Internet of Things and Data Brokers are all contemporary terms, buzzwords associated with what is being hyped in tech trends - except for the last one. Data brokers, although increasingly subject to debates, are as old as the Internet itself. After all, since information started being published online, there has always been a need for its aggregators, as there always have been those who were interested in their products: compilations from sparse sources generating specific dossiers about something or someone.

There is power in identity. According to Castells (2011), it is "people's source of meaning and experience." The most important link in the "who-what-where-when" tetrad, the unique identification of a being that is extracted from a data bulk (especially on the Web) has been, for decades, the desire of states and corporations. Prins (2006) points out that "a look at our contemporary, data-based society reveals that information about people is essential for a variety of economically and socially useful and crucial purposes: education, taxation, social benefits, health care, crime detection and terrorism prevention, commerce and marketing, to name but a few."

The goal of this paper is to recognize the new generation of laws that have been created worldwide, which establish principles, limits, and responsibilities to those who produce and consume data on individuals. Initially, concepts related to the data business as well as a brief history of such industry are introduced. Next, it is illustrated some of the events related to the leak of personal data that brought to the fore the need for regulations. Finally, we present some legislations that were recently enacted in the United States, Europe, and Brazil, and discuss their effects on the data brokerage industry.

2. Data Brokers

Data brokers are companies that collect information, including personal information about consumers, from a wide variety of sources with the intent of reselling this information for various purposes. These purposes include

*Corresponding author. Email: gbirckan@gmail.com

verifying an individual's identity, differentiating records, marketing products, and preventing financial fraud (Federal Trade Commission 2012). For the extraction (and presumption) of knowledge, statistical algorithms are used. Horvitz and Mulligan (2015) add that nowadays machine learning techniques are also used, which can facilitate making leaps across informational and social contexts, generating inferences.

There is no universal consensus on what Big Data is. However, De Mauro, Greco and Grimaldi (2016) describe it as "the information asset characterized by such a high volume, velocity, and variety to require specific technology and analytical methods for its transformation into value." The abundant volume of information that people generate every second did not take long to have its potential recognized, becoming, beyond a commodity, a whole new specialized business. As Sevignani (2013) explains, "commodification is the process of making things exchangeable on markets, either actually and/or discursively by framing things as if they were exchangeable." Roderick (2014) adds that "the growth of companies [...] has facilitated a shift in attention from production-oriented to marketing-oriented strategies, allowing companies to tap into and encourage (ir)rational purchase behavior." More than random or spontaneous data, especially nourished by the massive scale of the social networks, Big Data is also constructed from individuals' data, and that is where its real value resides.

Mosco and Wasko (1988) explain the essence of what is happening: "new technology makes it possible to measure and monitor more and more of our electronic communication and information activities. Business and government see this potential as a major instrument to increase profit and control. The result is a pay-per society." Figure 1 shows an ordinary flow of information to/from a data broker, describing commonly used sources for capturing data, as well as other public and private actors who participate in such ecosystem.

Although governments usually possess robust databases, eventually they also end up hiring data brokers' services. According to Stevens (2001), "private companies maintain and organize personal information on individuals in a manner that may not be legally available to government actors." As an example, there is the United States Privacy Act[1], which "establishes a code of fair information practices that governs the collection, maintenance, use, and dissemination of information about individuals that is maintained in systems of records by federal agencies."

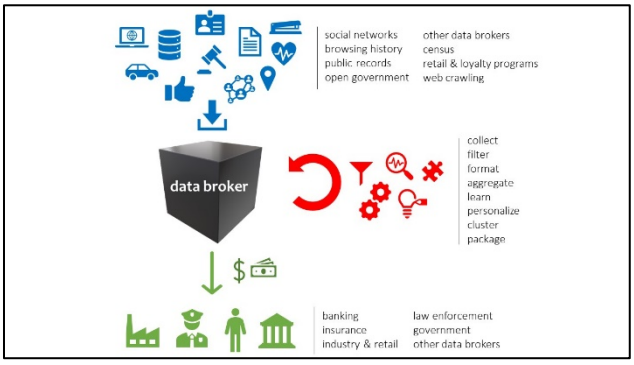

Figure 1. Ordinary flow of information from/to a data broker, from the work of Otto, Antón and Baumer (2007).

The comprehension of an individual's decision process through the employment of statistical models allows the establishment of behavioral patterns; additionally, it also grants the development of anticipation on trends (Ostrowski, 2013; Ngai, Xiu, & Chau, 2009). Examples of such a scenario include habit changes, fluctuation on demands, and interest shifts for specific goods. This predictive value complements the intrinsic importance of static data: the greater the collection of entries on someone, the more accurate the classification models will be, thus justifying the rush on digital mining. Zhang et al. (2019) add that firms have economic incentives to purchase data from data brokers in order to improve their ads delivery efficiencies, by increasing the probability of delivering and showing ads to the relevant consumers. This, in turn, would also benefit the customers themselves, by reducing their search cost. On that, Bourreau, Caillaud and De Nijs (2017) explain that offering more adequate products or more relevant ad messages, without any impact on prices, is presumably improving the consumers' welfare and delivering them better value. However, the increasing capacity of online players to extract value from consumers through more sophisticated price discrimination may reduce or even reverse the direct effect of personalization of services and of advertising.

Knowing information about people makes it possible to cluster them, which means label or assemble groups that share similar characteristics - or according to requirements and attributes pertinent to whoever is interested, from socioeconomic profiles to consuming inclinations. This capability comprises a latent ethical impasse, as it opens the possibility of the usage of variables that are not only merely demographic, but also may imply in questionable contexts. Features that nowadays are not seen as politically correct involve race, religion, gender, age, and income, among others - which in some legislations could also be a crime. Therefore, linking digital profiles to automated decision-making algorithms may (inadvertently or purposely) lead to discriminatory results, as pointed in the Big Data and Privacy report[2] made for The US President's Council of Advisors on Science and Technology in 2014.

[1]https://www.justice.gov/opcl/privacy-act-1974

[2]https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/PCAST/pcast_big_data_and_privacy_-_may_2014.pdf

The report examined the nature of current technologies for managing and analyzing Big Data, as well as preserving privacy. It considered how those technologies are evolving, and explained what the technological capabilities and trends imply for the design and enforcement of public policy intended to protect privacy in Big Data contexts.

3. Data privacy

Although a growing and virtually multibillionaire market, the dissemination of the data broker business model did not occur without any questioning. The debates take place around recurrent subjects, among which three are commonplace: transparency for when data is captured, loss of control over one's anonymity, and the sharing of the profits.

On the (lack of) transparency subject, Reyman (2013) describes the reality that "terms-of-use policies that describe data collection and use are required by law, but these are lengthy and difficult to understand when read at all". Besides, data is often obtained on social web technology trade-offs from "tacit agreements that users enter into, and a set of unspoken assumptions that govern who owns what is created and how it circulates". Gangadharan (2017) added that "marginal Internet users ignored privacy policies or terms of service agreements that they encountered", and that "when signing up for email, and despite instructors' advice to carefully review user agreements, students clicked through or past privacy policies and terms of service in order to complete the registration, suggesting these notification mechanisms functioned as meaningless accessories to the new learner's Internet experience."

The next common question refers to where do the data end up after all and who has access to it, which are key issues on the argument of anonymity control. In a non-regulated system, information can be sold and distributed without acknowledgments or even accountability of the transactions. In that prospect, Rachels (1985) explains that the value of privacy is "based on the idea that there is a close connection between our ability to control who has access to us and to information about us, and our ability to create and maintain different sorts of social relationships with different people." Roessler and Mokrosinska (2013) state that "the control and regulation of informational privacy should be viewed not only under the perspective of individual rights, but also as being necessary for social interactions themselves, and therefore as relevant to the integration of society."

Another consistent controversy concerns the earnings from third-party information: if companies in such business make extraordinary profits with data that are essentially generated by people, where are my paychecks? In that direction, Malgieri and Custers (2018) describe how "personal data of individuals represent monetary value in the data-driven economy and are often considered a counter-performance for 'free' digital services or discounts for online products and services". Furthermore, they point out that "individuals do not seem to be fully aware of the monetary value of their personal data and tend to underestimate their economic power within the data-driven economy and to passively succumb to the propertization of their digital identity."

In a nutshell, the essential aspect of the data broker industry, as emphasized by Crain (2018), is the asymmetrical loss of privacy: "people are opened up to increasingly extensive forms of monitoring, while the institutions doing the monitoring and the information they collect remain hidden from view. [...] Privacy asymmetry as a descriptive category is especially salient for the data broker industry, which has long operated without public awareness or direct regulatory oversight. The privacy of those under watch is undermined, while the watchers themselves operate with substantial freedom from scrutiny."

As technology advanced and the Internet's popularity escalated, despite the fact that the growth of online information has also increased the supply of informational inputs, data aggregators have always existed. After decades of progressive exploring, the beginning of the exposure of this practices, the so-called data breaches, triggered public objection and awareness, as we will see next.

4. The exposure of the data brokerage industry

A data security breach occurs when there is a loss, theft or other unauthorized access to sensitive personally identifiable information that could result in the potential compromise of the confidentiality or integrity of data (Stevens, 2012). This fact represents a deep problem in many areas, such as [29, 30, 31, 32, 33, 34, 35, 36]. Legislation that addresses such cases usually requires the events to be made public, and both the potentially affected individuals and regulatory agencies to be informed. The obligation to make the facts known is broad, reaching not only data brokers but also any private, non-profit or public organization, regardless of their area of activity (health, education, insurance, finance, etc.).

Stevens (2012) declares that security breach notification laws generally follow a similar framework and can be categorized into several standard elements: (1) delineating who must comply with the law; (2) defining the terms "personal information" and "breach of security"; (3) establishing the elements of harm that must occur, if any, for notice to be triggered; (4) adopting requirements for notice; (5) creating exemptions and safe harbors; (6) clarifying preemption and relationships to other federal laws; and (7) creating penalties, enforcement authorities, and remedies.

The significance of the expanding expenses in cybersecurity, with the added intent of also preventing - or minimizing - breaches, can be observed in Figure 2, which shows the USA annual spending in that area, from 2010 to 2018.

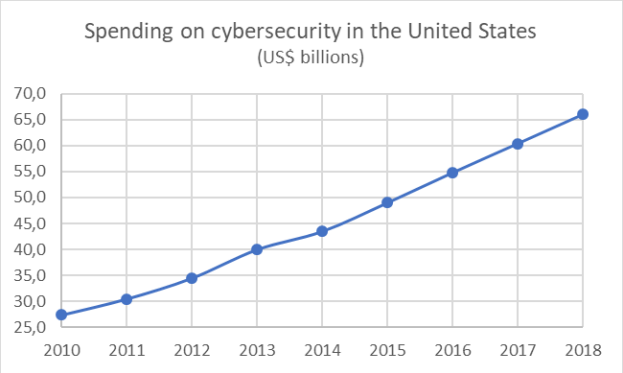

Figure 2. Annual spending on cybersecurity in the USA (data source: statista.com).

Figure 3 shows the annual records of data breaches in American organizations for the same period. In spite of steady expenses in cybersecurity, employee training and expanding regulations concerning individual data maintenance, incidents have been following a stable pattern. If, on the one hand, the statistics display an apparent regularity on the annual number of events shown in Figure 3, on the other hand, the volume of compromised records has been following a rising trend, as can be seen in Figure 4. Considering such a scenario, one possible explanation could be the increase in database sizes, proportional to the popularizing of social networks and the employment of Big Data technologies for capturing information.

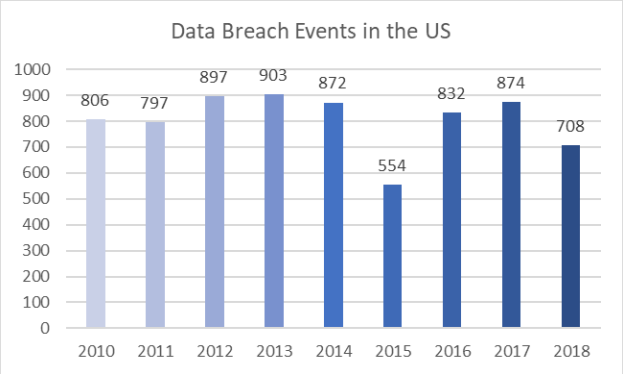

Figure 3. Data breach events in American organizations from 2010 to 2018 (data source: privacyrights.org).

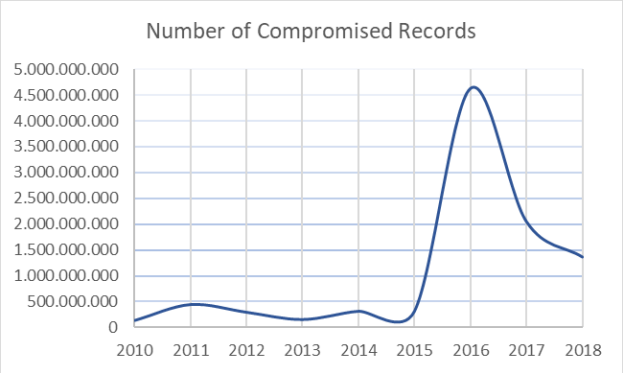

Figure 4. Progress of the number of compromised records, annually, in data breach events (data source: privacyrights.org).

Data breaches can be categorized according to their causes or origins, which are not limited to cyber-attacks from hackers and malware (although most of the incidents are based on those, see Figure 5). They are also considered the cases of unintended disclosure (sensitive information posted publicly, mishandled or sent to the wrong party), physical loss (paper documents or portable devices that are lost, discarded or stolen), insider (someone with legitimate access intentionally breaches information), fraudulent transactions involving debit and credit cards, and finally, the unknown cases. Unauthorized access to data causes direct losses due to financial fraud, identity theft, and industrial espionage, besides indirect losses such as reputation and asset depreciation.

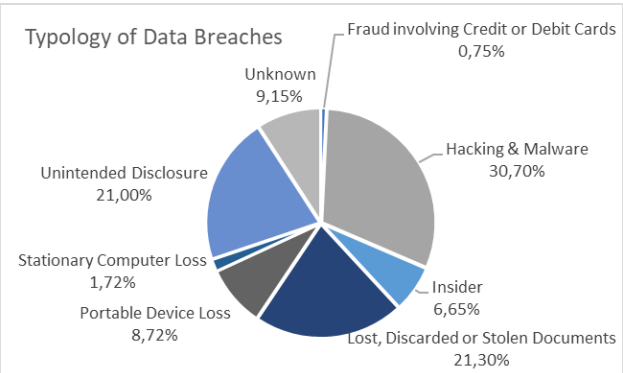

Figure 5. Typology of data breach events between 2010 e 2018 (data source: privacyrights.org).

According to Gordon, Rebovich and Choo (2007), identity fraud may be defined as the misuse of personal or financial identifiers for personal gain or to facilitate other criminal activity. Such gains may be obtained from online shopping scams, usage of stolen cards, or controlling and spending over someone else's accounts. Information leaks are often the fuel and/or the beginning of cybercrimes, and the increasing statistics of such frauds, added to the growing cases of data breaches, ultimately brought the attention to a market that up to this point had operated quietly.

Those who suffer from crimes of identity theft or misuse of personal data are customarily left, often with little or no assistance, with the necessary bureaucracy to reestablish their names and losses. In addition to the protest of the victims, there has been some support from non-profit research organizations that turned their attention to personal data protection. Among these organizations, we can highlight the EPIC - Electronic Privacy Information Center[3] and the Privacy International[4], which are both plaintiffs of several civil law-suits for alleged privacy abuse by technology companies.

It can be said that there has been a first era of exploring and exploitation of personal data on the Web, due to the lack of regulation and auditing. Recent movements, with the enactment of laws designed specifically with a focus on information privacy, point towards a paradigm shift, in which users are empowered while managing their data. In the next section, examples of the legal innovations that sustain this transition will be presented.

5. Personal data usage regulations

It has been observed that technology usually evolves faster than the legislative processes. Therefore, the classical narrative applied to the approach of a new problem is to try to address it with existing (older) laws, up until the creation and establishment of local specific regulations. In Brazil, recent examples of this reality are the laws 12.965/2014[5] (known as "Marco Civil da Internet", i.e. *Internet Civil Legal Framework*) and 13.640/2018[6]. The latter regulates the business model of transportation operated by private drivers through portable apps like Uber or Cabify. Before the "Marco Civil da Internet", duties and rights of users and internet providers were disciplined by the Brazilian Civil Code, Penal Code, Consumer Defense Code and the Constitution (all of them older than the Internet itself).

The reason behind the fact the privacy of personal data started to get more attention is related to scandals involving information security leaks, which were forced to be disclosed by laws that came at the beginning of the third millennium. One early example was the enactment of a Californian bill in 2002[7], which requires "a state agency, or a person or business that conducts business in California, that owns or licenses computerized data that includes personal information, as defined, to disclose in specified ways, any breach of the security of the data, as defined, to any resident of California whose unencrypted personal information was, or is reasonably believed to have been, acquired by an unauthorized person."

In light of the spreading number of exposure reports (shown in Figure 4) and crimes of identity theft, the data acquisition model consequently began to be challenged. The next generation of laws, after a maturing cycle that took ten to fifteen years, is more rigorous when it comes to managing personal data. The main example was the enactment of the European General Data Protection Regulation (GDPR)[8], in 2018, which affects companies that conduct businesses in the European Union, regardless of where their headquarters are. The major changes introduced by it are synthesized in Figure 6.

Clear language	
Before	After
Often businesses explain their privacy policies in lengthy and complicated terms	Privacy policies will have to be written in a clear, straightforward language
Consent from user	
Before	After
Businesses sometimes assume that the user's silence means consent to data processing, or they hide a request for consent in long, legalistic, terms and conditions - that nobody reads	The user will need to give affirmative consent before his/her data can be used by a business. Silence is no consent
More transparency	
Before	After
The user might not be informed when his/her data is transferred outside the EU	Businesses will need to clearly inform the user about such transfers
Sometimes businesses collect and process personal data for different purposes than the reason initially announced without informing the user about it	Businesses will be able to collect and process data only for a well-defined purpose. They will have to inform the user about new purposes for processing his/her data
Businesses use algorithms to make decisions about the user based on his/her personal data (e.g. when applying for a loan); the user is often unaware of this	Businesses will have to inform the user whether the decision is automated and give him/her a possibility to contest it
Stronger rights	
Before	After
Often businesses do not inform users when a data breach occurs, e.g. when the data is stolen	Businesses will have to inform users without delay in case of a harmful data breach

[3] http://epic.org
[4] http://privacyinternational.org
[5] http://www.planalto.gov.br/ccivil_03/_ato2011-2014/2014/lei/l12965.htm
[6] http://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/Lei/L13640.htm
[7] http://www.leginfo.ca.gov/pub/01-02/bill/sen/sb_1351-1400/sb_1386_bill_20020926_chaptered.pdf
[8] https://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1528874672298&uri=CELEX%3A32016R0679

Often the user cannot take his/her data from a business and move it to another competing service	The user will be able to move his/her data to another social media platform
It can be difficult for the user to get a copy of the data businesses keep about him/her	The user will have the right to access and get a copy of his/her data a business has on him/her
It may be difficult for a user to have his/her data deleted	Users will have a clearly defined "right to be forgotten" (right to erasure), with clear safeguards
Stronger enforcement	
Before	After
Data protection authorities have limited means and powers to cooperate	The European Data Protection Board grouping all 28 data protection authorities will have the powers to provide guidance and interpretation and adopt binding decisions, in case several EU countries are concerned by the same case
Authorities have no or limited fines at their disposal in case a business violates the rules	The 28 data protection authorities will have harmonized powers and will be able to impose fines to businesses up to 20 million EUR or 4% of a company's worldwide turnover

Figure 6. Key points established by the European General Data Protection Regulation (Source: https://ec.europa.eu/commission/sites/beta-political/files/data-protection-factsheet-changes_en.pdf)

In the past few years, United States institutions, including the Federal Trade Commission, Government Accountability Office, and US Senate, have released reports and held hearings on the practices and operations of data brokers. Despite having dispersed state acts, there has not been adopted a federal or country-wide law, hence, data brokers remain mostly unregulated (Yeh, 2018).

The United Kingdom, in turn, enacted their implementation of the GDPR, named "Data Protection Act 2018"[9] , which controls how personal information is used by organizations, businesses, or the government. Such act dictates that everyone responsible for using personal data has to follow strict rules called "data protection principles", making sure the information is (a) used fairly, lawfully and transparently; (b) used for specified, explicit purposes; (c) used in a way that is adequate, relevant and limited to only what is necessary; (d) accurate and, where necessary, kept up to date; (e) kept for no longer than is necessary; and (f) handled in a way that ensures appropriate security, including protection against unlawful or unauthorized processing, access, loss, destruction or damage. It is important to add that there is stronger legal protection for more sensitive information, such as race, ethnic background, political opinions, religious beliefs, trade union membership, genetics, biometrics (where used for identification), health, and sex life or orientation.

Brazil also followed the international movement and introduced an update to the 2014 "Marco Civil da Internet" by the enactment of law 13.709/2018[10], which addresses personal data protection specifically. This bill institutes ten fundamental principles, described in Figure 7, which are consonant with the content established in their European counterparts.

Principle	Description
purpose	data must be handled with legitimate, specific and explicit purposes, the user has to be informed, and data should not be processed later on for different reasons than it was initially acquired for
suitability	data must be handled in a way that is compatible with the goals informed to the user
necessity	data handling must be restricted to the minimum necessary to fulfill the purpose it was acquired for
free access	user's access to information about how and how long their data will be handled must be provided in a free and facilitated way
data quality	users must be assured about the correctness, clearness, relevance, and currentness of their data
transparency	users must be given clear, precise and easily accessible information about their data usage and handlers
security	establishment of administrative and technical rules to protect personal data from unauthorized access and from incidental situations such as destruction, loss, alteration, or disclosure
prevention	following of means to prevent damage due to data mishandling
non-discrimination	data cannot be used for abusive, illicit or discriminatory means
accountability	handlers must prove effective actions to obey and enforce such principles

Figure 7. Principles that must be obeyed on personal data handling (source: Brazilian law 13.709/2018).

[9] https://www.gov.uk/data-protection

[10] http://www.planalto.gov.br/ccivil_03/_Ato2015-2018/2018/Lei/L13709.htm

As a requirement for data acquisition and processing, the 13.709/2018 law establishes the obligation of the user's consent, in addition to his/her right to access, manage, correct, and eliminate his/her data. Among the penalties for transgressions, there is a fine that can reach up to R$ 50 million[11]. The milestone from the Californian bill that brought up attention to the data exploration business has not also been forgotten here: any security incident or breach that might lead to risk or damage must be disclosed to the corresponding individuals and the authorities.

Complementing the law 13.709/2018, in July 2019 it was enacted the law 13.853/2019, which created the Brazilian National Data Protection Authority. This agency has technical and decisory autonomy, is bound to the President's Office, and has the responsibility of watching over personal data protection, overseeing that the rules are followed properly, and applying penalties whenever is the case.

Yet, regardless of the Brazilian efforts, this country was not the first in Latin America to implement such set of actions. Uruguay has a personal data protection law since 2008 (Ley 18.331[12]), while Argentina has such legal framework since 2000 (Ley 25.326[13]). The importance of those mechanisms was recognized by the European Commission, which regarded both countries "as providing an adequate level of protection for personal data as referred to in Directive 95/46/EC," as per 2003/490/EC[14] and 2012/484/EU[15] decisions. Paraguay, in turn, has had its bill[16] ("Proyecto de Ley de Protección de Datos Personales") approved by the congress and sent to the executive for enactment. Over the next section, some ideas about adapting the data brokerage business to this new (strict) legal scenario are presented.

6. Proposals and perspectives

For decades, players known as data brokers operated in a market with little or no regulation, where transactions between corporations and governments were conducted without restrictions and away from public scrutiny. Digitally speaking, in that context pretty much everything was possible: capturing, buying, selling, and sharing of information. That information could also be mined and statistically inferred, such as the clustering of profiles or the prediction of trends.

In a first moment, laws were created to make those events known as data breaches public, which eventually brought attention to the market of personal data. In a global scale, the successive annual reports containing a growing number of compromised records (as illustrated on Figure 2 to 4) despite the spending on cybersecurity, in addition to the growing world claim for privacy and the tension from organizations that act on behalf of data protection potentially contributed to the recent advent of a new generation of harsher regulations.

Such fresh laws comprise restraints that directly affect enterprises in the data brokerage business, and suggest a possible change in practice regarding the time when personal information was processed as a commodity. Nonetheless, adaptations to this new paradigm more focused on privacy and data protection are feasible, respecting the users and their individual authority on controlling who they are (their digital identities) and what they produce (their generated data).

Values and principles that guided the creation of data protection laws comprise consent from the user (most important), transparency, and purpose, premises that must be complied with and also considered when designing new data-driven business models. By changing the practice of using third party data as a sheer input and bringing the original "data sources" (the users) closer as partners and suppliers, it is possible to envisage a healthier continuity scenario for several segments of data-driven activity in a foreseeable future. With a certain level of anonymity or voluntary exposure, products such as behavior prediction and consumer profiling or a wide range of classifiers might still be appealing and functional.

From a collaborative perspective, new proposals arose, such as the policy framework for user data sharing by Iyilade and Vassilev (2013), based on the idea of a market. In that concept, applications can "offer and negotiate user data sharing with other applications according to an explicit user-editable and negotiable privacy policy that defines the purpose, type of data, retention period and price."

Malgieri and Custers (2018) investigated different models for quantifying the value of personal data, analyzing whether consumers/users should have a right to know the value of their data. The authors also discussed active models of choice, in which users are offered the option to pay for online services, either with their personal data or with money. The conclusion, however, was that these models are incompatible with current data protection laws.

Tona et al. (2018) presented "a conceptual design for an artifact that will raise awareness amongst individuals about Big Data ethical issues and help to restore the power balance between individuals and organizations." Their proposal was constructed upon five dimensions derived from the European GDPR, such as consent, the right to be forgotten, the right to access, data portability, and data circulation. All those pillars are arranged over a foundation that would allow several collaborative interactions like replying, commenting, reviewing, rating, and tagging data. By observing the ubiquity of mobile smartphone usage and the ensuing massive generation of data from those devices (locations, movements, images, video, text, and even

[11] US$ 9.765.625,00 – exchange rate of 22/07/2020 where US$ 1 = R$ 5,12.
[12] https://www.impo.com.uy/bases/leyes/18331-2008
[13] http://servicios.infoleg.gob.ar/infolegInternet/anexos/60000-64999/64790/norma.htm
[14] https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32003D0490
[15] https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32012D0484
[16] http://silpy.congreso.gov.py/expediente/115707

health data, which is ordinarily uploaded to content-service providers), Mun et al. (2010) presented a privacy architecture named Personal Data Vaults. In their proposal, individuals retain ownership of their data, which can be reviewed and filtered before being shared, exploring three mechanisms for managing data policies: granular access control lists, trace-audit, and a rule recommender, which provides a high-level interface for setting sharing policies.

Oh et al. (2019) proposed an optimized trading model for data brokers who buy personal data with proper incentives based on the willingness-to-sell of providers, and they sell valuable information from the refined dataset by considering the willingness-to-buy of consumers, and the dataset quality. Their paper designs a personal data trading model in an Internet of Things data marketplace with multiple types of personal data (e.g., physical location, brand preferences, purchase histories, etc.) by considering economic benefits of personal data providers as well as satisfaction of personal data consumers. The proposal also includes a personal data quality model for the collection of heterogeneous types of personal data, and models a profit maximization problem with expected revenue and cost. The authors conclude that the proposed model is feasible even if the data brokers spend costs to gather personal data.

We can conclude that, in spite of the strictness of the new privacy laws, albeit fresh and assigning technology companies to an adjustment cycle, there are several studies and initiatives for the creation of tools (frameworks, models and architectures) that provide the users with more power to control their personal information. Such possibilities enforce a pattern shift away from the commoditization of data by the brokers, instead of the sheer extinction of their old model. Opportunities, therefore, might be in the effective deployment of collaborative data-driven platforms and products, which should have their primary focus on the value that has been so emphasized: the transparency.

References

[1] Birckan, G., Dutra, M. L., Macedo, D. D. J., & Viera, A. F. G (2020). Personal data protection and its reflexes on the data broker industry. In *EAI International Conference on Data and Information in Online Environments*. Florianopolis, Brazil.

[2] Castells, M. (2011). *The power of identity* (Vol. 14). John Wiley & Sons.

[3] Prins, C. (2006). Property and privacy: European perspectives and the commodification of our identity. *Information Law Series*, *16*, 223-257.

[4] Federal Trade Commission. (2012). Protecting consumer privacy in an era of rapid change: Recommendations for businesses and policymakers. *Washington, DC: Federal Trade Commission*.

[5] Horvitz, E., & Mulligan, D. (2015). Data, privacy, and the greater good. *Science*, *349*(6245), 253-255.

[6] De Mauro, A., Greco, M., & Grimaldi, M. (2016). A formal definition of Big Data based on its essential features. *Library Review*, *65*(3), 122-135.

[7] Sevignani, S. (2013). The commodification of privacy on the Internet. *Science and Public Policy*, *40*(6), 733-739.

[8] Roderick, L. (2014). Discipline and power in the digital age: The case of the US consumer data broker industry. *Critical Sociology*, *40*(5), 729-746.

[9] Mosco, V., & Wasko, J. (Eds.). (1988). *The political economy of information*. Univ of Wisconsin Press.

[10] Stevens, G. M. (2001). Data brokers: Background and industry overview. *Wall Street Journal*, *6*(5), 552a.

[11] Otto, P. N., Antón, A. I., & Baumer, D. L. (2007). The choicepoint dilemma: How data brokers should handle the privacy of personal information. *IEEE Security & Privacy*, 5(5), 15-23.

[12] Ostrowski, D. A. (2013, July). Identification of trends in consumer behavior through social media. In *17th World Multi-conference on Systemics, Cybernetics and Informatics: WMSCI2013, Orlando, Florida* (pp. 9-12).

[13] Zhang, X., Zhang, R., Yue, W. T., & Yu, Y. (2019). What is Your Data Strategy? The Strategic Interactions in Data-Driven Advertising.

[14] Bourreau, M., Caillaud, B., & De Nijs, R. (2017). The value of consumer data in online advertising. *Review of Network Economics, 16*(3), 269-289.

[15] Ngai, E. W., Xiu, L., & Chau, D. C. (2009). Application of data mining techniques in customer relationship management: A literature review and classification. *Expert systems with applications, 36*(2), 2592-2602.

[16] Gordon, G. R., Rebovich, D. D. J., & Choo, K. S. (2007). Identity Fraud Trends and Patterns. *Center for Identity Management and Information Protection, Utica College*.

[17] Reyman, J. (2013). User data on the social web: Authorship, agency, and appropriation. *College English*, *75*(5), 513-533.

[18] Gangadharan, S. P. (2017). The downside of digital inclusion: Expectations and experiences of privacy and surveillance among marginal Internet users. *New Media & Society*, *19*(4), 597-615.

[19] Rachels, J. (1985). Why privacy is important. *Ethical issues in the use of computers*, 194-200.

[20] Roessler, B., & Mokrosinska, D. (2013). Privacy and social interaction. *Philosophy & Social Criticism*, *39*(8), 771-791.

[21] Malgieri, G., & Custers, B. (2018). Pricing privacy–the right to know the value of your personal data. *Computer Law & Security Review*, *34*(2), 289-303.

[22] Crain, M. (2018). The limits of transparency: Data brokers and commodification. *New Media & Society*, *20*(1), 88-104.

[23] Stevens, G. (2012). Data Security Breach Notification Laws. Congressional Research Service.

[24] Yeh, C. L. (2018). Pursuing consumer empowerment in the age of big data: A comprehensive regulatory framework for data brokers. *Telecommunications Policy, 42*(4), 282-292.

[25] Iyilade, J., & Vassileva, J. (2013, June). A framework for privacy-aware user data trading. In *International Conference on User Modeling, Adaptation, and Personalization* (pp. 310-317). Springer, Berlin, Heidelberg.

[26] Tona, O., Someh, I. A., Mohajeri, K., Shanks, G., Davern, M., Carlsson, S., & Kajtazi, M. (2018, January). Towards ethical big data artifacts: a conceptual design. In *Proceedings of the 51st Hawaii International Conference on System Sciences*.

[27] Mun, M., Hao, S., Mishra, N., Shilton, K., Burke, J., Estrin, D., ... & Govindan, R. (2010, November). Personal data vaults: a locus of control for personal data streams. In *Proceedings of the 6th International Conference* (p. 17). ACM.

[28] Oh, H., Park, S., Lee, G. M., Heo, H., & Choi, J. K. (2019). Personal data trading scheme for data brokers in IoT data marketplaces. *IEEE Access, 7*, 40120-40132.
[29] Marcelino, L. V., Pinto, A. L., & Marques, C. A. (2020). Scientific specialties in Green Chemistry. Iberoamerican Journal of Science Measurement and Communication, 1(1).
[30] D. D. J. d. Macedo, A. V. Wangenheim and M. A. R. Dantas, "A Data Storage Approach for Large-Scale Distributed Medical Systems," 2015 Ninth International Conference on Complex, Intelligent, and Software Intensive Systems, Blumenau, 2015, pp. 486-490, doi: 10.1109/CISIS.2015.88.
[31] D. D. J. de Macedo, M. A. M. Capretz, T. C. Prado, A. von Wangenheim and M. A. R. Dantas, "An Improvement of a Different Approach for Medical Image Storage," 2011 IEEE 20th International Workshops on Enabling Technologies: Infrastructure for Collaborative Enterprises, Paris, 2011, pp. 140-142, doi: 10.1109/WETICE.2011.26.
[32] Prado, T.C., de Macedo, D.D.J., Dantas, M.A.R. et al. Optimization of PACS Data Persistency Using Indexed Hierarchical Data. J Digit Imaging 27, 297–308 (2014). https://doi.org/10.1007/s10278-013-9665-9
[33] A. de Souza Inácio, R. Andrade, A. von Wangenheim and D. D. J. de Macedo, "Designing an information retrieval system for the STT/SC," 2014 IEEE 16th International Conference on e-Health Networking, Applications and Services (Healthcom), Natal, 2014, pp. 500-505, doi: 10.1109/HealthCom.2014.7001893.
[34] A. Puel, A. v. Wangenheim, M. I. Meurer and D. D. J. d. Macedo, "BUCOMAX: Collaborative Multimedia Platform for Real Time Manipulation and Visualization of Bucomaxillofacial Diagnostic Images," 2014 IEEE 27th International Symposium on Computer-Based Medical Systems, New York, NY, 2014, pp. 392-395, doi: 10.1109/CBMS.2014.12.
[35] D. D. J. de Macedo, H. W. G. Perantunes, L. F. J. Maia, E. Comunello, A. von Wangenheim and M. A. R. Dantas, "An interoperability approach based on asynchronous replication among distributed internet databases," 2008 IEEE Symposium on Computers and Communications, Marrakech, 2008, pp. 658-663, doi: 10.1109/ISCC.2008.4625712.
[36] Eliza H.A. Gomes, Mario A.R. Dantas, Douglas D.J. De Macedo, Carlos R. De Rolt, Julio Dias, and Luca Foschini. International Journal of Grid and Utility Computing 2018 9:4, 322-332.