COLEGIO UNIVERSITARIO DE ESTUDIOS FINANCIEROS

GRADO EN ADMINISTRACIÓN DE EMPRESAS BILINGÜE

Trabajo Fin de GRADO

ETHICS OF BIG DATA: A CORRECT USE OF PRIVACY, TRANSPARENCY AND CONSENT OF USERS PERSONAL DATA.

Autor: Enrique Hidalgo Power

Tutor: César González Cantón

INDEX

1. INTRODUCTION

Over the last years, the use of digital technologies has grown exponentially, although this has limitless possibilities for the improvement of how companies operate and can have significant impact in the efficiency of companies, the lack of control can have atrocious effects on people's life. These people that are in danger of being subjects to the exploitation of their data include every individual that uses the Internet, thus the importance of severe monitoring of providers of online services.

The lack of control has been a result of the rapid growth of this market and the complexity of the market itself. The biggest danger customers face is violation of their privacy, due to the easy access to their information caused by the lack of regulated market and the power and value of their information.

Over the last years, there have been countless cases of misuse of customer's information and inappropriate behaviors by companies. Although the way the market should be controlled is becoming more defined thanks to collaboration of countries and organism, that have created legislations such as the GDPR[1], the changes that are being imposed do not follow the growth of the market itself. The laws that apply to data gathering companies still have a long way to go until they can guarantee the safety of customers, and as of today, many companies that have been subject of misconduct have gone unpunished.

This is why I have decided to focus on the importance of a correct behavior that companies should adopt for the protection of their customers, not only to avoid financial fines from the organism that are working on making the Internet a safer place, but because of the ethical implications, and customer's satisfaction should be a top priority for anyone leading a business.

The objective of this paper is to understand how different aspects of the big data world that directly affect customers are regulated, and if they adequately comply with

[1] General Data Protection Regulation, created by the European Parliament and Council of the European Union

business ethics. These different aspects will be privacy, transparency and consent. The reason why I have chosen this topic, is that this new field is growing at a huge pace and the impact it can have in society can be very positive, however I do not believe that current regulations manage to successfully protect customers interest.

The methodology I will use for the analysis will start by looking at how the GDPR defines each term. I will use real life examples for each section to demonstrate how companies treat the different aspects nowadays. Lastly, I will show the negative aspects that arise in the case that companies fail to meet the GDPR norms and the basic business ethics principles.

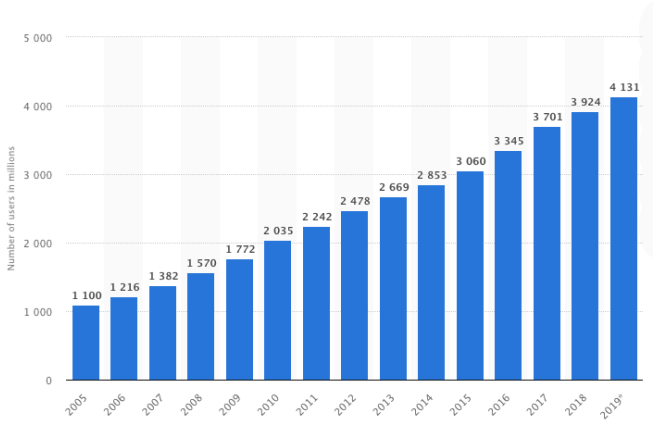
2. BIG DATA PANORAMA

2.1 ORIGINS OF BIG DATA

In order to be able to comment on the ethics revolving Big Data, we must have a deep understanding of what this relatively new concept is.

Although the concept of big data is relatively new, or at least it's gaining more recognition over the last years, the idea of analysis and gathering of extremely large amounts of data is nothing new. The first data centers date back to the 60s and 70s, however it was not until the 2000s with the boom of digital platforms that massive amounts of data were easily gathered for the further analysis of this information. It was when platforms like online platforms like YouTube, Facebook… realized how much data they were generating that Big Data started to grow, and the importance of data of costumers started gaining more recognition. (Oracle, 2020)

Overtime, as the use of digital devices has become more common, and necessary for our daily lifestyle, the amount of data generated by the average person has grown exponentially. Every year the collection of data becomes easier, thanks to the new technologies available to consumers to facilitate their daily life, products that are interconnected and accompany us throughout our daily chores collecting data in the process surround us. Some of these products include our smart phones, credit card, cars… Through the development of this interconnected products specifically designed to gather as much information as possible we generate amounts of information that far exceed their predecessors. Every two years we duplicate the amount of information generated per minute, hence the importance of Big Data. The graph to the right shows the growth in number of Internet user from 2005 until 2019. The first thing that we can see in the graph is that there is a constant growth, and it's estimated that actually (January 2020)



Graph 1 Number of Internet users worldwide

Source: Statista.com

there are over 4.54 billion users (>50% of world population) (Clement, 2020). The increasing number of Internet user indicates that every year more people are going to be generating data therefore increasing the amount of data generated, thus reinforcing the importance of Big Data.

2.2 WHAT IS BIG DATA?

According to Gartner Dictionary, Big Data is described as *high-volume, high-velocity and/or high-variety information assets that demand cost-effective, innovative forms of information processing that enable enhanced insight, decision-making and process automation* (Gartner, 2020). The key words here are volume, velocity and variety, called by some as the 3 Vs of Big Data. They are important because Big Data consists on huge amounts of information that are so complex and large that requires non-traditional computer applications in order for the information to be analyzed and transformed into useful information that could be used later used to create value for firms. As I aforementioned in the previous section, every year we generate larger amounts of data, hence why the Garner dictionary describes it as high volume. The image below shows a graphic representation of how much data was generated per day in 2019 from different devices and platforms.

Image 1 How much data is generated per day.

Source: weforum.org

Unit	Abbreviation	Decimal Value	Decimal Size
Byte	B	8 bits	1 byte
Kilobyte	KB	1,000 bytes	1,000 bytes
Megabyte	MB	$1,000^2$ bytes	1,000,000 bytes
Gigabyte	GB	$1,000^3$ bytes	1,000,000,000 bytes
Terabyte	TB	$1,000^4$ bytes	1,000,000,000,000 bytes
Petabyte	PB	$1,000^5$ bytes	1,000,000,000,000,000 bytes
Exabyte	EB	$1,000^6$ bytes	1,000,000,000,000,000,000 bytes

Table 1 Data measurements units

Source: Own elaboration with data from techterms.com

The table helps to understand the magnitude of information being generated per day, for example, just from Facebook alone, 4 Petabytes of data are generated per day. As the image above shows, the Data is obtained from many different devices, from smart watches to apps inside our phone. This explains the following two Vs of the 3 Vs of Big

Data. Since we own more devices whose objective is to generate data, the data collected in these processes come in many forms, in both formats, unstructured and structured, thus the high-variety and high-velocity, since we generate more information and in unusual forms. The objective of Big Data is to transform these immense amounts of data into useful information for companies to apply them to their business decisions. A useful transformation of structured and unstructured data can help companies in many forms, it can help reduce cost because it identifies for efficient ways of doing business; faster and more efficient business decisions, the reason for this happening is that companies obtain information about the market and consumers at a faster pace, thus speeding the decision process and not missing out on business opportunities. A better understanding of the market will also help develop new products and services since a niche will be easier to identify, or a need/want in the market will recognized faster due to all the available information about what customers look for in products and services and their satisfaction being taken into consideration.

2.3 NEGATIVE ASPECTS AND RISKS.

Even though it may seem as if Big Data can only help companies because it help place products in the market, thanks to a better understanding of patterns of consumption from the industry, consumers… it is also worth mentioning that there are some disadvantages to the use of this technology. As I mentioned before, the information being transformed doesn't always come in structured format, it may also be semi-structured and unstructured. The problem of the conversion of the data being analyzed is that it's estimated that only 20% of the information being analyzed comes in structured form, this implies that the vast majority of information comes from other sources, those being videos, audios, reports, documents… and the conversion of this data into valuable information for companies be inefficient or with worse quality. In 2015, a study that was carried out in Britain that showed that approximately 60% of consumers had purposely handed out wrong information regarding various aspects of their life's such as job title, age, address, company they work for… in order to protect their privacy (Steyerl, 2017), this is known as *dirty data*. Other notable risks associated with big data include cyber security. Since the impact that big data can have on

companies is so large, it's not rare for competing companies to try to obtain this information in illicit ways to benefit themselves, or for third parties to try to steal the information to sell it to competitors. This is a very important issue because when handling personal information of consumers, the privacy of this type of information should be guaranteed. The last risk worth mentioning is regarding the volatility of information. Since the information is changing at a very rapid pace, it must be analyzed fast before it's outdated. A fast analysis might lead to wrong conclusions that could later on guide companies in the wrong direction.

2.4 TYPES OF INFORMATION

Big Data collects information from many different sources. The most noticeable ones are the ones that are shown if the image below.

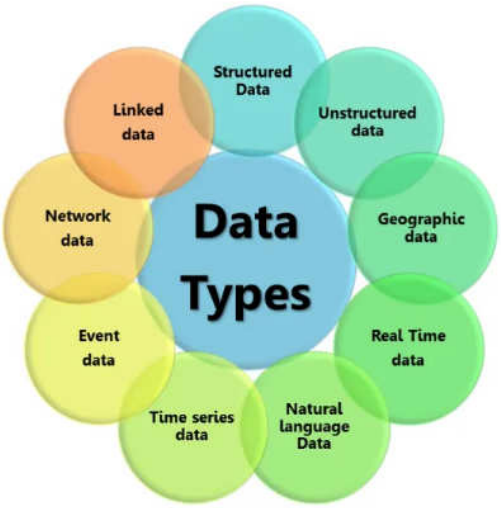


Image 2 Types of data

Source: Agroknow, Big data: Types of data Used in Analytics

Having a deeper understanding of the types of data there are will allow us to know exactly what kind of information they have collect from us. Therefore I will be giving an explanation on the content of the 4 most common types of data, those being Structured, Unstructured, Geographic and Real-Time Data. Knowing what data is being collected and processed is necessary to understand the importance of companies making a good use of it

Structured Data:

According to Big Data experts, this type of data amounts to briefly 20% of all the data being analyzed. We refer to structured data as the information that comes in defined format. This type of information is the most likely to be affected by Dirty Data, since it involves information such as Name, Age, Gender, Address, Phone Number, Currency, Dates… Structured Data is characterized by being easier to analyze and to access than other forms of data, since its stored in data warehouses and relational databases (a set

of formally described tables from which **data** can be accessed or reassembled in many different ways without having to reorganize the **database** tables)(Rouse, 2020).

Unstructured Data:

This second type of data is estimated to account for about 90% of the data generated in the 21st century. Unstructured data is made up of those pieces of information that are not structured via a pre-defined data model. The information being analyzed here includes text files, emails, social media (this is one of the reasons for its growth in volume over the last years), websites… Thus it can come in many formats, i.e. text, video, image, sound… Due to the variability of the unstructured data, it makes it riskier than structured, since there can be errors in the transformation of unstructured data into valuable information, however due to its large volume it plays a huge role in the Big Data world. As Mohammed Zouhair Al Taie says in his article, the difference, and advantage, between unstructured data and structured data resides in fact that data collected from social media has a personal taste embedded to it. (Zouhair Al Taie, 2016)

Geographical Data:

Geographical data, also known as location data, refers to the data collected from individuals regarding their movement. Electronic devices that track the position of the individual gather this information. The information collected includes the latitude, longitude or altitude; the direction the individual is going; the time at which the individual moved. The main purpose of this information is to help in the development of transportation routes and urban planning (Information Commissioner Office, year).

Real-time media.

This type of data is collected from streaming of either live of stored media data. The volume of this type of data has grown significantly over the last few years thanks to the increase in content of streaming platforms such as YouTube, Twitch, Vimeo… These platforms contain video, pictures and audio that are used to collect information regarding viewer's preferences and likes. Other types of data being collected in this

section includes online conferences, i.e. Online meeting through Zoom, Skype or other two-way video and audio transmission applications.

3. A CORRECT USE OF BIG DATA

This section will focus on the analysis of the basic elements of the ethical framework to analyze the ethicality of corporate practices. The elements are privacy, transparency and consent. The analysis will start by explaining the General Data Protection Regulation (GDPR) guidelines of each area, followed by an analysis of how the reality of each one is, and in the case of problems in any area, possible sources of improvement. I have chosen the GDPR because it is the European Regulation for the correct use of people's personal data and data circulation; they are in charge on setting the guidelines of privacy and data usage. It is important to mention that other organisms have tried to set rules regarding how big data should be used, but the GDPR has been a regulatory landmark in this new field, serving as an example to other countries and regulatory organisms. The analysis will vary for each area since, even though they are supposed to work together to protect customer rights, they are very different from one another.

3.1 Privacy:

3.1.1. What is privacy?

As I mentioned before, we live in a digital world where everything we do leaves a digital footprint behind. This footprint is the data that is being generated every second by customers. This data can be of great value for companies allowing them analyze it and turn it into information that allows them to forecast future trends in consumer behavior. However, this information that is being generated every second can oftentimes be tremendously personal and the individual generating the data would not be in favor of having this information circulating around the web.

Therefore, data can be considered a two edged sword, with the correct use it can drastically impact businesses by allowing them to have more successful marketing campaigns or by allowing them to have a better understanding of the market, however, this data is susceptible to being stolen or leaked. Here is where the privacy concept arrives.

The GDPR Privacy policy obliges companies to disclose information regarding how the data is being used and who has access to it. Privacy rights not only give consumers the right to know how their information is being used and with whom this information is

shared (*Right to access GDPR Art. 15)*, but also entitles them to other rights such as the *Right to erasure, Right to Restrict processing, Right to object*… These rights ensure that the consumers can decide when he/she wants to delete their data from the controllers data base, change the way the controller analyzed the data of the customer and even stop the controller from processing the data that is being collected from the customer.

The privacy paradox is a theory that states that what consumers claim to value their privacy is not consistent with how they act on the online marketplace. The theory suggests that customers publicly acknowledge the importance of their privacy, but in reality they willingly perform activities that put their privacy at risk, creating a discrepancy between attitudes and behavior. However, a study carried out by Kirsten Martin contradicts this theory, and underlines the importance of keeping personal data private, since customers value more the privacy of their data than the expected benefits from trading this information (Martin, 2020). Another study emphasizes the importance of trust between consumers and online businesses. The study revealed a relationship between loyalty and trust, and how a lack of trust will draw consumers away from the platform and affect their future purchasing behavior. A violation of the customer's privacy will affect the trustworthiness of the company, and affect their profitability. (Flavián, C., & Guinalíu, M. 2006)

3.1.2 How to achieve privacy?

Once we already know what the big data environment takes into consideration when defining privacy, we can elaborate on how to achieve it. Privacy's biggest issue is the leak of personal information. There are many ways companies can reduce the risk of personal information being leaked. The techniques vary greatly in terms of complexity and effectiveness. The least efficient way to protect personal information would be to reduce the amount of personal data that companies gather from customers. As more information is being collected, the value of such information increases, therefore it is a bigger target for leakers. This method of reducing the amount of data collected is not the most suitable solution for two main reasons:

1. As the size of the data being collected from a customer decreases, the company will have less information to analyze, thus decreasing the value of the

information that was going to be used for internal purposes such as marketing campaigns…

2. No changes in security. This technique is based on the idea that as information loses value the likelihood of this information being a target for leakers decreases as well.

This is just an example of a simple way of attempting to protect personal data. Another technique, that is more complex but more efficient is *Data Anonymization*. This technique has several different methods, those being:

- Data Masking: This method consists on altering values of the database with symbols. A mirror of the database with the customer's information is created. In the mirror version certain personal values are replaced with symbols making it impossible for a reverse engineer process to find our what the data is. This advantage of this method is that it assures that data will remain private. The disadvantage is that it must go a certain amount of changes in order to guarantee that reverse engineering will not be useful to find out the hidden values of data.
- Pseudonymization: This second method is much simpler. It consists on changing the identity of the customer with a fake pseudonym, this way if information is leaked, the identity of the consumer will not be given away. The advantage of this method is that it maintains the validity of data, since no significant values are being modified. The main disadvantage is that is more susceptible to reverse engineering, by using information that has not been modified, such as the zip code, birth, gender…
- Generalization: In this third method, the goal is to eliminate information can be used to identify a subject and does not provide significant value to the firm. An example of this technique is frequently used in addresses, where data collectors keep the name of the street where the customer leaves, but delete the number of the apartment. The advantage of this method is that the accuracy of data is significantly modified while eliminating some of the characteristics that can be used to identify the customer. The disadvantage is that some relevant information can still be accessed.

- Data Swapping: This method has different names; it can be called data swapping or shuffling and permutation. This method consists on rearranging values of the original database without affecting the statistics of the data. This effectiveness of this method when protecting privacy of customers depends on what data is being swapped.
- Data perturbation: This technique changes the database without affecting the statistical values. They do so by applying different techniques such as rounding numbers and adding random white noise. The main advantage of this technique is that the data is being modified without affecting the positive impacts that it can generate in the company. The main disadvantage is that the effectiveness of this technique when trying to achieve anonymization depends greatly on the size of the data set.
- Synthetic data: This technique is the most secure when trying to achieve complete privacy. A new artificial set of data is generated thanks to observations made in the original database, this way no values have to be modified and the privacy is being ensured. The artificial database is created through statistical observations that are found in patterns of the original dataset.

Anonymization Technique	Effect on data validity	Effect on privacy	Observations
Data Masking	Loss of information in the mirror version.	Sensitive to reverse engineering when not done properly	As protection of data increases, the value of it decreases.
Pseudonymization	No relevant information is lost in the process	Sensitive to reverse engineering.	No loss of value of data, however risk of privacy breach is present.
Generalization	Loss of accuracy but not to a significant level	If data leak, specific information will not be disclosed	Effect depends on what value is being generalized.

Data Swapping	No effect on statistical values	Degree of anonymity depends on values being swapped	If not done correctly, it is very susceptible to reverse engineering. Not the most recommended method
Data Perturbation	No effect on statistical values	Depends of the size of database	Effect depends on the size, so very efficient with large database, but not recommended for small ones.
Synthetic Data	Even though the dataset is artificial, the validity of data remains the same.	Best solution to keep privacy and security of customer data	More complex than the other methods but good results when marinating privacy

Table 2 Summary of anonymization methods

Source: Own elaboration from: Imperva

The goal of these techniques is to protect customer's data in case of data breach from hackers, data leaks… However, in many cases the data that is being shared without the consent of customers comes from the company itself. In this case it is the responsibility of the company to be held accountant for this actions and the misuse of personal information.

3.1.3.Privacy Breaches

The purpose of this section is to demonstrate how a breach in privacy and leakage of personal information can come in many different ways. The following cases are some of the most notable privacy leaks of the last few years, outlining the cause and the impact.

Marriott Starwood Hotel:

This case is very significant due to the amount of people's personal information affected by this breach in security. It is estimated that around 500 million guest's data

was stolen when hackers broke into databases of the hotel. In a statement released by the hotel chain they stated the following ‘‘for approximately 327 million of the guests, the information includes some combination of name, mailing address, phone number, email, passport number…’’

This is your digital life Quiz:

This was an online quiz that would pay those who completed it. The targets of the quiz were Facebook user. The goal was to collect information about those who completed the quiz by obtaining relevant information from their Facebook profile. The problem with this quiz was that it was allowed (Because of the Facebook Privacy Policy) to not only collect information from the respondent, but also the entire group of friends that the individual would have on Facebook. The total number of people affected by this scandal amounted to 87 million. The data was later sold from the developer of the quiz to Cambridge Analytica.

The cases before are very different one from the other. The first case is an example of a breach from external source, such as hackers. The second case was a result from poor privacy policies from Facebook, that allowed the developer to collect information from the social network website to sell it to third parties. The objective of this comparison is to demonstrate that protecting the data from external sources isn’t enough as long as the internal policies of data protection don’t have the customer as their priority. The damage caused in the Marriott case could have been mitigated if applying certain techniques that were mentioned in the previous section. By applying such techniques, even if the hacker accessed the data, the amount of personal information would have been lesser, reducing the impact on privacy breach.

3.1.4. Drawing the privacy boundary:

The previous sections consisted on how to protect the data of customers from leakage, but for customers, sometimes the own company they are working with, and to whom they agreed to share their data can be a privacy threat. In this section I will discuss privacy from the customers perspective. From the types of data I discussed previously, there are some aspects that can be sensitive and that would not like to be shared with

others, even if this personal information is just used for internal purposes of the company. So the question for firms is where to draw the line?

Starting from the beginning, structured data doesn't suppose a threat to customer's privacy and the information that is gathered should be analyzed to help the company in its process of becoming more efficient, this information however should be protected from leakers using some of the previous methods mentioned previously. The issue arrives with unstructured data. This segment is so broad that there is not an exact definition of what data is ethical to collect and what not. The problem with today's technologies is that companies collect information even when you are not in their websites, through cookies. To prevent the violation of privacy of customers, companies should focus on collecting data from direct interactions between the company and the customer. In 2017 Winston Smith sued Facebook from tracking his data on a cancer patient website with the purpose of profiting from this by later on posting advertisement on Facebook. Companies should be able to gather data and collect information about customer when they are interacting together, but anything beyond this point is something most customers would not be comfortable with, and as Martin's study mentioned above showed, privacy is a major concern for customers, and a violation of this privacy can have negative effects for any company. The problem in today's society is that the value that most companies give to profit and becoming efficient blinds them from the primary focus that companies should have, keep customers satisfied. Therefore, in order for companies to be ethical regarding the privacy of their customers, they should inform them of tracking activities through the web if they are doing so, and customers should have the ability to disable data collection at will. To do so, its necessary for companies to be socially responsible and take into consideration the damage that achieving their goal of efficiency will cause. The end does not justify the means.

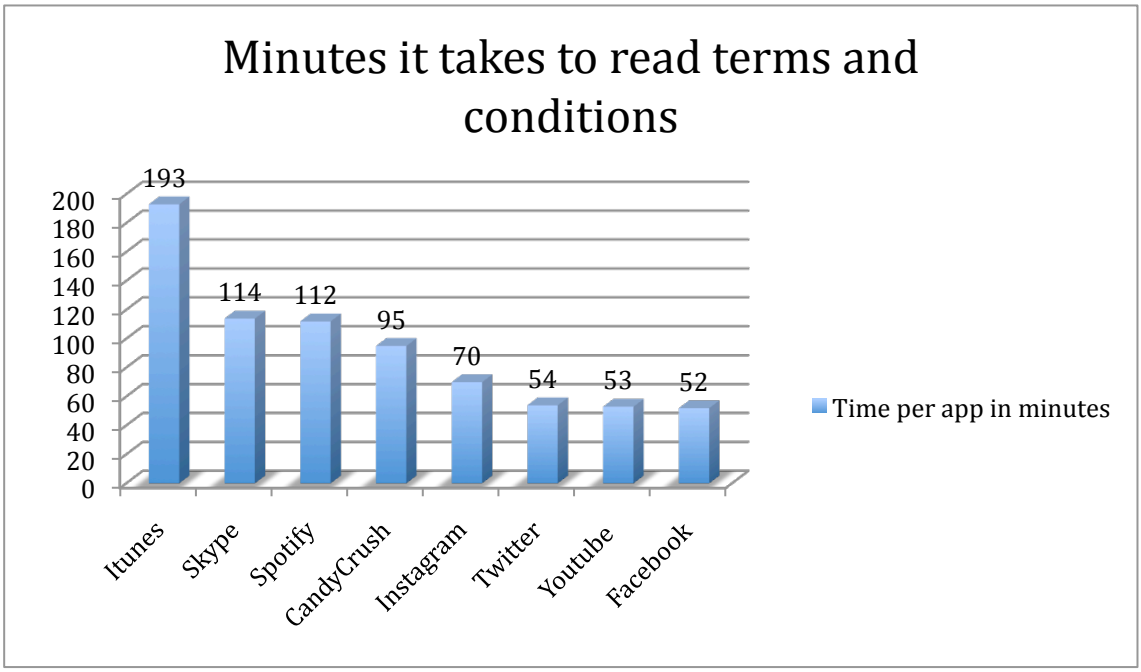
3.2 Transparency:

3.2.1. What is Transparency?

This aspect of Big Data is not given as much importance in terms of how it's controlled. The GDPR discusses this topic in Art 12, according to this section of the regulation, when dealing with transparency and communication they state that companies are obliged to disclose any type of information that they have collected and processed about customers if they demand to do so. The information disclosed must be presented in a very specific format; it must be concise, clear, intelligible and easily accessible form. The information that customer are entitled to request includes how they are processing the data of the customer and what is the purpose of the data collection, as well as with whom it is being shared. Companies are obliged to make it easy for the customers to request the information and the delivery of such information must be done in an efficient manner.

Although this section is useful because it provides a clear definition of how communications between customers and companies should be when they request information, they exclude a very significant part that is of crucial importance when talking about data collection and transparency. The regulators of data protection make no reference about how companies should ask customers for their data. Companies disguise some of the information that they will be collecting by offering contracts that are extremely long and not clear at all. In 2016 a group of Norwegians launched a campaign lead by the Norwegian Consumer Council with the sole objective of finding out how long it would take to read the terms and conditions that Smartphone applications force customers to accept for the use of their service (Noack, R. 2016)

When looking at the results, we can observe that the contracts are convoluted and of an extreme length, and the average person does not have that much time to read it thoroughly. Although some of the applications were local Norwegian application, there were many that the average person has on their phone. The following graph shows the results of how much time it would take to read some of the terms of agreement of these applications.



Graph 2 Minutes it takes to read terms and conditions contract

Source: Own Elaboration with data from Norwegian Consumer Council (2016)

How is this related to transparency? If companies have to be transparent about the information they collect once customers request it, why not be transparent about what you are asking them to accept? These contracts do not meet the GDPR requirements of concise, clear and intelligible. Do companies really expect people to sit and read 95 minutes the terms and conditions to play Candy Crush? Or almost two hours just to listen to music in Spotify? The contracts that customers are signing are designed to be so tedious to read that customers are forced to accept it without having the possibility to actually read it. Some could argue that this would define as transparent due to the fact that by the extent of the contract, everything they will be doing with your information is written down and presented to the customer before the contract is accepted, but the truth is that the lack of clarity violates the transparency principle. The same way the GDPR defines in a clear way how companies should disclose information when customers request it, they should define how companies should inform customers of the information they will be collecting from them when using their service.

3.2.2. Importance of Transparency

Although this topic is not given as much importance in comparison with the other two that have been previously discussed, it is important to mention Transparency plays a crucial role in the success or failure of a company. From the data collection and processing perspective, transparency is importance for three main reasons: Trust, allowing customers to make informed decisions and respect of customer's rights.

There is a direct relationship between transparency and trust, which leads to an increase in satisfaction of people. In 2007 there was a study performed in South Korea that focused on analyzing the impact that transparency had in governmental corruption (Park, H., & Blenkinsopp, J. 2011). The results showed that increasing transparency was a good strategy to discourage corruption and eliminate bad behaviors, this lead to an increase in trust by the public, which translated into higher level of satisfaction. If we translate this into a business equivalent, the governmental corruption would be the companies failing to inform customers of their practices, the bad behaviors would be misusing data and information which would cause a decrease in trust, causing customers to reconsider using the service, and losing market share would damage the companies profits.

The second reason focuses more on the customer's perspective. If informed properly, the customer is given the certainty that the decision he/she is making will be the correct one, since all the information available to make such decision will be at his/her reach. This links once again with trust and satisfaction, two pillars that companies should have present when managing customers relations. In addition, allowing customers to make informed decisions will ensure fairness, since the company will at no point be taking advantage of the lack of knowledge from the customers.

The last reason is related with the GDPR code of conduct. Due to the way the online services work, companies can collect information from people who have not asked for their service directly, since many companies work with technological partners who are in charge of operating the data collection and processing aspects. It is the responsibility of both the company that is working directly and the third parties

involved to inform the customer of such practices to ensure that customers do not lose their rights that have been established by the GDPR, such as right to access, right to erasure…

In conclusion, transparency's main link is trust, however other aspects such as ensuring customers rights are involved, hence it is of vital importance to guarantee transparency.

3.2.3. Why avoid Transparency

The reason mentioned above should be enough incentive for companies to enforce a policy of transparency with customers, however there are also reasons for companies to avoid doing so. Being completely transparent can have various negative effects on the company. The negative effects could include losing trust of customers, executives profiting from hiding information, reputational image… To illustrate this, Equifax is a perfect example of not being transparent about a data breach and how it can backfire.

Equifax case:

In 2017, Equifax, a credit rating agency was victim of cyber attack and the hackers successfully obtained data from approximately 150 million clients. The attack was a result of lack of security in their data storage software, the company was aware of this weak point since the US-CERT[1] had notified the company of the necessity to redesign this software to protect the company from possible breaches. The company failed to protect the data storage software and were victims of the attack mentioned above, and the personal information of approximately 150 million consumers was stolen. The answer of various upper level management members was to keep the information private from the public alleging that disclosing the information would make them a target for new attacks, however during this time four senior executives, among which we the CFO could be found, sold shares of the company for a value that amounted to 1.8 million USD. (Rasalam, J., & Elson, R. J. 2019)

The executives profited from hiding the information to the public since after the information was disclosed, the price of shares dropped around 50 USD in a period of

time of two weeks. This practice was not only unethical because of their wrong behavior of executives, but by failing to inform customers of such information, trust by customers was lost, resulting in a loss in company's value. The problem of Equifax was that executives gave more value to their wellbeing than to the right of customers.

3.3. CONSENT:

3.3.1. What is consent?

In the Big Data world, consent is necessary to gather information from customers, since otherwise it would be a violation of their privacy. Consent is commonly defined as agreeing to something, however, when we talk about giving consent to companies to gather information about us, there are more factors included in the GDPR than just simply agreeing to the terms the companies set. The GDPR is very quite specific regarding this topic. When discussing consent and the requirements for it to be valid the GDPR states the following:

''The basic requirements for the effectiveness of a valid legal consent are defined in Article 7 and specified further in recital 32 of the GDPR. Consent must be **freely given, specific, informed and unambiguous**. In order to obtain freely given consent, it must be given on a **voluntary basis**. The element "free" implies a **real choice** by the data subject. Any **element of inappropriate pressure or influence, which could affect the outcome of that choice, renders the consent invalid**. In doing so, the legal text takes a certain imbalance between the controller and the data subject into consideration''

Based on this definition it is clear that simply agreeing to share your data is not enough for consent to be valid, and the responsibility of ensuring the validity of such relies on companies. The most characteristics the stands out the most are freely given, specific, informed and unambiguous. Freely given implies that the customer is willingly to accept with the terms that are being shown because he/she agrees with them, and not because other factors are influencing the decision. Specific and informed go hand in hand. To ensure the validity of consent, the customer must know what data will be obtained and processed, how the data will be used, with whom the data will be shared and with what purpose… The length of contracts that was discussed earlier in the transparency section should also be taken into account when discussing consent. The format should be concise and clear, which proved not to be true by the Norwegian Consumer Council. In this section, it is also important to mention that the GDPR specifies that companies can gather information but must inform customers of a right called *right to object* (Art. 21), which gives the customer the right to ask the data controllers to not process their data. The last requirement, unambiguous, focuses on ensuring that customers give a clear affirmation of the agreement with the terms that

the companies propose. From an ethical point of view, this is the requirement that raises the least amount of issues that can affect the validity of consent, since the way this ''clear affirmation'' is given is very specific. The following image is an example of how this confirmation is made on most websites.

Image 3 Facebook method of obtaining consent from customers

Source: Screenshot from Facebook's homepage obtained the 14 of April 2020

This image is a screenshot from a Facebook's website. As soon as a user enters the website a window opens with information about the use of cookies explaining how if they continue using their platform they are agreeing to the terms of data collection and processing. By adding this popup, it is practically impossible for the user to no see this text, so if it is ignored it is the customer's fault for disregarding the message.

3.3.2 Issues customers face regarding consent requirements:

Although the GDPR makes it clear that consent should be given in a specific way, companies find ways to twist the definition for their own advantage. The GDPR specifies that consent should be given freely, without external factors affecting the decision. In reality this is not always true, since many online service providers do not allow customers to use the service unless they agree to the terms of agreement. So from an ethical point of view, we must ask ourselves if this meets one of the most important ethical principles, Fairness. When customers are forced to accept data sharing just because they want to use a service does not really meet the free choice requirement, since customers may be reluctant to share some information, but have a necessity to use that online service so they are left with no other option that accepting the terms that the company is offering.

3.3.3. *''Goodbye Big Five''*

To put that into perspective, Kashmir Hill, a journalist for the technological weblog Gizmodo, carried out an experiment to show the lack of power that customer has and

how dependant we are on online services. The experiment consisted on blocking the big five tech companies (Amazon, Apple, Facebook, Google and Microsoft) for a week each. The goal was to try to prevent these companies from getting access to her sales, data and attention. In order to block these tech giants she would block the websites owned by them (Hill. 2019). Once she concluded her study, she came to the conclusion that in today's digital world, its virtually impossible to live without having the service providers mine data about you, even for simple tasks like talking to a relative or communicating with a coworker, we rely on the service they provide. This raises awareness regarding the free choice, because customers should ask themselves if they agree because of personal beliefs of data privacy and are willing to authorize processing of their information, or simply they are forced to agree in order to not ostracize themselves in today's society, where if they do not use these services they would be an outcast of society. As a collaborator of the study named Daniel K. Gillmor says, ''I have the capacity to make this choice (not get his data mined). I know a lot of people would like to sign off but can't for financial reasons or practical reasons''. Gillmor believes that quality of live of people would increase if their data was mined and monetized, and although this though is a big generalization, it would be ideal if the customers had the power to make the decision for themselves, and not to be able to be a functional being in today's society.

3.3.4. Possible solution to freely given consent

The solution to this problem is similar to the one proposed in the section 2.1.4 Where to draw the line. The solution is a bit idealistic, because in reality companies are not willing to give up most of the data they collect, but some companies apply it, so if some are able to succeed using this technique, why cannot others do the same. In the case of El País, a Spanish newspaper, they allow customers to decide whether they are willing to share information and personalize to a certain degree their privacy settings.

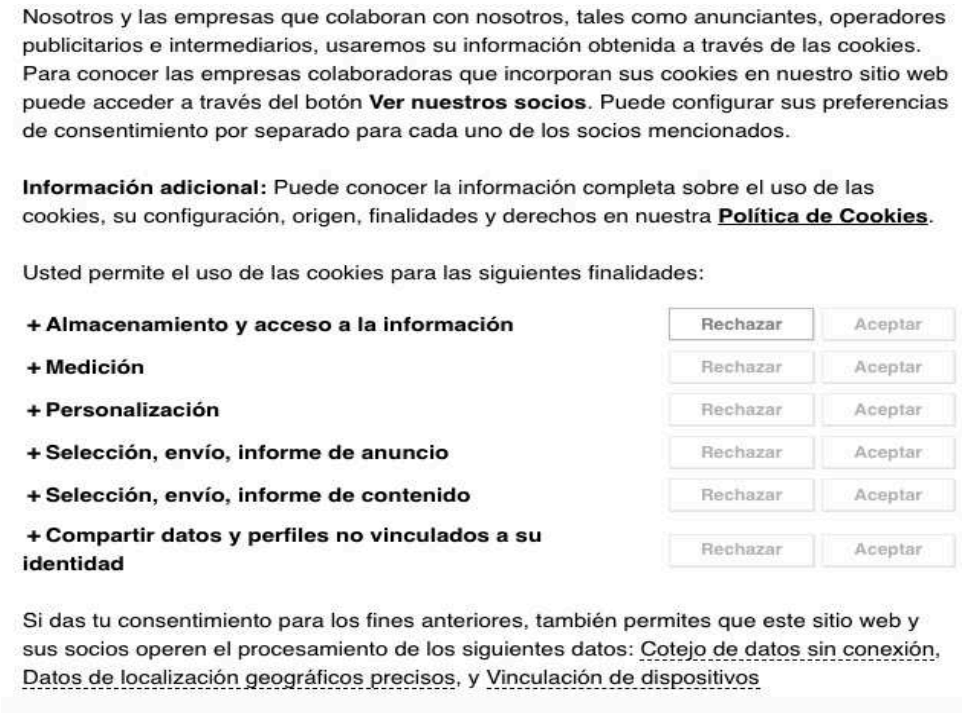


Image 4 ElPaís.com method of personalizing data collection from readers

Source: Screenshot from Elpaís.com consent settings obtained the 16th of April 2020

Many other online service providers use this type of setting when asking for consent, but it is not a universal method. This would ensure that consent is given freely, since in this case the customer has more control about the decision he/she is making. However, for this to happen companies would have to be willing to give more control to customers, which is not common in many cases due to the impact it can have in profits of the company derived from the data collected. Data collectors value more the information that they will gather than the free choice of customers.

3.3.5 Freemiums

There is a business model that is gaining popularity with the new online services called *Freemiums*. There are different approaches to this model, the most common, is providing a fraction of the total service, so by using the free version the customer will only have access to certain access of the service. Another approach is to differentiate premium user from free users by not placing ads in those who have the ability to pay. The issue with this method is that in order for companies to maximize the profit from placing ads, they gather data from users to make the advertisements that they show to

the free user more attractive, since the effectiveness of ads will be greater if it includes goods that the customer likes or has knowledge about. This raises once again the issue of inequality and fairness. From an ethical point of view, it does not seem fair that having higher purchasing should allow to keep your data more private, since the *Freemium* service provider will not have an incentive to gather as much information from this type of user because they cannot place advertisement for them to see. The data of those who lack the ability to pay for the premium service and are forced to use the ''Free'' version will be more valuable for the company, hence having an incentive to gather and process more data from this users. In conclusion, this method destroys the free given consent requirement, since now the financial status of an individual will affect the decision of deciding which type of service free/premium to use. This method of the Freemium model violates rights of Internet users from an ethical point of view because creates inequality among users, since those that cannot afford the premium service will be obliged to disclose information.

4. CONCLUSION.

In this section of the paper, the overall conclusion will be presented from each of the three sections that have been mentioned, Privacy, Transparency and Consent.

Privacy:

Privacy is the core of all problems of the data world. The value that data processing adds to companies is too important for them to decide to minimize its use. As data collection will continue to grow due to the new technologies that are available, the control of it should follow. In 2020 the control over this aspect of customers live is far from the ideal point. Definition of privacy in data terms is very vague and customers are the ones suffering this. Companies are deciding what data to obtain and what to do with it, and with the value that data has nowadays they have no incentive to take into consideration the customers perspective. This results in a lack of ethical behavior from companies in terms of privacy of customers; however, this is the fault of the organisms in charge of regulating this subject. Much progress is required for customer's privacy to be safe.

From the customer's perspective, there is not much that can be done if they want to be a functional member in today's society and keep private information from companies. The lack of guidelines indicating the difference between customer's private data, information that they would not want anyone to know, and public data, information that they are wiling to share with the companies they work with, prevents customers from having the choice about their privacy and how to protect it. The GDPR should work on segmenting data into what should be accessible to companies and what data should be considered to access.

From the company's perspective, it is their responsibility to maintain an ethical behavior regarding how they obtain data, what data they obtain and how they use it. There is a trade off between profits from data processing and privacy of customers. As a result of the value that data has, most companies are reluctant to give up such data, in those cases where companies are not willing to give up collecting sensitive

information, companies should pay special focus to minimize the risk of exposure of customers personal information in case of a breach.

Transparency:

Much like privacy, transparency is a victim of lagoons when the GDPR and other organisms treat this topic. The Equifax case is an example of the importance of transparency and more clear guidelines should be given regarding this topic.

Transparency begins before customers and company sign and agreement, thus the GDPR should work on improving this area, since transparency of contracts has proven to be an issue for customers and a way for companies to take advantage of the situation to maximize their potential benefits. Guidelines should be created clarifying how the companies should ask customers for specific things, assuring that the format is clear and concise, otherwise companies can exploit it and customers suffer from this abuse.

In terms of transparency when handling breached and incapability to keep customer data private, although it may seem as a possible solution to not notify customers, in the long run the negative effects will be greater than if the company adopted a policy of transparency.

Consent:

In my opinion, although it may seem that consent is the section (out of the three analyzed in the paper) that has been defined in a manner that guarantees customer their rights and works to ensure their freedom, this is not entirely true.

The way that customers give consent to the use of their personal data still has room for improvement to empower the customer and guarantee a free given consent. The way consent is asked by companies does not face any ethical issues, however the companies that oblige customers to give consent in order for them to use their service should be looked upon, this limits the freedom of customers since some services are necessary for their daily activities and forcing them to accept terms they might not

agree just so they can fulfill a need does not seem ethical from my point of view. Consent asking method like ElPaís.com allow customers to make a free choice, giving more power and freewill to the customer. All companies should adopt this model if their objective is to offer a fair treatment to customers.

The way that regulation treat the topics discussed is far from ideal, the companies have significantly more power than customers. Many of the services that companies offer that result in the collection of data are of vital importance for its users, therefore to protect their rights the balance should redistributed. Companies oftentimes look out for their own interest, disregarding the impact that their actions might have in the community, in order to change this, stricter regulation is required and more detailed regulations have to be developed to create a safe environment for customers.

5. Bibliography

Colmenarejo Fernandez, R. (2017) Una ética para Big data: Introducción a la gestión ética de datos masivos [Online], UOC, Viewed 19 February 2020, <https://books.google.es/books?hl=es&lr=lang_es&id=Y45ODwAAQBAJ&oi=fnd&pg=PT4&dq=%C3%A9tica+contratos+big+data&ots=7_IfvPl4C-&sig=Nz9Rdlvo6QiVDo0e-KCStmrsxiw#v=onepage&q&f=false>

Desjardins, J. (2019). *How much data is generated each day?* Retrieved February 2020, from https://www.weforum.org/agenda/2019/04/how-much-data-is-generated-each-day-cf4bddf29f/

Flavian, C., & Guinalíu, M. (2006). Consumer trust, perceived security and privacy policy: Three basic elements of loyalty to a web site. *Industrial Management and Data Systems , 106*, 601-620.

Falkvinge, R. (2013). *How Does Privacy Differ From Anonimity, And Why Are Both Important?* (Privacy News Online) Retrieved March 2020, from https://www.privateinternetaccess.com/blog/how-does-privacy-differ-from-anonymity-and-why-are-both-important/

GDPR. (n.d.). *Consent*. Retrieved March 2020, from https://gdpr-info.eu/issues/consent/

GDPR. (n.d.). *A guide to GDPR data privacy requirements*. Retrieved March 2020, from https://gdpr.eu/data-privacy/

GDPR. (n.d.). *Art. 21 GDPR Right to Object*. Retrieved March 2020, from https://gdpr.eu/article-21-right-to-object/

GDPR. (n.d). *The Principle of Transparency*. Retrieved April 2020, from https://gdpr-info.eu/recitals/no-58/

GDPR. (n.d.). *Transparent information, communication and modalities for the exercise of the rights of the data subject*. (General Data Protection Regulation) Retrieved April 2020, from https://gdpr-info.eu/art-12-gdpr/

Hill, K. (2019). *Goodbye Big Five*. Retrieved March 2020, from https://gizmodo.com/c/goodbye-big-five

Hill, K. (2019). *I Cut the 'Big Five' Tech Giants From My Life. It Was Hell*. Retrieved April 2020, from https://gizmodo.com/i-cut-the-big-five-tech-giants-from-my-life-it-was-hel-1831304194

ICO. (n.d.). *Location Data*. Retrieved February 2020, from https://ico.org.uk/for-organisations/guide-to-pecr/communications-networks-and-services/location-data/

ICO. (n.d.). *Principle (a): Lawfulness, fairness and transparency*. Retrieved April 2020, from https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/principles/lawfulness-fairness-and-transparency/

Jain, P., Gyanchandani, M., & Khare, N. (2016). Big data privacy: a technological perspective and review. *Journal of Big Data* .

Kemp, K., & Manwaring, K. (2020). *Australia vs Facebook: Why the privacy watchdog is right to be taking on the tech giant*. Retrieved March 2020, from https://www.smartcompany.com.au/business-advice/legal/australia-privacy-facebook/

Leiva-Gomez, M. (2017). *Why Do Executives Hide Data Breaches?* Retrieved March 2020, from https://www.maketecheasier.com/why-do-executives-hide-data-breaches/

Martin, K. (2019). Breaking the Privacy Paradox: The Value of Privacy and Associated Duty of Firms. *Cambridge University* , 65-96.

Matheus, R., & Janssen, M. (2015). Transparency Dimensions of Big and Open Linked Data. *9373*, 236-246.

n.a. (n.d.). *Anonymization*. Retrieved March 2020, from https://www.imperva.com/learn/data-security/anonymization/

n.a. (2018). *Marriott: Data on 500 million Guests Stolen in 4-Year Breach* . Retrieved March 2020, from https://krebsonsecurity.com/tag/starwood-breach/

n.a. (2019). *Your Data Is Shared and Sold... What's Being Done About It?* Retrieved February 2020, from https://knowledge.wharton.upenn.edu/article/data-shared-sold-whats-done/

Noack, R. (2016). *How long would it take to read the terms of your smartphone apps? These Norwegians tried it out.* Retrieved March 2020, from https://www.washingtonpost.com/news/worldviews/wp/2016/05/28/how-long-would-it-take-to-read-the-terms-of-your-smartphone-apps-these-norwegians-tried-it-out/

Oracle. (2020). *What is Big Data?* Retrieved February 2020, from https://www.oracle.com/es/big-data/guide/what-is-big-data

Park, H. & Blenkinsopp, J. (2011) 'The roles of transparency and trust in the relationship between corruption and citizen satisfaction', *International Review of Administrative Sciences*, 77(2), pp. 254–274.

Rasalam, J., & Elson, R. J. (2019). CYBERSECURITY AND MANAGEMENT'S ETHICAL RESPONSIBILITIES: THE CASE OF EQUIFAX AND UBER. *GLOBAL JOURNAL OF BUSINESS PEDAGOGY*, *3*(3), 8-10.

Rivers, C. M., & Lewis, B. L. (2014). Ethical research standards in a world of big data. *F1000Research*, *3*.

Rouse, M. (2020). *Relational Database*. Retrieved February 2020, from https://searchdatamanagement.techtarget.com/definition/relational-database

Shamsi, J. A., & Khojaye, M. A. (2018). Understanding privacy violations in big data systems. *IT Professional*, *20*(3), 73-81.

Statista. (2020). *Number of internet users worldwide from 2005 to 2019*. Retrieved February 2020, from https://www.statista.com/statistics/273018/number-of-internet-users-worldwide/

Steyerl, H. (2017) Duty Free Art: Art in the Age of Planetary Civil War [Online], Verse , Viewed 29 February 2020 <https://books.google.es/books?id=7yFaDwAAQBAJ&pg=PA51&lpg=PA51&dq=60%25+of+uk+consumers+have+intentionally+submitted+inaccurate+information&source=bl&ots=QlaIWlQ0c-&sig=ACfU3U0gl0tCDb16Nzk7XrgsBxyiddCaBA&hl=es&sa=X&ved=2ahUKEwiaheq36-_oAhWhDmMBHdyECsEQ6AEwAXoECAsQLA#v=onepage&q=60%25%20of%20uk%20consumers%20have%20intentionally%20submitted%20inaccurate%20information&f=false>

Taie, M. Z. (2016). *Big Data: Types of Data Used in Analytics*. Retrieved February 2020, from http://blog.agroknow.com/?p=4690

Turilli, M., & Floridi, L. (2009). The ethics of information transparency. *Ethics and Information Technology*, *11*(2), 105-112.

Woollacott, E. (2016). *Man Called Winston Smith Files Lawsuit Against 'Big Brother' Facebook*. Retrieved February 2020, from https://www.forbes.com/sites/emmawoollacott/2016/03/19/man-called-winston-smith-files-lawsuit-against-big-brother-facebook/#5f76a4007e28

6. INDEX OF TABLES, GRAPH AND IMAGES