UNIVERSITAT POLITÈCNICA DE VALÈNCIA

etsinf

Escola Tècnica Superior d'Enginyeria Informàtica

Escola Tècnica Superior d'Enginyeria Informàtica
Universitat Politècnica de València

Legal part of the Big Data

TREBALL FI DE GRAU

Grau en Engenyeria Informàtica

Autor: Diego Montagud Rodriguez
Tutor: Filomena Ferruchi
Juan Carlos Casamayor

Curs 2016-2017

Acknowledgments

First of all, I want to thank my parents for all the support they have given me, especially to my father who has helped me with his knowledge in the legal field becouse his degree in law.
Then, I want to thank the help of Filomena Ferruchi, my tutor from Salerno and Juan Carlos Casamayor, my tutor from Spain.
Besides, I thank my colleagues José and Hector for all the help.
I especially want to thank Chema Alonso who is Chief Data Officer (CDO) in Telefonica, becouse of all the content about the big data that he has published, being one of the greatest experts in Spain.
I also want to thank Marta García for helping me with the project translation.
And finally, I want to acknowledge in general all those people who have made this project possible and all the people who have supported me in an emotional way.
Thanks to all of them, this work has been possible.

Synthesis

The big data is one of the most popular computer topics of recent years however, the legislation goes one step back. I have always been passionate about the Big Data world and so I have wanted to do a research project as complete as possible to make clear to all who read it how far the law regulates this new technology called Big Data. The laws that are going to be studied will be of European Union although, will also be seen the laws of many other countries.

I also want to study Big Data in depth becouse it is necessary to build a good regulation in the future. I would like to come up with ideas for a good regulation in the future. We need to be all together to be able to regulate one of the most complexes areas of IT. Our privacy is increasingly disappearing and we have to put solutions right now. That is why this paper summarizes all laws relating to Big Data until 2017.

In this document is going to study everything related to the legal aspects of Big Data until now, some cases of companies or the government and some proposals for improvements. The lack of information on this topic has been a drawback but, with the help of some experts, it has been possible to arrive to the end of the matter. In this project is exposed everything related to the legal aspects of the technology that is revolutionizing the world.

Content Index

Index of the figures

Index of Tables

CAPITOLO 1
INTRODUCTION

This document will adressed the legal aspects of Big Data. In recent years this new technology has emerged that only few people is able to define it in a correct way. Like all the important changes, this new technique has appeared with a series of risks and legal problems.

Because this technology is relatively new, most users do not know how it affects them directly. In this way, the objective of the document will be to inform about all the new techniques of the Big Data that affect the privacy of the users, as well as all the legal aspects that they undermine with emphasizing the European legislation.

Being such a broad topic, we will review all the different fields that can affect the legal part of a society that increasingly uses this technology in a habitual way. It tries to solve all the legal doubts that comprise this subject and also that it serves as prevention so that in the near future the Big Data is regulated in the companies of a clear way and without attacking the privacy of the clients.

1.1 Structure of the project

To solve the problem it will begin by explaining thoroughly the meaning and the uses of Big Data. This is due to the great confusion that exists right now even in advanced subject matter users. For this reason we will proceed to clarify all the doubts on this subject.

It will also explain how this practice is used in different types of companies and what impact this has on the privacy of the population. In addition, a tour of all the legal problems involved in this technique, and the different methods used today that are in a legal vacuum. You will also see the regulations from the base of Big Data, real cases and some of the possible solutions for a near future.

The structure that we are going to use is as follows:

1. Big Data concept
2. Big Data Utilities
3. Current Use by Multinationals and Marketing Companies
4. Privacy in the Big Data
5. Legal problems
6. Laws defined so far
7. Big Data, an in-depth regulation
8. The Future of Big Data
9. Possible solutions for Big Data in the future
10. Possible solutions in the legal field
11. Conclusion

1.2 The objective

The objective of this research project is none other than to inform of all the dangers that this new technique supposes and to try to help to create a regulation in the future. This will review all the regulations regarding this practice so far and propose new solutions.

There is an increasing lack of awareness among users about all of this. It is becoming increasingly complex and for this reason it must be addressed as soon as possible in an objective manner and protecting the privacy of those affected.

CAPITOLO 2
BIG DATA'S CONCEPT

2.1 ORIGIN AND EVOLUTION

From the 90's decade started to increase exponentially the information stored in computing formats. The popularization of internet supposed big troubles when it came the time to gather data.

In 2001 appears for the first time the concept of Software as a service (SaaS), the ERP modern systems and the Web services got strength. In 2007 a International Data Coportation study (IDC) predicts that in the next four years the information accumulated will reach 988 exabytes, warning, this way, of a future problema. If we consult reports of 2011 that quantity has been overtaken.

In 2010 appeared the ERP systems on the cloud, a huge revolution for the Big Data.

In 2014 was characteristic because of the huge increasment on the Internet of Things (loT), at that moment, existed 3700 millions of 'things' on the Internet. [1]

All these events are the former of this new fashion. At the point where we are, the information stored has been reached any prediction, most of the bid companies are saving every possible data even though they are not treating them now, in a near future is posible to take advantage of them. Most of the studies announce that this is the beginning. The quantity of data will be multiplied in the coming years an that is why we have the need to study them.

In recent years, Big Data it is being spoken a lot. Is a tendency that has been consolidated between the years 2011 and 2013 because of the kind of factors that have been commented earlier and whose have been incresing as the quantity of devices connected on-line, the significant increase of the broadband, popularization of social networks and a huge rise on the ease to acquire products that can be connected to the net.

If we go back around a decade, this concept was unviable. Besides of lack of data, the main reason was the lack of power at the time of managing all that information. With the decrease in the price of the computers in general, from now to the future, practically any Company will be able to manage enourmous quantity of data.

2.2 Big Data's concept.

The generalized idea that we have of Big Data is about a huge quantity of information that can be generated by a Company in his private data base. The user actions that can be saved, any informtion about the clients etc. But the facts are that concept is mistaken. When we talk about Big Data we are talking about thousands or even millions of data that can be generated every second and require a special treatment for their study.[2]

Chema Alonso, CEO of Telefonica, the most important telephone Company in Spain and founder of Eleven Paths, a Company which is focused on security, defines Big Data as "*The ability that is been ofered to us by the technology to process large volumes of data at a minimun cost and in an useful timing to be able to make choices*". It is important to mark the last part of this sentence, because this is the complex thing. The processing time is which makes the difference for a Big Data's study to be productive.[3]

At first Big Data is used by the companies as a help when it arrives the time to make choices. Once the information is processed, this Company can take one measure or another and it will be more efficient tan another Company that keep on making choices based on intuition or in another kind of data. The problema arises when this information is used in a way that violates the privacy of clients or of workers.

2.3 Big Data evolution data

Many companies are betting very hard, others say they have been using it for years. It is the future of the information without a doubt and the professionals of the sector are very interested in the subject. The following graphs show the constant evolution of this new technology.

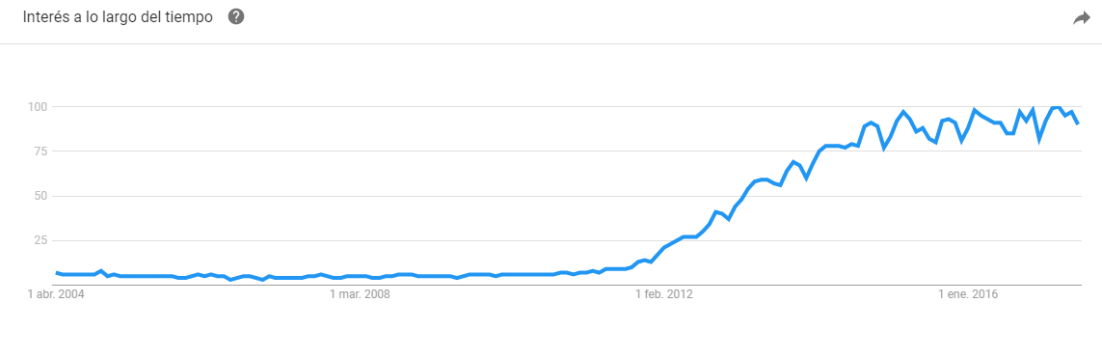


Figure 1 Frequency with which the words 'Big Data' are searched in Google provided by Google Trends [4]

Google Trends shows us how often searches are performed. For the term Big Data we can observe the result in the figure 1. If we look at the years in which the searches are frequented we can see how it was studied in point 2.1 Origin that coincide the years in which this technology becomes popular.

Since 2015 the search for information on these two words has reached 100% compared to other years. Since that year it has reached 100% practically every quarter of the following years.

These are just some facts about Big Data's rapid growth:

- YouTube users upload 48 hours of new video every minute of the day.
- 571 new websites are created every minute of the day.
- Brands and organizations on Facebook receive 34,722 Likes every minute of the day. (Source)
- 100 terabytes of data uploaded daily to Facebook.
- According to Twitter's own research in early 2012, it sees roughly 175 million tweets every day, and has more than 465 million accounts.
- 30 Billion pieces of content shared on Facebook every month.
- Data production will be 44 times greater in 2020 than it was in 2009.
- In late 2011, IDC Digital Universe published a report indicating that some 1.8 zettabytes of data will be created that year.
- In other words, the amount of data in the world today is equal to:
    - Every person in the US tweeting three tweets per minute for 26,976 years.
    - Every person in the world having more than 215m high-resolution MRI scans a day.
    - More than 200bn HD movies – which would take a person 47m years to watch. [5]

CAPITOLO 3
BIG DATA UTILITIES

3.1 Non-profit utilities

Big data utilities cover a wide range of possibilities. We will distinguish between the non-profit utilities and the utilities that Big Data has in companies and his problems. There are a lot of non-profit utilities, among which we can find:

- **Medicine and health:** Stadistic studies on this field are very improtant. The novelties and tolos that facilitate the desing of strategies and to materialize personalised diagnostics for each person.
- **Climatology:** the quantity of sensors installed currently on satellites as in the Surface allow us to know perfectly what happens on real time in our planet. Besides that with the technological advances we can increasingly make more accurated forecasts of weather and natural disasters.
- **Sport:** the applications that thanks to the Big Data allow us to know and share our route and relevant information are increasingly common and accurate. These applications not only gather information at user level, but they run stadictic studies between all connected users.
- **Defence and security:** security against atacks and cyber attacks, fraud, terrorism, have been affected thanks to this new technology. Very used also in the army to design strategies, to make diagnoses and to elaborate treatments with greater detail and sophistication.[6]

3.2 Profits in companies

Generally small companies can't afford the technology needed for large-scale statistical study. On the other hand big companies are making a great investment to be able to store and to treat the data. It is a complex and expensive task but one that with any doubt has a great benefit for the company.

Big companies are finding new ways to get value to the data they have stored. Each company seeks different goals with data processing, however, we could summarize the majority of cases on making better choices, improving operations and reducing risks. But before we will see the problem with the three "Vs": Velocity, volume and variety.

Let's start by defining each one of the three database attributes for the Big Data study.

- **Velocity:** Refers to the speed at which data is processed. It often presents a problem given the amount of data that is generated per second. For organizations, the speed with which information is handled is important. Sometimes even if information is analyzed in real time if you have the right technology. This is very important to be able to make decisions on time and reduce risks.
- **Volume:** Indicates covering, processing and storing an enormous amount of data for its later processing. Unlike the traditional statistic that studies a sample of the population, in this case is studied each and every single one data. With this amount of information you can no longer analyze with common tools like SQL or Excel. It is a quantity of data that doesn't fit on a conventional hard disk. Large data warehouses are needed, sometimes distributed throughout the world facilities equipated with necessary technology.
- **Variety:** Finally we face the problem of variety or difference of data. The data to be processed or saved is not all the same, this means that it is necessary to store from banking transactions to satellite images, information from social networks, web page contents, geolocation of mobile devices, etc. This is a challenge for the study of the whole. [7][8]

A few years ago these three attributes of databases were totally incompatible among each other. Normally a combination was chosen in between two of the three according to the objectives of the company. Nowadays with the current technology they are no longer contradictory. In fact all three are used, in combination, for a better study of the information. The treatment of the data in each of the 3 attributes is a very complex and expensive challenge.

Now, we are going to see some of the great benefits that companies achieve with the treatment of Big Data. Obviously it varies from company to company but mostly the study of data is used to:

- **Greater knowledge of the client:** For this the companies rely on external and internal information.His goal is to understand and predict customers behavior.
- **Security:** Both to predict attacks and to detect fraud, predicting real-time attacks is very important matter.
- **Analysis of operations:** For decision makers the study of data is the key. It increases intelligence and efficiency.
- **Increase value:** The general idea is: the more data the better. In many cases they can not be processed at that time but maybe in the future. It is estimated that companies only analyze 12% of the data.[9]

These were the main benefits, however, this list shows many other benefits of implementing Big Data.

- Provide ideas from huge amounts of data from multiple sources.
- Real-time monitoring and forecasting.
- Ability to find, acquire, extract, analyze and connect data.
- Identification of important information.
- Ability to mitigate risks.
- Identification of the root causes of failures.
- Full understanding of the potential of data-based marketing.
- Offers generation to customers based on their shopping habits.
- Improved customer engagement.
- Customizing the customer experience. [10]

For all these reasons, companies are increasingly betting to trust the Big Data. In the book Big Data, Analysis of Large Data Volumes in Organizations is described the following:
"*According to a study by IDC Spain, sponsored by EMC, JasperSoft, Microsoft and Sybase, the Big Data market is booming in the country. Data collected from 502 interviews with Spanish experts confirm that 4.8% of Companies have already incorporated these processes into their business, and forecasts indicate that in 2014 the adoption will be 19.4%, which represents an increase of 304% compared to 2012. The Big Data is beginning to show as an essential factor In 2010. The benefits generated by this technology in 2010 were around $ 3.2 billion worldwide, according to IDC estimates, this figure could reach 16.9 billion euros in 2015.* " [11]
In this extract from the book we can observe several facts. The first is the rapid growth of Big Data in the business sector. As we said before, not all companies can afford this technology,

and yet the forecasts indicated 19.4% adoption in Spanish companies. At the global level, according to data from SAS (US Business Analytic and private equity software), 12% of companies globally had this technology in 2015. [12]

Second, we can see how it is possible to quantify the benefits of this technology, which amount to $ 3.2 billion worldwide and with a forecast of almost $ 17 billion by 2015.

This data is a little short because in 2015, from IDC (International Data Corporation) calculate that the benefit generated by the Big Data each year are 122,000 million euros, anything to do with the comparison of the previous forecasts cited. [13]

In addition, an increase of almost 50% is foreseen for the year 2019, surpassing thus the 187,000 million dollars. Dan Vesset, vice president of analytics and information management at IDC, said:

"*Companies that are able to take advantage of the new generation of business analytics solutions will be able to take advantage of digital transformation and thus adapt to disruptive changes, creating a competitive differentiation in their markets.*" [13]

Because of all these reasons, Big Data has grown so quickly in virtually every industry. But this entails a cost, and large-scale data processing is a sensitive issue. In this way problems arise both legal and ethical some contemplated by the law and others that not currently, which we will see later.

CAPITOLO 4
CURRENT UTILIZATION BY MULTINATIONALS AND MAKETING COMPANIES

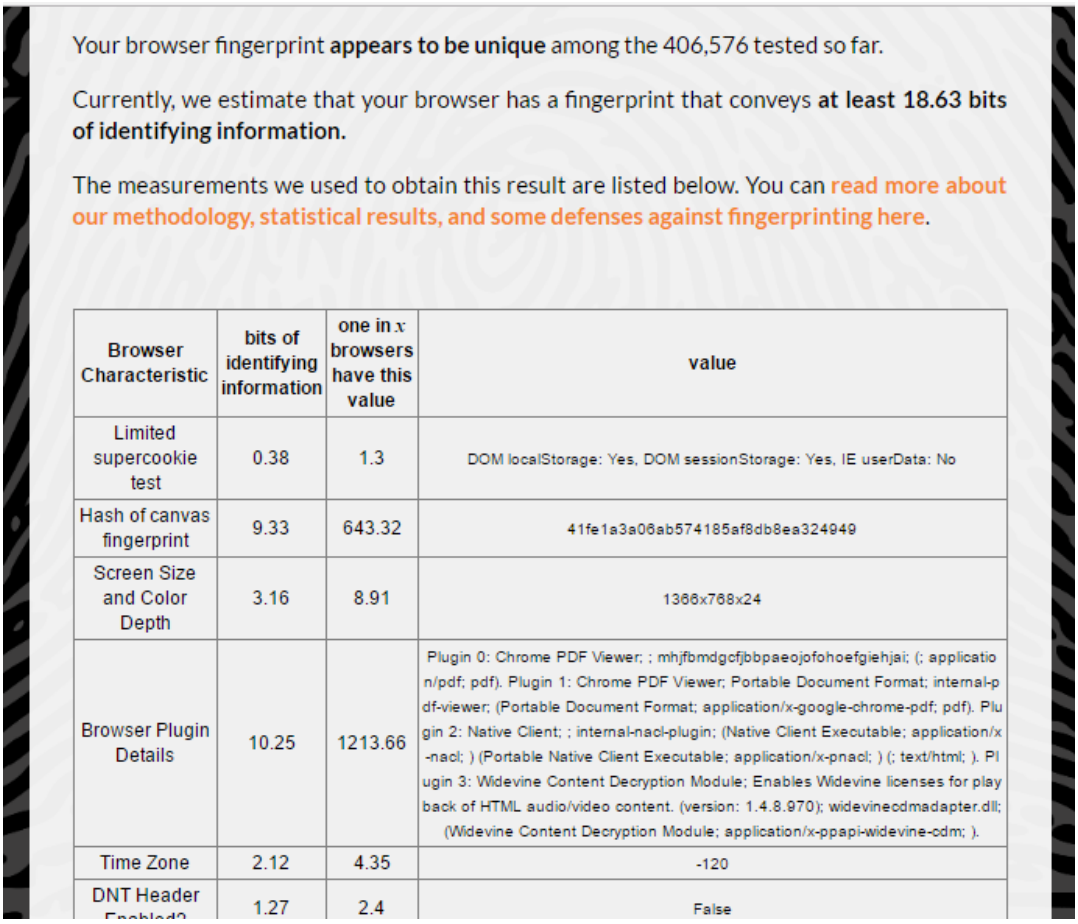
4.1 Online Marketing

As we have seen, there are many benefits that Big Data provides to companies. We will see next the way that the companies have to use this technology in its workers as in the clients.
"*Imagine an online store. There are all kinds and sizes. In principle we will think that only internet companies of the size of Amazon or eBay will make use of these technologies (and they do) but it does not have to be this way. A small store which is Online now uses web analytics applications to know what users are doing until they buy a product, which have been the most effective and efficient advertising and communication channels, or to discover usability or information architecture failures that prevent The user to find and use what the person responsible for the website would like to find and use.* "[14]
As Jorge explains in the article, not only big companies make use of Big Data. Any online store has a free of charge of a large number of applications and services that allow monitoring in real time or not and, in this way, can perform statistical studies for better decision making.
It should be taking in to account that the article is from 2011, in is the year in the entire analytical topic on the web still hadn't been exploited, and yet we found that there were great facilities for low-scale data processing. Marketing has always been a fundamental element in any business, whether small, medium or large, investment in this area is essential. Therefore,

it is important to know which campaigns are working more or less and which will need to invest more or discard them. So a crawl of each user in the web services is made to know from where it enters, in which advertisement has clicked and at what time, etc.
Online advertising has come a long way, customer tracking is increasing and we can see how personalized ads are, at this time, totally essential for campaigns on the web.
This advertising is tailored users-size is related in the majority of the cases with the account of Google, Apple etc. So if the same user connects to a device that they do not use regularly or that is the first time they use it but enter their account, all information related to that account will be downloaded to that new device. Likewise, both personalized advertising and their likes, preferences, etc. Will now be connected to the terminal.
The whole theme of personalized advertising is not a new topic and it is totally legal. The user has given the appropriate permissions and does not affect his personal privacy. However, the next step, the "fingerprinting", has emerged in the last year thanks to the development of Big Data and that this probably affects privacy and for which we dont have legal solutions at the moment. The concept and possible risks of the "fingerprinting" will be explained in detail below.

4.2 The fingerprinting

As explained above, each user is related to one or more Google accounts, for example. However when we connect first, these servers save all the information that our terminal sends. Thanks to this information it is no longer necessary to reconnect to know when a user is surfing the Internet. This is possible because of the fingerprinting.
Each time we click on a link or send a request the same information is sent. In addition devices have a series of public variables that the websites can get Access to. Thanks to all this a description of each terminal is formed. In this description enter values that differentiate one user to another without having to have any account at that time. They are values that are generally not changed and of which we will expose later. For a better understanding of the idea, there are different pages that show you all these values with additional information. For example the page https://panopticlick.eff.org shows you all the necessary data needed to form your fingerprinting.



Your browser fingerprint **appears to be unique** among the 406,576 tested so far.

Currently, we estimate that your browser has a fingerprint that conveys **at least 18.63 bits of identifying information.**

The measurements we used to obtain this result are listed below. You can read more about our methodology, statistical results, and some defenses against fingerprinting here.

Browser Characteristic	bits of identifying information	one in *x* browsers have this value	value
Limited supercookie test	0.38	1.3	DOM localStorage: Yes, DOM sessionStorage: Yes, IE userData: No
Hash of canvas fingerprint	9.33	643.32	41fe1a3a06ab574185af8db8ea324949
Screen Size and Color Depth	3.16	8.91	1366x768x24
Browser Plugin Details	10.25	1213.66	Plugin 0: Chrome PDF Viewer; ; mhjfbmdgcfjbbpaeojofohoefgiehjai; (; application/pdf; pdf). Plugin 1: Chrome PDF Viewer; Portable Document Format; internal-pdf-viewer; (Portable Document Format; application/x-google-chrome-pdf; pdf). Plugin 2: Native Client; ; internal-nacl-plugin; (Native Client Executable; application/x-nacl;) (Portable Native Client Executable; application/x-pnacl;) (; text/html;). Plugin 3: Widevine Content Decryption Module; Enables Widevine licenses for playback of HTML audio/video content. (version: 1.4.8.970); widevinecdmadapter.dll; (Widevine Content Decryption Module; application/x-ppapi-widevine-cdm;).
Time Zone	2.12	4.35	-120
DNT Header Enabled?	1.27	2.4	False

Figure 2 Some of the data that the browser can provide to the server in each request, given by https://panopticlick.eff.org

As can be seen in figure 2 we have certain values that confom our footprint since they are generally static. For example, the screen resolution will always be the same. The time zone will be the same in all the navigations that the user makes. You can change practically all these characteristics with different techniques, however, this is not the concern of the average user and its statistical use is the same.

The characteristics of the processor, which are responsible for the differentiation between users are as follows:

- Limited super cookie test
- Hash of canvas fingerprinting
- Screen Size and Color Depth
- Browser Plug-in Details
- Time Zone
- DNT Header Enabled?
- HTTP_ACCEPT Headers

- Hash of WebGL fingerprinting
- Language
- System Fonts
- Platform
- User Agent
- Touch Support
- Are Cookies Enabled?

These are the features showed in Google Chrome. Each one of these increases the differentiation between one user and another. All this information is sent online without the user showing its approval as they are public data of the system.

In case of my own device, as can be seen in the image, it is unique among 406,576 devices. This means that every almost half a million devices there is another with an equal configuration. If the configuration were unique among 10 devices, it would be impossible to recognize the terminal because 10% of computers would have the same configuration.

However, if the connection is made for example from Barcelona with a million and a half inhabitants, statistically, only 3 users would have the same configuration. It would then be very easy to trace to any user.

This new technology was not possible before, because all this information is generated millions of times per second. However with the advances of Big Data it is possible to analyze all this data and to know when a user is surfing (the internet) or not independently that it has entered in his habitual account of the navigator.

All this area has been professionalized and there are companies dedicated to analyze fingerprinting offering a new service. However, this shows a legal problem because at the present day we dont have laws which make studies in this area. The fingerprinting is a series of variables, as explained before, that are public in each terminal, for example, your device provides the consent to send this data. However, in the protection of personal data provided by European legislation we find the following regulations regarding the privacy of each user:

The EU data protection rules mean that your personal data can only be processed in certain situations and under certain conditions:

1. If you have given your consent (you must be informed that your data is being collected)
2. If the processing of the data is necessary for a contract, a job application or a request for a loan
3. If there is a legal obligation to process your data
4. If the treatment of the data is of vital interest to you, for example, if a doctor needs to access your private medical data when you have had an accident

5. If the treatment is necessary to carry out actions in the public interest or for the management of the administration, the tax agency, the police or other public organism. [15]

Each of these conditions are applied to consent to process personal data. If we study the points carefully, then we see how there are certain doubts and gaps that should be solved. For example point number 1, you must give consent and must be informed in advance. This is not the case but, as we said, they are public data of the terminal. The following 4 points are not met for the purposes of these companies that collect data for fingerprinting.
The main controversy in this matter is whether the question is personal data or not. They are only device variables and there is no personal information, however, we can know in most cases which user is the one that is connected and can follow it exactly without his consent.

4.3 Cookies without cookies

As the fingerprinting, there are some techniques, although not many, that thanks to Big Data allow you to discover which user is behind every request without his or her legal consent. These technologies will be very dangerous in the future if they are not legally treated and is only possible today thanks to the high computational power of modern equipment.
Cookies without cookies is a technique based on the eTAG, that means, In the cache of images, thanks to which, it is possible to know exactly which terminal has been connected whenever and not the first time and thus be able to register its behavior without requiring the user to accept a policy of cookies.
"*An entity tag (eTAG) is an HTTP header used for Web cache validation and conditional requests from browsers for resources*" [16]

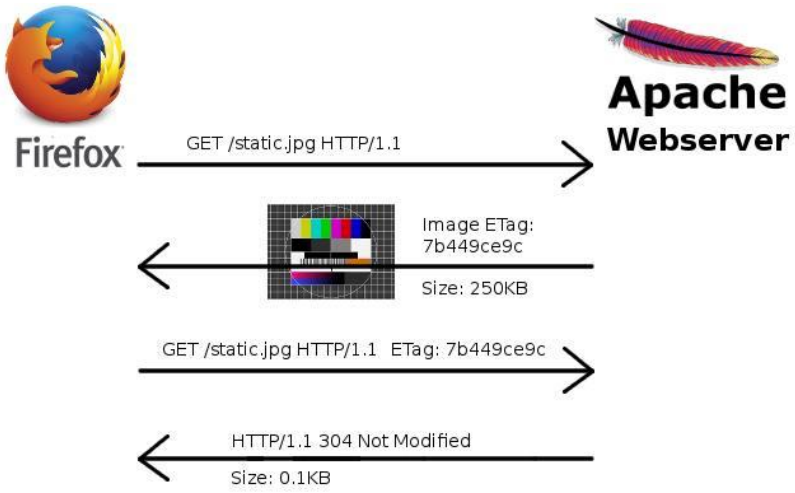


Figure 3 Scheme of the operation of the technique cookies without cookies [17]

Operation of this technique is very easy and with a computational cost not so high. To understand it, you need to know how the caché works in browsers. When a host sends an image to be decoded in the browser of any user, this image is stored locally so that the next time you visit that site does not need to re-download the same content.
If we look at figure 3 we can see the series of steps to make it possible to identify the user. At first request of the client to the server we do not see anything strange, simply a GET of a picture that it is necessary to be able to solve the HTTP. In the response of the server we see that it sends the image that previously has given a unique title for each terminal. This way you have a database with the different titles that you have issued and the corresponding actions that you want to save.
When the user returns and asks for the same image by providing the eTAG, in this case 7b449ce9c (any normal identifier, can follow any algorithm), it is then when the server processes that user is not the first one that comes and already knows exactly which terminal Is with your eTAG number. He simply returns a "not modified" response and has finished his work.
Both the latter technique and the fingerprinting, are in a legal grey sphere as it violates the user's privacy with procedures usual in virtually any website. All this information today is accumulated by some companies to save it in a large database of Big Data. However most of these data are not processed for lack of interest.

4.4 Use of Companies

As we already know some of the different techniques, the question arises for what companies want all this information without the customer knowing that they are following up. Although most of the information will be stored for different uses in the future, they are already being used for the following services:

1. **Fraud detection:** A company can have the fingerprinting of each of its customers and when it detects that there is a sharp change in this footprint, as discussed: the change of screen resolution, change of system fonts, etc. You can generate security alerts or simply prohibit for example a money transaction.
2. **Advertising providers:** It is one of the main applications on the topic of user differentiation. Obviously it generates much more benefit to be able to put personalized publicity to a user although this one is not connected to his session. [2]
3. **Tracking of hackers or terrorists:** There have been cases of being able to locate hackers or terrorists thanks to these methods. They are people who act from the shadows and try to protect themselves above everything else. They never have the

session started but they have a fingerprinting just like the rest of the world and thanks to this it is easier to follow them.

Although these are the main applications, it must be remembered that they are relatively new technologies and that, at this time when Big Data databases are being created with a large number of users with or without the session started. As Big Data advances, new ways of using all this information will be found.

Legally speaking, there is no regulation today that makes it clear how all this data should be handled. The truth is that it is a very delicate subject and that there is still a long way to go, for example, there is still doubt whether IP addressing represents a user, it is even less clear if the fingerprinting, which is a set of Each terminal also represents it. [2]

We are far both technologically and legislatively from being able to control this massive data capture. An average user can not do anything literally so that his information is not collected even if he does not want it and for an advanced user it is a practically unfeasible topic, for example, for the fingerprinting should change in each navigation the variables of his system and his Browser and the eTAG should directly change terminal.

4.5 Big data and user privacy

In this part we will proceed to make an example about the privacy of users. Social networks are generally a free service. This is not entirely true as the user gives in exchange for the service his data. It seems obvious that a social network, for example Facebook, users give their data in exchange for what the application offers. But it will show all the information that a company has when using Facebook as a marketing services provider. Not only companies can advertise but any type of user in a very simple way. These are the steps to follow:

Firstly, simply in the left menu, at the end you can see an access to Ad, as you can see in figure 4, this will be our ad manager on Facebook. This functionality has accounts of any type, there are no restrictions and anyone can advertise with this service.

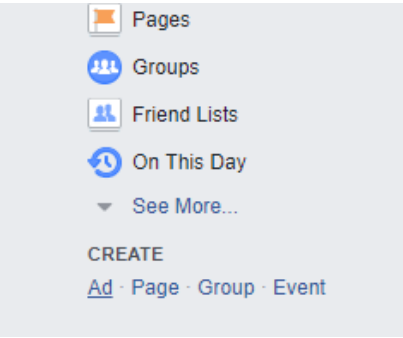


Figure 4 Place in the menu where the ad manager can be accessed

This directs us to the ad manager, you can see all the features in figure 5. At this point there are several tools according to the goal of the advertiser. From visits to a physical store, visits to a video or just traffic to a web address. Whatever the choice, the page will open to configure the advertisement.

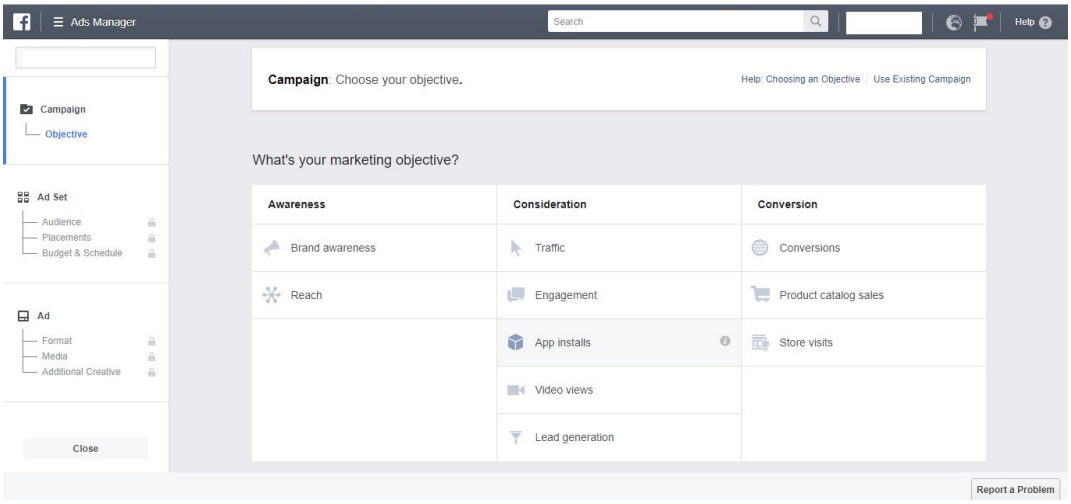


Figure 5 Facebook Ads Manager Menu

In addition to set up an image or video, and text information, Facebook allows us to target a specific audience. Here comes Big Data and all the information that Facebook has been compiling all these years. As you can see in figure 6, you can select a specific range of users with variables such as your country, age, gender, etc. Up to this point everything is correct. However we will focus on the "Detailed Targeting" option.

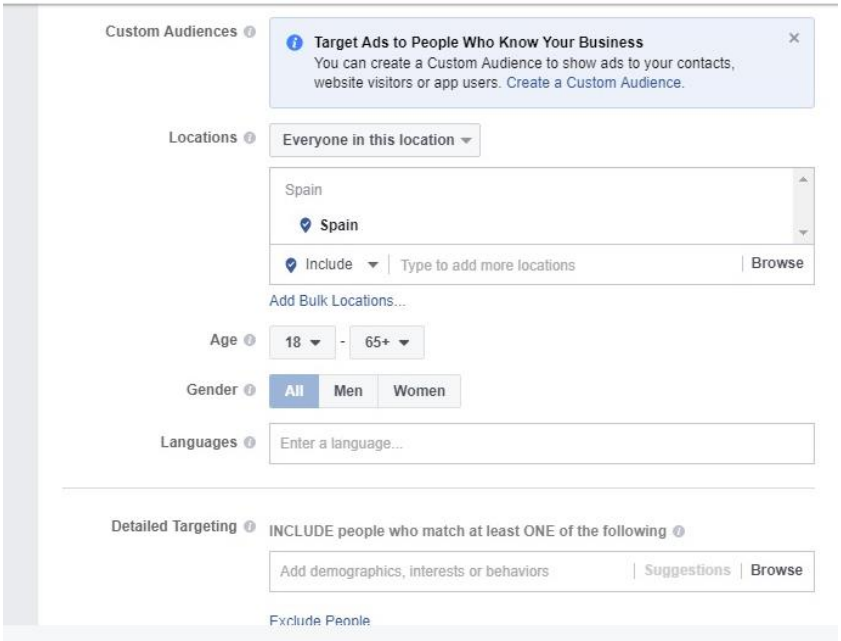


Figure 6 Screen of filters to apply to the public of the advertisement

Clicking displays a menu with four categories. In the figure 7 you can see each one of these. The advertiser can write a condition or simply browse through each of these four options. All the data that we contribute to the social network without realizing it are now embodied with the intention of limiting the maximum possible public the advertisement. All this information is very valuable as it is one of the keys in the Marketing sector.

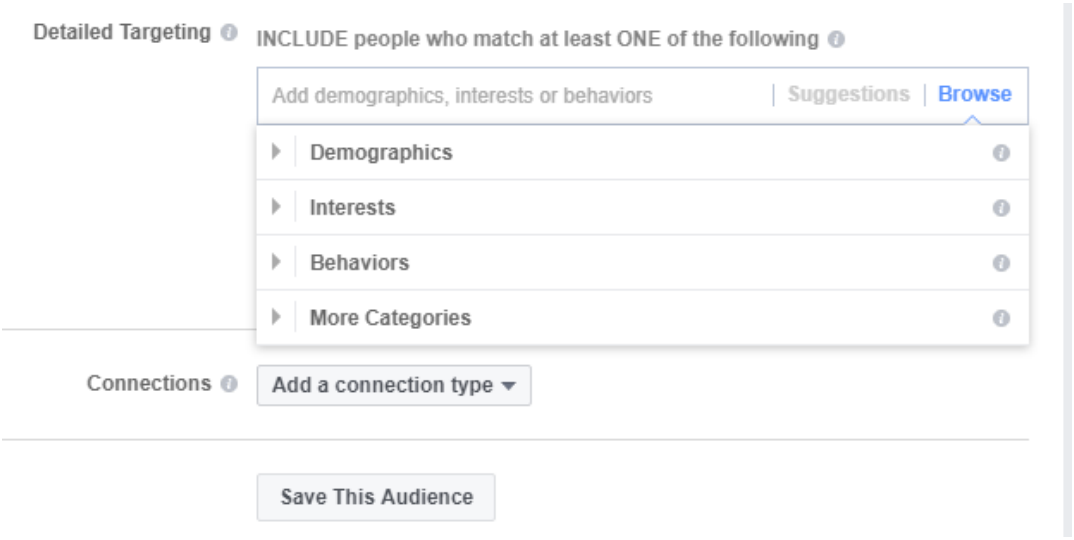


Figure 7 Types of filters applicable to the ad

We find dozens of options that can be put together to get the right audience. As can be seen in figure 8, there are conditions such as people having a distance relationship. In this premise there are just over 20 million active people on Facebook.

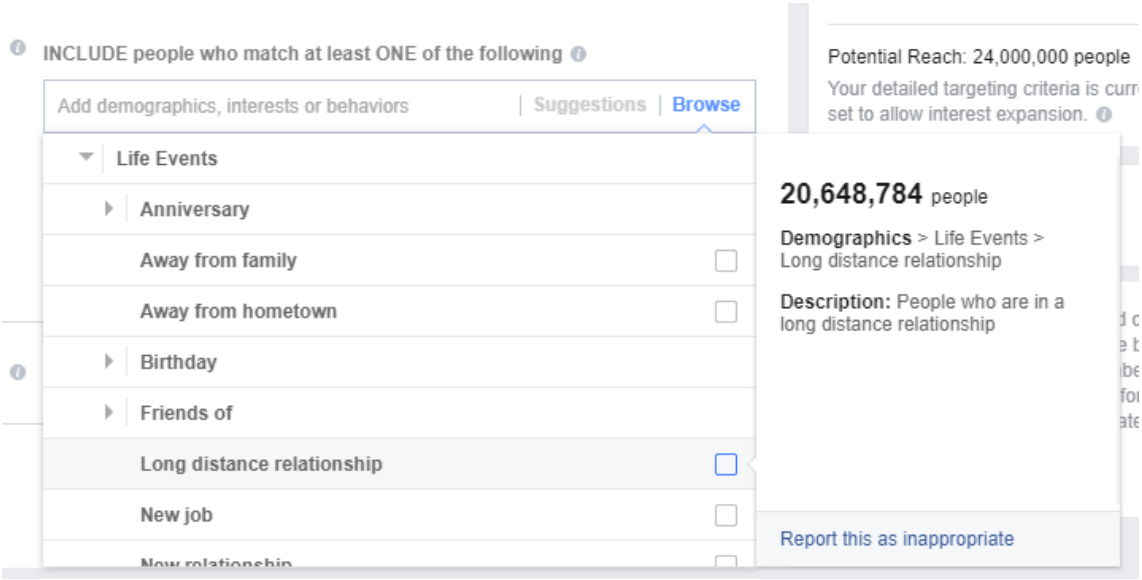


Figure 8 Example of a filter in the ad

This is totally legal, because in the terms and conditions of Facebook has already been previously informed of all the information they are going to collect and deal with. It is known by the users of this social network that they are going to collect their data, however there is an important point of ignorance of how far they collect the data, about how much Facebook is able to know about their own users. Here are some of the options that any advertiser can find to limit the audience who will see the ad later:

- Away from family.
- New job
- Recently moved
- Likely to engage with political content (conservative)
- Likely to engage with political content (liberal)
- Likely to engage with political content (moderate)
- Divorced
- Gyms
- Nutrition
- Engaged Shoppers

And a long etcetera of conditions in which the users are segmented. The average user of Facebook is not aware of this practice since few are the ones who read the Terms and conditions. By notifying Facebook in its policy is something totally legal and that users who do not want their registered information should simply stop using this social network.

CAPITOLO 5
PRIVACY AT THE BIG DATE

5.1 Privacy in the United States

There is a lot of evidence about the global surveillance network. Thanks to Edward Snowden who unveiled a complex espionage network of spy agencies in the United States in collaboration with more countries. This case was in between the years 2013 and 2015 and had a huge impact all over the planet, as the current lack of privacy was discovered.

The truth is that this topic was already foreseen and some experts advanced it, but before it came to light, all this were conspiracy theories. In the United States because of a series of terrorist attacks and actions, security was imposed before privacy. In fact, when Snowden filtered out the millions of files available on espionage, the US passed a rule that allows all companies on US soil to be permanently tapped and receive requests for information from public administrations. These administrations refer in most cases to the FBI, CIA and the NSA.

This regulation affects the rest of the planet in a very direct way. The United States has the highest quality servers in the world, and it is very likely that many pages that any user visits throughout the day are hosted on these hosts. Even if they are not hosted in the US, it is likely that some of the information will be forwarded for some type of web service and in any case, the government can process the information.

It is one of the clearest cases of privacy violation at user expenses. A person who has not accepted that they treat their personal information can do nothing, because United States' laws, and is true that in many occasions the information must pass through there. A European citizen has practically no defense against this violation that takes place outside its country and, therefore, it can not nor vote not to question the laws. [18]

Big companies, people with great reputations, users from around the world, and civil rights entities from the United States joined a fight to get regulation in the Big Data. This is a challenge for the legislation of all the countries of the world, because which is considered today on this advanced subject is very little.

5.2 Data Protection

It is a clearly complex task to be able to create legislation that encompasses all infinite possibilities of Big Data. As we have seen, most of the information that is covered in the Big Data is used statistically. However, data protection regulations apply when a user's information makes it possible to identify. The we have a new problem, since the data processed at statistical level do not contemplate this case, however we have seen techniques such as cookies without cookies and the fingerprinting that were able to identify a user thanks to their information.

Big Data threatens data protection regulations due to several reasons:

- **The regulation is not adapted to the new technological environment.** This is one of the biggest problems. It only considers certain cases in computer cases when they are currently those that charge more importance.
- **The principle of data minimization does not meet the practice.** As explained above, there is a point where the amount of data collected by the user can already be identifiable even though the data is anonymous. This is what happens in the fingerprinting. The principle of data minimization explains that only the very necessary data should be collected for its purpose. Data which does not serve the market purpose should not be gathered. This principle is currently not met and, despite being regulated, no penalty is observed in the Big Data.
- **The rules rely too much on the informed consent of the user.** This problem exists since the beginning of data collection in computer-related applications. Most users accept daily conditions that have not been read previously. There is too much responsibility on average users to accept from privacy conditions, cookies or a simple installation on the mobile phone.
- **Anonymization has been shown to have limitations.** Although it is a good solution, it can not encompass all the possibilities covered by the Big Data. It is a subject to be discussed in the next point.
- **Big Data increases the risk of decision making choices in our lives.** Big Data is already used to make different decisions by companies such as when to apply for a credit, to take out car insurance, life insurance etc. The problem is that this

decision making is performed by algorithms that can start from false information and the user does not know when the algorithm is who has taken a decisión.[8]

European Union, due to all these inconveniences, proposed a 'data protection' amendment in 2012. On 8 April 2016, the Council adopted its position at first reading. European Parliament then adopted the texts on 14 April 2016. A necessary reform due to data that they themselves give us on their page.
This data is as follows:

- "*57% of Europeans consider disclosure of personal information to be an important issue.*
- *70% are worried about companies using the information for purposes other than those for which it was collected.*
- *Only 15% consider that they fully control the information they provide online.*
- *90% of Europeans believe that it is important to have the same rights and protection in all EU countries.*" [19]

Despite this amendment, it is an area that must be constantly updated due to the promptness of the changes and that, generally, first come the computer advances and later comes the legislation trying to solve some of the problems that it faces .

5.3 Anonymisation of data

This is one of the most sensitive and key points throughout the Big Data revolution. There are two types of data that can be collected:

- **Anonymous data:** It must be impossible to know with the data who the user is.
- **Personal data:** These are personal data and through which you can know what person is behind them. In addition, they are subject to data protection rules.

Formerly, anonymization was a process that consisted of two phases:

1. Delete the data that were identifying traits. Personal Identificable information – PII - This data could be the name, address, date of birth, ID code, etc.
2. Modify other data that can act as identifiers. This explains why asterisks are used in some occasions to cover the numbers of credit cards, email etc.

This worked very well before the appearance of Big Data. However, with this new technology, by increasing the calculation power, the quantity, quality and diversity of the information

facilitates the possibility of identification of the users even after the data have been anonymized. Federal Trade Commission of the United States has stated the following: [8]

"*There is sufficient evidence to show that technological advances and the possibility of combining different data can lead to the identification of a consumer, computer or device, even if these data by themselves do not constitute personally identifiable data. Reidentify data that are not personal identifiers through various means, but companies have strong incentives to do so*." [20]

Obviously in most cases this technique is unethical but very useable for profits in companies. The question is whether it is a legal technique or not. At first glance it is not so, however we are going to study the legal framework on this subject.

The truth is that European regulations do not regulate anonymous data or the process of anonymising information. This is understandable since not many years ago the anonymous data meant that same. Data through which it is impossible to identify a particular user. It is now with the appearance of Big Data and the excess information when we encounter this problem that needs to be solved.

The most direct reference is found in Recital 26 of Directive 95/46 / EC, which reads as follows:

"*Principles of protection shall apply to any information relating to an identified or identifiable person. In order to determine whether a person is identifiable, all the means that can reasonably be used by the controller or by any other person, To identify such person The principles of protection shall not apply to data made anonymous in such a way that it is no longer possible to identify the person concerned(...)* "[21]

As can be seen, it explains that the difference between a person who can be identified and another who is not, who is anonymous, must be considered the set of measues that can be used by the responsible of the treatment or by any other person to identify that person. Therefore, as it was said the fact of identifying a person with anonymous data should be on most occasions punished. However, there are still doubts about this regulation. As for example, in the case where the treatment is not carried out by a person but by an algorithm. Or several, or if to decrypt the user are put together several sets of anonymous information. Or anonymous information is gathered with public. It is undoubtedly an issue that should be studied more in detail as not all possibilities are considered.

To confront this issue in more detail we must define the threshold by which we consider that the data are anonymous or not. There is a difference of opinion on this issue which is treated differently by the courts of each country. What is meant by anonymous data varies between:

- **Absolute anonymization:** It implies the zero possibility of directly or indirectly reidentifying any of the subjects.
- **Functional anonymization:** Implies an insignificant risk of reidentification. [8]

In the first case, absolute anonymity a few years ago was more common than today. It was often impossible to identify a user with anonymous data. However, with the advent of Big Data it is becoming more complicated absolute anonymization being, in some cases, practically impossible to obtain. It because of this reason that we must start regulating all information processing and redefine what data are anonymous and which are personal, in addition to being able to measure in which cases it is easier to reidentify users with their previously anonymous data.

5.4 Responsibility in the user

5.4.1 terms and conditions of the contract.

The responsibility of the user in the Big Data field is enormous. However, sometimes the user can not do anything so that they do not violate their privacy. One of the most important legal points is the contract terms. Most companies hide in the "Yes, I accept" the user.

Here are a number of issues to address, such as the conditions discussed above. If you observe the conditions of most of these applications, we find a long list that becomes a rather expensive job in time and effort for the user. According to statistics 88% of users do not read the terms of contract, so it becomes a somewhat dubious issue when it comes to blaming the person who accepts.

We are faced with cases like the following: A person in a tone of mockery or complaint tried to beat a record in reading the Terms and conditions of Kindle.

"*This is an initiative of Choice Australia magazine, which has hired a person to read, aloud and at a stretch, the 73,198 words contained in that important document.*" [22]

This user has taken 8 hours and 59 minutes to complete. This is a clear case about how little these conditions serve. It would be impracticable for a person throughout his life to read each and every one of these terms and conditions. However, they are the only ones that have legal validity. This means that companies are virtually in control of putting the conditions they want as it is very likely that users will accept them without reading them.

At legal level these Terms and Conditions have exactly the same validity as any signed contract. However it is socially accepted to be careful when signing a physical document and on the contrary to accept any report of terms and conditions without consideration.

The Internet Security Office (OSI) in Spain gives us several examples of the conditions of use that almost all users have accepted and yet few have read:

- "***Dropbox:*** *this tool, to have accessible from multiple devices all our information, reserves a clause that, knowing it, can make change our use of the same. Basically, the company reserves the right to suspend or cancel services in any moment. (...)*
- ***Google:*** *Any information that Google obtains from us, can be used for any of its services. This is a good thing to improve your searches and adapt them to our needs, but also use it to teach us advertising tailored to each of us. (...)*
- ***Facebook:*** *This social network is perhaps the most controversial in its conditions of use. Mainly makes it clear in its terms of use that everything you upload to your social network (photos, videos, states, information etc.) happens to be your property. (...)*
- ***Youtube:*** *In a similar way to what Facebook does with your information, so does YouTube with the videos that we upload. You have the option to delete them, but Youtube reserves the right to retain them even if they do not reproduce them.*
- ***Yosemite:*** *As is often the case with newer versions of Apple's operating systems, they are free. But you have to know that when you upgrade to the new version, you are not acquiring the operating system. Apple lends it to you to use it, and in case you change ownership of your computer, you have to leave it with the operating system with which you bought it, because it is not lent to the equipment, lends itself to the person who goes to use. (...)*
- ***Twitter:*** *Although this social network have good evaluations made on its policy of terms and conditions of use, it reserves the right to change them without consulting its users. With this it reserves an ace in the sleeve in case in the future I would like to own the rights of our tweets and photos as Facebook does, for example.*
- ***Game Station:*** *This is the funniest case and although it is a joke, it makes us reflect on the fact that by accepting the conditions of use of a service without first reading them carefully, we can actually be accepting abusive conditions without being aware of it. Game Station, to draw attention to this fact, included as a condition: "By sending a purchase order on the web the first day of the fourth month of 2010, Anno Domini, you agree to grant us the non-transferable option to claim, For now and forever, your immortal soul. " More than 7500 people accepted these terms*." [23]

At Big Data the most valuable information for companies is customer data. The more data the better as this is usually your business model. In the case of Facebook, the company offers you a free service, in exchange for each user to indicate their personal data and little by little their likes. In this way companies that advertise on Facebook pay a little more with the condition of reaching the public they dictate. Something similar happens on Youtube and Twitter.

5.4.2 Application Permissions

Other issue that leaves much responsibility to the user are the permissions in the applications. Any application, simple as it seems, requires a lot of permissions on our Smartphones. As you can see in the figure 9, the Whatsapp messaging application requires access to, for example, the camera, the contact list or the location.

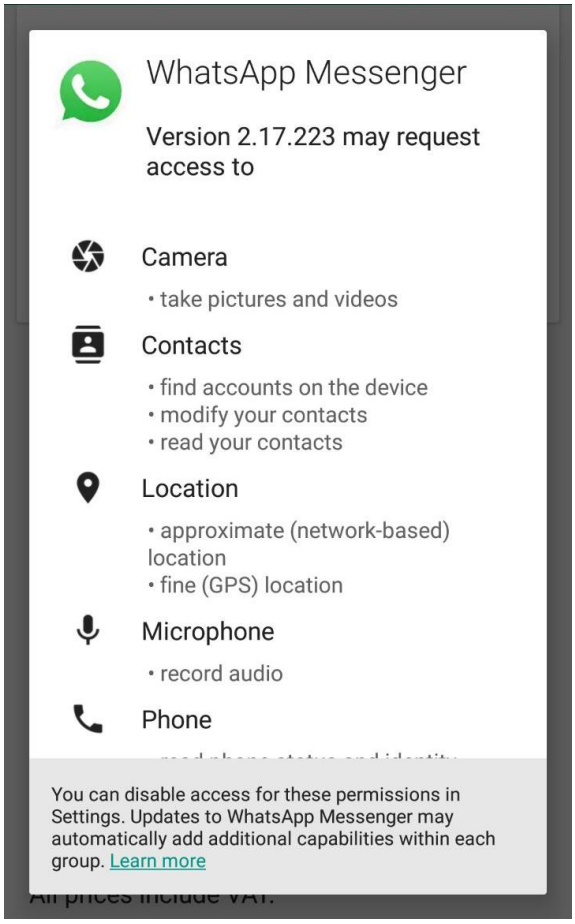


Figure 9 Permissions that the Whatsapp application requires in version 2.17.223 in the version of Android

In this case everyone accepts these permissions because it is a trusted application and the user knows what each one is for. The GPS permit for example is necessary to be able to share your location in a precise way.

On the other hand, not all users look at the permissions they accept. It is a similar case to accept the terms and conditions of the contract. There are a lot of applications that require permissions, although they are not used in any program functionality. Many of them send data for later treatment. Surely they have warned the user through the terms and conditions

and thanks to the permits a pack is created in which companies have free access to be able to study the user in detail with his legal consent.

One of the most common permissions is the exact location. Although there are several techniques to know where a device is without the need for a GPS sensor, thanks to this permit, the work is much easier. Many companies want to know the location of users.

Legally, the person who accepts the location permit is consenting to know much more than you can imagine. And that is why they rely too much on the responsibility on the users when there is no knowledge about the information that can be deduced only with the position. If you send data continuously from the exact location you can deduce for example:

- Where do you live, where do you go on vacation, where do you buy
- If you are walking, by car, by bicycle or if it is a combination
- If you go to the gym, which restaurants do you like to go to?
- If you go to the doctor, if you have a chronic illness
- Your circle of friends, your social status [24]

And there is much more information that can be deduced from the location. That is why the user who had accepted the permit of the location or who may not even know the existence of this permit, has a responsibility that is totally unknown. Nevertheless all this information is completely legal for the companies since they have deduced it from a contract that the user has accepted.

5.5 Conclusion of privacy in the Big Data

Summarizing, the most important question is: Is it possible to have privacy with Big Data? Most experts tell us the answer is no. The reasons why anonymity has been lost have been discussed previously and the points are as follows:

- **The regulations in the United States:** Being one of the countries with the best infrastructure of web servers, it affects us directly since the majority of requests that are made from Europe pass through American soil. Its regulations allow public institutions such as the CIA, the FBI or the NSA to treat information as they see fit, so it is impossible for our privacy not to be violated in most web surfing.
- **Outdated Data Protection Policy:** As explained above for several reasons. All these legal grey areas are used by companies to be able to collect information and to be able to study it later.
- **Poor regulation of anonymous data:** Anonymous data today rarely maintains anonymity and the limits of data processing must be redefined. It is a complex but necessary task to be able to maintain the privacy when surfing the internet.

- **Public information:** And not only do we publish, but also enter here, social networks that also encompasses what other users publish about us, IoT (The Internet of Things) and our configuration that serves, for example, to help The fingerprinting.

All these reasons, among others, are responsible for making it virtually impossible to remain anonymous on the network. At the legal level, companies have legal obligations in data processing among which are the following:

- In the case of Spain, register the files in the Spanish Data Protection Agency (AEPD)
- Prepare and keep updated the Safety Document
- Obtain the consent of those affected
- Take care of data quality
- Regulate access to data by third parties
- Consider international data transfer
- Ensuring secrecy
- Working with anonymised data [25]

On the other hand, as has been seen, all these obligations serve for an average amount of data. However, when we talk about a massive amount of data in which, many data are public or anonymous there are so many legal loopholes that it is impossible to maintain privacy since at the legislative level there is still a long way to go before you can surf the internet In a way that our data is safe.

CAPITOLO 6
LEGAL PROBLEMS

6.1 Approach to the problem

Data Collect. It's done when we get Access to online services, when we download programs or applications, when we hire a flight, when we buy things, and so on, we give our data to the companies that provide them.

What kind of data is delivered? From our name, surname, address, registration number of our vehicle...

These data enter into a database that these private or public companies use for their daily activities. According to data provided by I.B.M. 90% of today's data, worldwide, have been generated in the last 2 years. This company quantifies the data in 2.5 quintillons of bytes per day. One example, about 10,000 credit card transactions occur every second, around the world. Also 340 million tweets are transmitted every day. [26]

To sum up we can say that Big Data is the statistical measurement of the data collected.

In Spain these databases are protected by the Organic Law of Data Protection (LOPD).

Ultimately, we can consider how new technologies in the information society affect our private lives.

The analysis is very complex because there are a multitude of advantages and disadvantages of our technological society, we are going to describe them consequently. Indeed, Big Data is the set of technologies or procedures that allows the treatment of massive amounts of data in order to look for a utility that by itself provides a value. This can reveal patterns of behavior of users or customers in an organization that allows advertising or predict economic trends, as Elena Gil tells us in a more general way. For other authors, it is useful for companies, with improvements in communication networks, competitive advantages in distribution that

allows optimizing the behavior of customers in real time, improving efficiency and business costs and improving business management. [8]
This entails an inherent risk, and is that the data collected is based on the privacy of the people from whom it is obtained. Bear in mind that Big Data can transform into information many unstructured data or aspects that previously could not be quantified or analyzed.
The use of Big Data in the legal field can carry many risks but basically can be summarized in two:

- Automated decision making, based on algorithms, without the supervision of people.
- Privacy in the legal field.

For the first, a possible solution would be for company employees to review this data, programming the algorithms in a way that protects as much as possible the peoples data. It is preferred that this control procedure is performed right after of obtaining the results.
For the second, the solution is legislate a demanding regulation. Let's take a deep look at this privacy issue.
According to Article 3 of Organic Law 15/1999 of 13 December on the Protection of Personal Data and Article 5 of its Regulations, personal data is defined as "any information concerning identified or identifiable natural persons". That means, it is referred to personal data, to that given that whatever its nature (economic, cultural, etc.) allows to identify the person or obtain from him his economic data, address, etc.
Following Elena Gil in his work Big Data, Privacy and data protection data characteristics, would be:

1. That is any information, in a broad sense, considering its content, nature, or type of medium used.
2. Identified or identifiable person, through the combination of processed data, distinguishing it from other people.

The nature of being a fundamental right is that privacy is a right "... inalienable and the fact of prevailing over other non-fundamental rights," as determined by the Constitutional Court Judgment 292/2000 of November 30. In European legislation, this right is set out in Article 16 of the Treaty on the Functioning of the European Union (TFUE) and Article 8 of the Charter of Fundamental Rights of the European Union.

6.2 Legal issues in the Big Data

In all the reasonement exposed above, we can see all the legal problems of Big Data. It is also very easy to see how it doesn’t have an easy solution. There are countless cases and it is

virtually impossible to regulate all the options. This is a review of key points for legal issues related to Big Data:

- **Privacy issues:** It is in this area that more emphasis has been placed. The problems related to this topic are practically infinite and very difficult to control. Main problems are, as mentioned previously, the following:
    - The regulations are not adapted to the new technological environment.
    - The principle of data minimization is not met in practice.
    - The rules rely too much on the' informed consent' of the user.
    - Big Data increases the risk of decision making in our lives.
- **The danger of manipulation:** Kevin Werbach, professor of Legal Studies and Business Ethics at Wharton, said: *"It is easy to use gamification to manipulate. Do such a thing because it is fun when there is actually some objective that Does not necessarily coincide with the interests of the player.Therefore, it is fundamental in ethical gamification that the project is transparent in relation to those objectives.*" The problem is when gamification is used in new technologies and yet it has some kind of unethical objective. With the large amount of data moving through the network, an application that could get user data down and then use it in something out of the blue could have catastrophic consequences.
- **Great legal responsibility:** In the forecoming years, the entire Big Data issue will be regulated, however, as we have seen, there is no consistent legislation or practically any penalty. This means that all the companies that are currently gathering our information are not violating the laws or, if they do, they do in a way that is very difficult to penalize. At the moment a relatively low percentage of all of the captured data is processed but its usefulness is not known tomorrow. For this reason a certain speed is needed in order to put an end to the lack of control of data that is now circulating on the network.

These are just three of the many legal issues Big Data is giving birth to. Undoubtedly, one of the greatest ethical-legal problems occurred in the case of global espionage. This case represented a major legal problem and no repercussions have been taken. It was quoted above and will comment as the case broke all laws that was proposed for its purpose.

6.3 The case of global espionage

6.3.1 Introduction to the case

One of the biggest cases of privacy breach is the revelations about the global surveillance network from 2013 to 2015. This case brought to light how US intelligence agencies along with allied countries had for years Global surveillance.

All this information was revealed thanks to Edward Snowden who copied and leaked thousands of highly secret documents from these companies. It was a set of more than 1.7 million files that contained information about countries such as Australia, Canada or the United Kingdom.

The US government had respected the privacy of all its citizens, however, after a few attacks in the country it was decided to give more importance to national security than to privacy. This caused him to take action in different areas and the intelligence agencies realized that the biggest source of information came from Big Data.

Different methods were used to spy on both mobile phones and computers, such as:

- Introduction of Spyware in popular mobile applications such as the famous game Angry Birds or Google Maps.
- Security breach on iOS and Android.
- BlackBerry encryption violation.
- Spy mails in Hotmail, Outlook or Gmail.

Through these documents, it was also possible to discover how some of the world's leading telecommunications companies collaborated with intelligence agencies on a voluntary basis or in exchange for millions of dollars in addition to access to their servers. This list shows only a few of these companies.

- Microsoft
- Google
- Apple
- Facebook
- Yahoo!
- AOL
- Verizon
- Vodafone
- Global Crossing

All this information was revealed throughout 2013 and the documents were published by the media worldwide.

TOP SECRET//SI/TK//NOFORN

Resource Exhibit No. 1A
National Intelligence Program
Funds by Program
FY 2011 – FY 2017
This Exhibit is SECRET//NOFORN

(Dollars in Thousands)

Program	FY 2011 Actual	FY 2012 Appropriated	FY 2013 Base	FY 2013 OCO	FY 2013 Request	FY 2012 – FY 2013 Change		FY 2013 – FY 2017 Total[1]
						Funds	Percent	
CCP	10,737,163	10,514,035	10,036,851	730,914	10,767,765	253,730	2	50,652,537
CIAP	14,652,379	15,332,901	12,037,708	2,672,317	14,710,025	-622,876	-4	64,567,982
CIARDS	292,000	513,700	514,000	—	514,000	300	—	2,570,000
CMA	2,063,394	1,870,255	1,676,387	—	1,676,387	-193,868	-10	10,274,665
DHS	275,136	307,359	284,332	—	284,332	-23,027	-7	1,462,089
DoD-FCIP	517,720	505,895	456,475	72,485	528,960	23,065	5	2,487,905
DOJ	2,978,329	3,010,795	3,019,958	—	3,019,958	9,163	—	15,596,944
Energy	163,700	186,699	188,619	—	188,619	1,920	1	943,095
GDIP	4,767,009	4,815,583	3,655,662	774,480	4,430,142	-385,441	-8	19,901,677
NGP	5,227,945	5,041,569	4,339,195	539,735	4,878,930	-162,639	-3	22,786,959
NRP	11,401,745	10,411,335	10,268,773	53,150	10,321,923	-89,412	-1	54,842,860
SRP	1,466,792	1,267,751	1,099,820	33,784	1,133,604	-134,147	-11	6,010,922
State	68,773	68,203	72,655	—	72,655	4,452	7	377,056
Treasury	27,422	27,123	27,297	—	27,297	174	1	138,274
NIP Total	**54,639,507**	**53,873,203**	**47,677,732**	**4,876,865**	**52,554,597**	**-1,318,606**	**-2**	**252,612,965**

[1]FY 2013-2017 Total includes the OCO Request for FY 2013 only.

Figure 10 Capture of a filtered document in the case of global espionage [27]

It can be observed in the figure 10 the different programs at the left with really alarming numbers that indicate that it has undoubtedly been the biggesr project of espionage at global level. This figure is an excerpt of the thousands of reports that were leaked by the media.

6.3.2 Data of interest

The effect that had these documents were very diverse in different parts of the world. Different groups demanded NSA and several organizations put pressure into Obama's administration to protect Snowden, who finally had to get exiled.

Obama stated that "It wasn't the citizenship who was being spyed (because) United States didn't have a national spy programme".

In the chart 1 it can be observed diverse surveillance programmes who came to light, and the relationship between programs, partners and business partners involved.

Table 1 Relationship between programs, partners and business parterns involved in the global espionage case [28]

Program	Collaborators	Business Parterns
Prism	- Australia	- Microsoft
	- United Kingdom	
	- Netherlands	
U.S XKeystore	- Germany	
	- Sweden	
United Kingdom	- U.S (NSA)	- British Telecommunications
Tempora		- Interoute
		- Global Crossing
		- Verizon Business
		- Viatel
		- Vodafone Cable
United Kingdom	- U.S (NSA)	
Muscular		
Germany Project 6	- U.S (CIA)	
Stateroom	- Australia	
	- United Kingdom	
	- U.S	
	- Canada	
Lustre	- U.S	
	- France	

It is impossible to determine the exact size of the files, however several officials have approached the following data

- 15,000 archives of Australian intelligence
- 58 000 archives of British intelligence
- 1.7 million files of US intelligence.

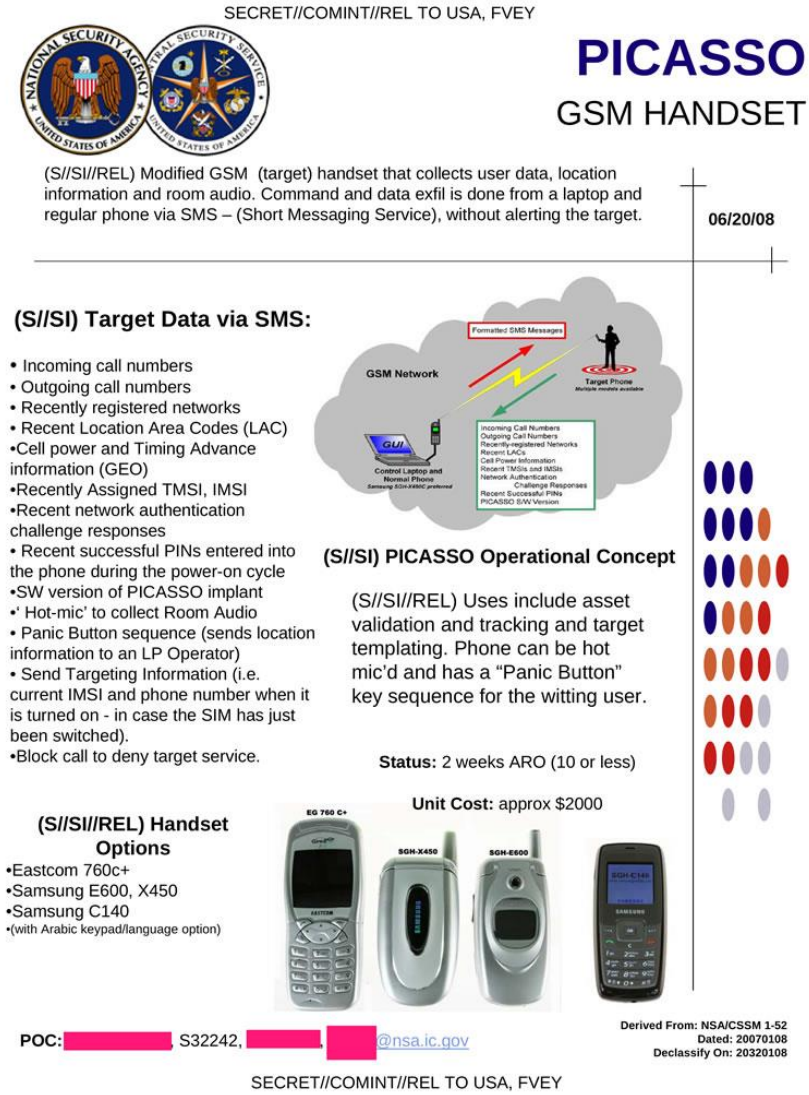


Figure 11 Techniques for global espionage, document leaked by the German magazine.

The TAO division of the NSA developed high-tech devices modified to be able to realize captures of screens and captures of the pen drive to later be transferred through radio waves. In figure 11 we can see a document leaked by the German magazine Der Spiegel, which published a whole catalog of products made to measure for espionage.

6.3.3 Consequences

At first glance it should seem obvious that such espionage should be penalized without any consideration. However, the political leaders of the different countries acted as follows.

- Barack Obama said he had not spied on Americans. The government stated that all of the above programs had been authorized by Congress.

- The UK Foreign Secretary stated that "we are very careful in balancing the privacy of individuals and our duty to safeguard the national and public security of the United Kingdom" [29] and that all programs were necessary because they have saved many lives of citizens.
- The Australian Prime Minister stated that they acted according to the law.
- In Germany, Angela Merkel defended all programs that had leaked and positioned itself in favor of the United States.
- In Sweden, it was also stated that all these programs are necessary for the security of the country.

In addition to the declarations of the main collaborating countries, the United States introduced several new regulations, one of which allows public companies to make any type of request at any time to the servers of private companies that are on US soil. [18]
The ethics of the situation have been discussed on countless occasions. In all the governments it is defended that this practice is still done and that it is totally necessary for the security of the citizens. It is clear that there is a violation of privacy on a global scale and that it is legally impossible to impose this continuous surveillance. Most citizens are still unaware of this issue and the other side simply accepts it, if the government side consents, the legal part is on a lower step.

CAPITOLO 7
LAWS ALREADY DEFINED

7.1 Regulations applicable in Spain.

In relation to data protection by companies, bodies or citizens in Spain the following regulations are applicable:

- The Spanish Constitution.
- The Organic Law 15/1999 of 13 December on the Protection of Personal Data.
- Law 62/2003 of 30 December, on fiscal, administrative and social measures.
- Royal Decree 1720/2007 of 21 December, approving the Regulation implementing Organic Law 15/1999 of 13 December, on the protection of personal data.
- Royal Decree 428/1993 of 26 March approving the Statute of the Spanish Data Protection Agency
- Other regulations (Resolutions).

The Spanish Constitution, as the supreme rule of the state, regulates in its Chapter II, Rights and Freedoms, the content of those rights of maximum protection by the legal order, especially in the private sphere. The article states that: [30]
"*1. It guarantees the right to honor, to personal and family privacy and to the image itself.* "

The legislative development of this article occurs with Organic Law 15/1999 of 13 December on the Protection of Personal Data [31] and its regulatory development (Royal Decree 1720/2007 of 21 December). [32]

These generic concepts of the right to honor and personal privacy are delimited by articles 6 and 7 of that Law 15/1999 establishing a more concrete protection, such as respect for personal belief, ideology, religion, race, Health and sexual life.

Despite the doubt among jurists as to whether personal data should be protected or not, after an express written consent, there is finally a definitive consensus of positive agreement, in the sense that these constitutional principles must be defended.

This is because after a study by the State Administration, almost half of Spanish companies violate the Organic Law on Data Protection (LOPD), recognize the illegal treatment and their transfer of their files to third parties under a low level of security. [33]

The data protection law begins with the definitions of what is considered a file (organized set of personal data, regardless of the form of its creation, storage, organization and access), data treatment (operations and technical procedures of Whether or not they allow the collection, recording, preservation, elaboration, modification and transfer of data) and of the different authors involved in this elaboration (interested, in charge of the treatment, etc.).

"*Personal data*" is defined as any information relating to identified or identifiable natural persons. Companies in the development of their own economic activity with their clients or suppliers use personal files that are affected by the set of regulations mentioned above.

The principle that governs the regulations is that the data they treat in their companies are not owned by the same, but by their owners, so for the correct application of the law we must take into account all phases of computer processing a) . Of data, b) Processing of data, c) Use, assignment or communication to third parties). For this reason, the obligation of the company is not reduced to a certain moment, but to all the time that the company carries out its activity. [34]

Also the functions of the people involved in the previous process of collection, treatment and use or transfer of data is defined in the law. The person in charge of the file is the person who is responsible for the knowledge and compliance with the regulations, to know the security measures, awareness and training of users and ultimately is responsible for the sanctions for breach of the law.

However, we understand that the most important of this law comes in Articles 5 and 6 since in the first of them it contemplates the rights of information in the collection of data of the interested parties in which they must have full knowledge of the existence of the file. The processing of the data to be performed and the consequences of obtaining or refusing to supply them.

This regulation affects all professionals, companies, public and private organizations that store or manipulate personal data, regardless of the hardware used.

Article 6 of this law already establishes that the consent of the person concerned is mandatory ("*The processing of personal data will require the unequivocal consent of the person concerned, unless otherwise provided by law*.")

Such consent may be revoked when there is Justified cause.

An exemption to this consent is established by the receipt of data by the Public Administrations for the exercise of their own administrative functions.

Once the file has been created, the procedure established by the law is that the person responsible for the file must write a security document with the measures adopted, submitting the forms to the Data Protection Agency for registration. From now on, the Agency will monitor and control the databases created.

Title III of said Law 15/1999 establishes the rights of persons in relation to databases to: [35]

- Challenges of valuations, where it is allowed to challenge administrative data whose assessment of their behavior is based on the processing of personal data.
- Right of access of the interested party, to obtain free information of their personal data undergoing treatment.
- Right of rectification and cancellation within ten days when your data is inaccurate or incomplete.
- Opposition, access, rectification or cancellation procedure that will be established by regulation, without requiring any consideration for their right of opposition.
- Protection of the rights that can be claimed before the Data Protection Agency and the resolution of this, can be appealed in contentious-administrative procedure.
- Right to compensation as a result of non-compliance with Organic Law 15/1999 of 13 December.

During Chapter IV of the law regulates in Chapter I the creation of a publicly owned file which, in Chapter II, establishes the creation of privately owned files. It is important for the lucrative purpose of the activity, the provisions of article 30, which regulates treatments for advertising and commercial prospecting purposes. In this, it is specifically determined that only advertising or commercial activity may be carried out, when the names and addresses have been provided by the interested parties or obtained with their consent. If they have been obtained from public data, the interested party will be informed of the origin of the data and of the identity of the person responsible for the treatment, as well as of the rights that assist him.

Title VI regulates the control body of this activity that falls under the Data Protection Agency, public law entity, with its own personality and full public and private capacity that acts with full independence of public Administrations in the exercise of its functions. It is governed by this Law and its own Statute.

The organization chart is formed by its Director and an Advisory Board and its functions are basically to ensure compliance with the legislation on data protection and control its application in particular the rights to information, access, rectification, opposition and

cancellation of data as well Such as responding to the requests and claims of the people affected.

The law ends with a catalog of infringements and sanctions (Title VII) that are classified as minor, serious and very serious that can cause sanctions of 900 to 40,000 euros the first, from 40,001 to 300,000 euros the second and from 300,001 to 600,000 euros Respectively.

The Regulation of Organic Law 15/1999 is complemented by the provision Royal Decree 1720/2007 of 21 December, which approves the Regulation of development of Organic Law 15/1999 of 13 December, on the protection of personal data personal. This rule has the function of expanding and developing those aspects that the Organic Law does not contemplate. [32]

The purpose of this regulation is to establish the technical and organizational measures necessary to guarantee the safety of automated and non-automated files, treatment centers, premises, equipment, systems, programs and persons involved in the treatment Of the personal data. [36][37]

7.2 Regulations applicable in European Union.

The current legislation on data protection is Directive 95/46 / EC, which was approved in 1995 for two purposes. [38] On the one hand, to defend the fundamental right to data protection and on the other to guarantee the free movement of such data, at a time when the free movement of people, goods and capital was a crucial issue. This Directive constitutes the reference text at European level on the protection of personal data, creating a regulatory framework designed to strike a balance between a high level of protection of the privacy of persons and the free movement of personal data.

This Directive considers that data falls under the lawful category only if the following requirements are met: [39]

- That the person concerned has unequivocally given his consent, or
- That the treatment is necessary for the fulfillment of a contract in which the interested party is part; or
- That the treatment is necessary for compliance with a legal obligation of the controller; or
- That the treatment is necessary to protect the vital interests of the person whose data is involved; or
- That the treatment is necessary for the fulfillment of a mission of public interest or inherent in the exercise of the public power conferred on the controller or a third party.

- That the treatment is necessary for the purposes of the interest of the controller or of a third party, provided that such interests do not prevail the interests of the interested party in the area of rights and freedoms that require protection.

The principles of quality are as follows:

- That personal data will be treated in a fair and lawful manner and collected for specific, explicit and legitimate purposes.
- That special categories of treatment exist, and those that reveal racial or ethnic origin, political opinions, religious or philosophical beliefs and membership in trade unions or that reveal data relating to health or sexuality are prohibited.
- The interested party has the following rights:
- The right to obtain information: the controller must provide the following information (identity of the controller, purposes of processing, receiver of the data, etc.)
- The right of access of the data subject.
- The right to object to the processing of data, for legitimate reasons.

In the same way provides that national laws must provide a judicial review in cases where the controller does not respect the rights of the parties concerned and obtain compensation for the harm suffered. As we have already exposed, in Spanish legislation this is fulfilled, given that after the corresponding administrative complaint against the State Data Protection Agency (APD), the corresponding contentious-administrative appeal may be lodged with the Contentious Jurisdiction. This causes that a judicial body i sable to supervise the whole procedure of data processing and declare it, adjusted or not to law.

This law was complemented by another, Framework Decision 2008/977 / JAI as protection of personal data in police and judicial cooperation.

We should also analyze Directive 2002/58 / EC of the European Parliament and of the Council of 12 July 2002 on the processing of personal data and the protection of privacy in the electronic communications sector (Directive on the Privacy and electronic communications - Official Journal L 201 of 31.07.2002).

This Directive was enacted in conjunction with other laws, with the aim of protecting electronic communications, including provisions for the conservation ofConnection data by the Member States for the purpose of police surveillance (data retention), sending unsolicited electronic messages, using chivaos (cookies) and the inclusion of personal data in public directories.

In case of a breach of personal data that compromises safety, providers must inform the competent national authority and in certain cases subscribers or individuals about the breach. Regulation (EU) 611/2013 provides for the technical measures to be taken to comply with these obligations, which include, inter alia, reporting the incident within 24 hours of

detention, reporting on the measures taken and the nature of The files or data affected (financial, economic, etc.).

Given the evolution that the information society and digital is changing our way of coexistence, and given the massive increase of data of the companies, it was necessary to generate confidence in the online environments.

In this new need the European Parliament and the Council have adopted Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of individuals regarding to the processing of personal data already The free movement of such data and repealing Directive 95/46 / EC (General Data Protection Regulation - RGPD) with the aim of unifying all the regulations of Member States in this area. Published on May 4, 2016 in the Official Gazette of the European Union entered into force on May 25, 2016, although it will not be mandatory until two years have passed since that date.

These regulations are governed by the following principles:

Principle of responsibility, implementing mechanisms to check all measures of the process of data collection, manipulation and use. It is up to the companies or organizations to accredit the legality requirements.

Principle of protection of data by default and from the design. It must be guaranteed from the birth or beginning of the creation of a company, product, service or activity that the treatment of the data has been correct.

Principle of transparency. The privacy policies have to be simpler and simpler as well as the mechanisms of correction, modification and exclusion of the existing data.

New aspects that this regulation implements are the following:

a) In some cases, it will be necessary to designate a Data Protection Officer (DPO), internal or external, that will help companies to comply with the standard.
b) Certain impact assessments on privacy, on certain personal data, should be carried out in order to eliminate the risks.
c) Multinational enterprises shall have only one national interlocutory authority for the data processing. This is called a one-stop shop.
d) Security problems must be reported to the supervisory authorities within 72 hours.
e) The catalog of specially protected data is expanded, which now includes genetic and biometric as well as criminal convictions.
f) The choice of a data manager becomes more strict, requiring that it be suitable for compliance with the law.
g) Additional mechanisms and guarantees for the international transfer of data outside the European Union are established.
h) They increase penalties that can reach 20 million euros or 4% of global annual turnover.

One criticism that we must make to these regulations is that, there are many generic elements that are too ambiguous in which there is no concreteness and others that are pending legislative development.

Notwithstanding the foregoing, indicate that the provisions of this Regulation are directly applicable in each of the Member States, without the need for new laws, which obliges private companies and Public Administrations.

This General Data Protection Regulation does not automatically repeal the LOPD and its Regulations. It simply shifts these to what extent they are incompatible with it.

7.3 Regulation of different countries

Few countries have regulated certain aspects of Big Data. Being such a novel subject should be raised a series of new laws that is as much as possible. This is a review by the different countries both in Europe and abroad with its laws relevant to the regulation of Big Data:

- **Australia:** The Freedom of Information Act 1982, the Archives Act 1983, the Telecommunications Act 1997, the Electronic Transactions Act 1999, the Data-matching Program (Assistance and Tax) Act 1990, the Privacy Act 1988, the Privacy Amendment Protection) Act 2012, the Privacy Regulation 2013.
- **Brazil:** An amendment to the legislation on data protection is currently being developed. The government has released a draft bill for this law, entitled: "On the processing of personal data to protect the personality and dignity of natural persons".
- **China:** China does not have overarching privacy legislation such as is present in many European countries. At the end of 2012, the Chinese parliament drafted a resolution consisting of 12 articles and regulating privacy and data protection: the 'Decision of the Standing Committee of the National People's Congress to Strengthen the Protection of Internet Data'.
- **France:** The 'Loi Informatique et libertés'1978, which has been amended several times since its introduction.
- **Germany:** The central data protection legislation in Germany is the Bundesdatenschutzgesetz, originally dating from 1990.
- **Japan:** The Act on the Protection of Personal Information from 2003, in 2013 was amended to this law, inter alia because of Big Data.
- **USA:** The United States does not have an overarching law for the regulation of privacy, and certainly not for the specific regulation of Big Data. Besides the constitutional protection, the United States has a system of sector-specific regulation of privacy risks. The Consumer Bill of Privacy Rights was introduced in 2012. This is

not legislation in the sense of being enforceable, but more of a guideline for the business sector. [40]

7.4 Actual cases of companies using the Big Data irregularly

The cases of exploitation of companies by the processing of information is inexhaustible. We have already indicated previously that in a study of the Spanish Administrations, the irregular or illegal use of corporations and companies arrive almost to the half. We are going to present four real cases that have appeared in the media and that allow us to observe the possibilities of the procedure.

7.4.1 CASE 1: Father who finds out that his daughter is pregnant before she tells him

In Minneapolis, in 2010, a parent demanded explanations from the head of a facility about why his daughter, who was still in high school, had received maternity clothing and discounts for baby clothing, understanding that the trade was encouraging pregnancies. Later the responsible one went to excuse itself with the father, indicating him that he had just learned that his daughter was pregnant.

This prediction was designed by a scientist who created an algorithm that found that pregnant women purchased up to 25 pregnancy-indicating products, including an unscented cream at the beginning of the second trimester of pregnancy, which could also be predict child's birth time. [41]

7.4.2 CASE 2: The Nordstrom chain studied the intimacies of its customers

Taking advantage of the free Internet access of its customers and the assignment of a Smartphone to each user, using software called Elucide and the local surveillance cameras, analyzed data about their sex, places they frequented, the time they stayed in each place, etc. When the users were aware of the surveillance to which the clients were subject and by the complaints of these, the company had to finish the study after eight months, although Nordstrom indicated that its privacy policy did not allow to advertise this data. Actually, this same system is used by many companies like Amazon that analyzes navigation websites using cookies to identify age, sex, social status, residence, etc. [42]

7.4.3 CASE 3: Older people were left without health insurance

Walter and Paula who lived in the US and bought large quantities of drugs in stores like Wal-Mart and Randalls, especially antidepressants and blood pressure. When they wanted to insure themselves in health insurance, they were denied inscription for high prescription drugs. The case came out in a Bloomberg report in 2 008. [42]

CAPITOLO 8
BIG DATA, A REGULATION IN DEPTH

8.1 Data types

To better understand how data processing works, you should see everything Big Data encompasses. In order to deal with the legal problem, it is necessary that each case should be studied in particular so as to be able to propose better solutions in the future.
First, we will start by differentiating the types of data that this technology can study. This is a complex task because of the large amount of information it collects. IBM has classified the Big Data data types into 5.

- **Web and Social Media:** When it comes to information coming from social networks. An example of this type would be for example Facebook or Twitter.
- **Machine-to-Machine:** Also known as M2M, this type of data talks about technologies that allow it to connect to other devices. Usually we find sensors, cameras, etc. An example of this type would be sensors of temperature, speed, meteorological variables among others.
- **Big Transaction Data:** This type of data includes all the information that is collected in the transactions. For example all the information that is collected in the centers of attention telephone, of finance, in a banking transaction, attention to the clients, etc.
- **Biometrics:** Everything that includes biometric information. It is mainly used in intelligence and security centers. This technology is responsible for fingerprinting, retinal scanning or facial recognition.

- **Human Generated:** This last type of data collects all the missing information. That is all the data created by people collected from various sources such as emails, medical results, fines, electronic documents, etc.

At the legal level none of these 5 types of information is regulated as such. Laws should make a special focus on Web and Social Media, Big Transaction Data and Human Generated as they are the most dangerous and relevant types of information regarding privacy and data protection.

But it is not enough just to regulate and standardize the data types that Big Data comprises. The real legal problem is in the treatment of these. In this way, in addition to regulating the types, the same should be done with the analysis of this information. The different ways of analyzing the types of data previously seen will then be studied.

8.2 Analysis of information

This information analysis is treated in different specialized software. One of these programs is the Sisense that serves as a solution for the treatment of massive data. We can see it in figure 12 and on their own website we explain the following:

"*Sisense Cloud gives you the analytics and insights you need, without worrying about hardware or IT. Now you can focus on what really matters: giving your business users, analysts and customers the data they want to see, analyze and visualize, when and where They need it.*" [43]

Figure 12 Capture of the sisense program used for Big Data applications [44]

These programs, since they have been developed relatively recently, continue to improve every day with more functionalities for the study of the data. There are many programs for each of the data types, but the information can be treated in several ways:

8.2.1 Machine Learning

Analyzing information from Big Data is a really complex process. In a conventional database specialized algorithms are created in each one of the types of data that are able to take advantage of the hosted data. The problem arises when these data are created thousands per second, then comes the need to construct other models of data processing.

Kovahi and Provost in 1998 defined this type of data processing as the discipline that deals with the construction and study of algorithms that can learn from data. That is, these algorithms try to construct models based on massive data. An example of this treatment would be the voice control of Smartphones.

As the figure 13 shows, the first step is to handle a large amount of data so that an algorithm is able to create an organized data map. Later on new information can predict or make decisions.

An example of this technique, as discussed earlier in the treatment of voice, would follow these same steps. First of all you must possess a large number of syllables or phonemes recorded by many types of people from different locations. Through an algorithm the system learns the idea of each of these syllables or sounds. Subsequently, a foreign user records a sound and the system that has a large classified database and has learned all the possible sounds is able to know which of them has said.

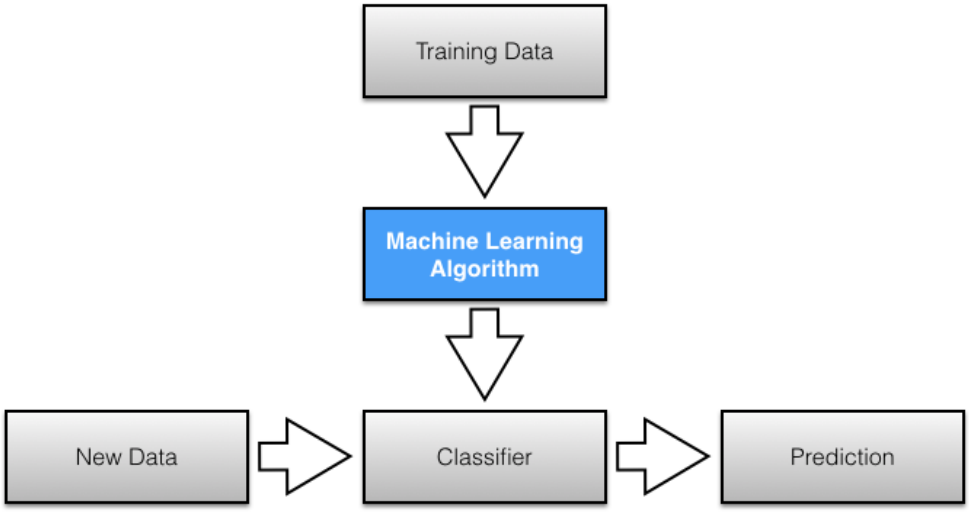


Figure 13 Scheme of the operation of machine learning technology [45]

It is also widely used in data mining, artificial intelligence or statistics. There are different types of automatic learning such as:

- **Supervised learning:** The algorithm produces a function that matches the inputs and outputs of the system.
- **Unsupervised learning:** The process for constructing models is done on a set of examples formed by inputs to the system. It also recognizes patterns for new entries.
- **Semi-supervised learning:** It is a combination between the two types of algorithms above, which takes into account the marked and unmarked data.
- **Reinforcement Learning:** This system learns from sensors by observing the terrain where it is located. Information is the feedback you get from the outside world. This algorithm is able to learn how according to the actions it does it obtains different answers, an example could be an expert robot.
- **Transduction:** This type is similar to supervised learning. However the data here are declassified. This algorithm is designed to predict the categories of future examples based on the input examples.
- **Multi-task learning:** This algorithm performs the simultaneous resolution of different tasks. An example would be the imputation of incomplete data.

As in data types, none of these treatments is regulated. The algorithms of automatic learning are found in our daily life and facilitate us work in different fields. However, used in an unethical way could be very dangerous. That is why it is necessary to study and regulate all these algorithms as soon as possible.

8.2.2 Big Data and Statistics

Another way to analyze Big Data data is by using different statistical techniques. There are many forms that can be summarized in:

- **Classification:** Assign a class or label to a particular object, or individual. With this system for example you can label whether a product is good or bad.
- **Regression:** It is a generalization in the part of the regression. The output of this study is usually a number or a vector of numbers
- **Clustering:** This technique is used to organize objects or individuals into groups. These groups may be hierarchical or not. [46]

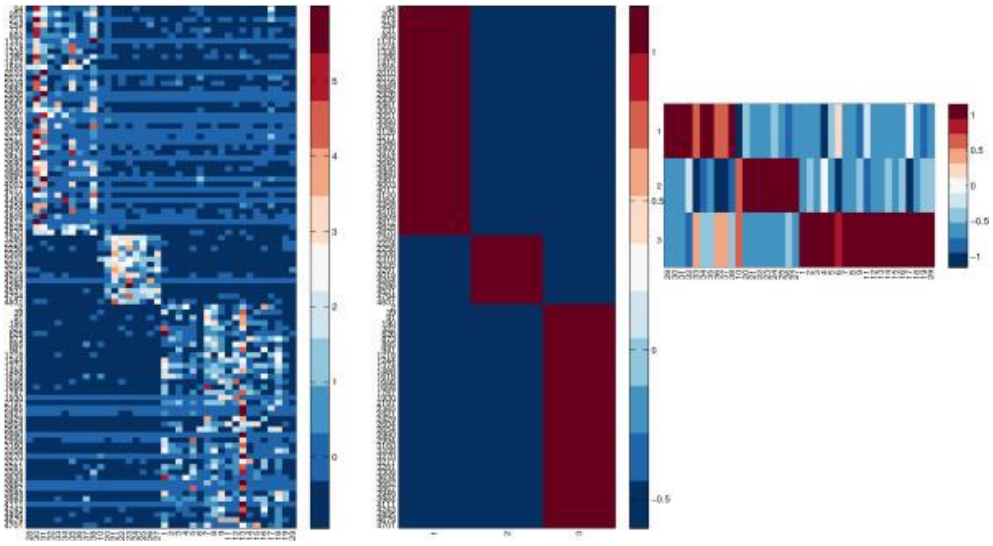


Figure 14 Heatamp biclustering technique used in the big data static study [47]

In the example of the figure 14 we find a color map that allows the interpretation of data and variables. This is effective since from a visual you can observe coincidences and patterns. This is the ''Heatamp biclustering'' technique.

In these statistical studies there is a more appropriate regulation. The data must be anonymous whenever it collects personal information. In addition the European Union has legislation on the statistics that are grouped in the following areas:

- General, financial and institutional matters
- Customs union and free movement of goods
- Farming
- Fishing
- Free movement of workers and social policy
- Right of establishment and freedom to provide services
- Transport policy
- Competition policy
- Taxation
- Economic and monetary policy and free movement of capital
- External relationships
- Energy
- Industrial policy and internal market
- Regional policy and coordination of structural instruments
- Protection of the environment, consumer and health
- Science, information, education and culture
- Company law

- Common Foreign and Security Policy
- Area of freedom, security and justice
- Citizens' Europe [48]

Each one of these chapters contains a regulation on the statistical studies of the subject that treats. This applies to statistical studies either by conventional methods or using Big Data tools.
However statistical studies with massive amounts of data can also be used in an unethical way because statistical studies could find information that was previously anonymous for example.
A regulation must be created that covers the new technologies of Big Data also in statistical studies because, beyond the subject matter they may be used for many purposes that may violate the privacy of the participants or the users who are Find in the database without even having their consent with techniques such as the fingerprint or cookies without cookies.

8.3 Internet of Things

Another issue to deal with in Big Data technology is the Internet of Things. Data protection authorities define this phenomenon as: "Global infrastructure in which sensors capable of interacting with each other and with other systems are incorporated into everyday devices in a way that collects, processes, stores and transfers data using capacities Network interconnection".
It also refers to 3 types of devices:

- **Devices available or wereables:** Accessories in clothing that include sensors.
- **Quantifying devices of the activity of the person:** Devices designed to store information about the habits of the people.
- **Devices on home automation:** Devices that can remotely control an object, have sensors and can connect to the internet.

In the idea of the Internet of the things more ideas exist besides these three, nevertheless they are the points to which opinion 8/2014 of the protection workgroup of information WP223 refers.
This opinion is in charge of creating rules in the Internet of Things (IoT) and can be summarized as:

- Directive 95/46 / EC on the protection of individuals with regard to the processing of personal data and on the free movement of such data.

- Directive 2002/58 / EC on the processing of personal data and the protection of privacy in the electronic communications sector. Amended by Directive 2009/136 / EC on universal service and users' rights relating to electronic communications networks and services.
- Legislation specific to the member state, in this case the organic law of personal data protection (LOPD), in general. [49]

In addition a set of criteria specifies for regular power like that the correct use of this technology:

- Evaluation of the impact on the privacy, with character before the launch of any device or application.
- The responsible companies will have to erasethe global information obtained of the devices, once it has obtained the information that they need for hisprocessing.
- Respect of the principle of privacy for design and fault.
- The holders of the information and the users must be able to exercise hisrights arch (information, access, rectification, cancellation and opposition).
- The methods by what information is supplied and the assent is requested, as well as the policies of the site, must be written in a clear and simple way.
- The devices and applications. [49]

As can be seen, this is a subject that is still developing. In cases where these regulations can not cover, the same law as the one related to the protection and privacy of personal data will apply.
The Internet of Things is a subject that is regulated in part. Obviously it is impossible to cover everything that encompasses this theme when thousands of new products related to this subject are released every year. However this regulation works quite well.

8.4 Physical Storage

In addition to all types of data and their treatment, it is worth mentioning the very storage of these. These massive amounts of data must be stored in very large databases and also need security so that they can not be filtered in any way.
These data come from very different sources, such as from hospitals, meteorological information, social networks, etc. Some of this information comes from sensors that can or can not be treated in real time. Useful information should be obtained from these data so they should be treated very carefully.

On the other hand, all this information must be housed in physical spaces to which people can access. This hardware must be protected in a number of ways. Natural disasters, deterioration of materials and endless events that can put an end to information must be taken into account.

In addition to the conservation of information, what should be regulated in a very strict way is the protection of these physical systems. Here are some of the potential problems that may occur in large database structures:

- **Eavesdropping:** The process by which information is intercepted. It can be done in various ways and in different media such as wireless or Ethernet. One of the most adopted solutions is to encrypt all the information with the use of different algorithms.
- **Non-electronic supports:** It must also be protected in the same way. In many cases the results, a part of the data and any type of information of the database is reflected in paper among others. This information should be handled with the same care as electronics.
- **Backups:** The issue of backups is important in all databases. These copies are usually kept in those that must be protected physically since, although the information is not updated, it still stores personal data. [50]

There are three main aspects as security measures:

- **Confidentiality:** Only authorized persons can know the data or the information stored.
- **Integrity:** Only authorized persons can modify or delete data. In addition, any movement must be recorded.
- **Availability:** If authorized persons can have the information at the time indicated.

In addition to the security mechanisms in the database:

- Access control
- Flow control
- Inference control
- Encrypted [51]

CAPITOLO 9
THE FUTURE OF THE BIG DATE

The speed with which Big Data technology grows reaches unsuspected limits. That is why it is important to note that if today there is no regulation that can solve all legal problems, tomorrow will be even more difficult. This means that real and clear rules are needed as soon as possible. The legal part of all countries should grow at the same rate as Big Data if we do not want our privacy to be in danger.

Figure 15 Growth of information saved until 2015 and forecast for 2020 [52]

Here are 10 predictions that experts are sure will happen in the future in relation to Big Data:

1. The growth of information is unstoppable: Every year humanity generates an enormous amount of information. As you can see in the figure 15, it has grown exponentially since the appearance of Big Data. In the future this can be a problem due to the large amount of data stored.

2. Tools will improve. This is obvious, because Big Data is becoming more and more important in most industries. The big companies will bet on this new technology with the objective of obtaining a dominant position in the market.
3. The simplest analytics: Data manipulation is expected to be so easy in the future that any department can do so without relying on engineers and analysts.
4. Analytics will be integrated into basic business software: Data processing tools will be incorporated into the core business tool packages.
5. There will be more failures: This is one of the drawbacks of the growth of this technology. When you are in the hands of virtually every business, people without proper preparation run the risk of producing erroneous conclusions from complex studies.
6. Computers will not only collect data: They will also be able to make connections between study data that a user would be unable to. This is due to the continued growth of Artificial Intelligence.
7. It will affect privacy: As we have seen, a regulation should be created as soon as possible. Otherwise experts say that in the future this situation will be much worse.
8. Demand for data communication experts will grow: Obviously whenever technology grows, demand for experts will grow. The analysis team will be an obligatory department for medium and large companies.
9. Algorithms will enter the market: Companies will be able to buy just one algorithm and not all study or analysis software like now.
10. Completely new areas will open: With the growth and the appearance of new computer technologies will appear studies in the Big Data since all the statistical information from the users is of great value in the companies. [53]

In addition to these predictions, there is a lot of data about how technology has changed our lives today. Studies show how in the coming years, statistics on the use of Big Data will increase dramatically. These are some of the data that are believed to come until 2023.

- 2018: Half of the users will interact with cognitive computing technologies.
- 2019: It is estimated that as in 2015, 38% of the population was connected to the internet, to 57%.
- 2020: Decision management platforms will grow to 60%. This can cause a legal problem if these platforms start from the wrong premises so a regulation would be needed.
- 2021: Predictive analysis will be a key part of any business.
- 2022: The price of sensors will fall to almost half of what was in 2015. In addition, the machines will process all the information and serve products to the users at the right time and place.

- 2023: 50% of smart cities will be in Europe and North America. [54]

These are some of the data that are foreseen for the future of Big Data. They are data that demonstrate a great advance of the humanity however, the legal part is less hopeful. If users' privacy is now violated secretly and passively, in a future where everyone is connected with all the sensors that will come out new to the market can be devastating for the privacy of the users.

In this way, for a better future where Big Data is going to find in the normal way most of the companies, certain legal subjects for a radical change that is looming in the next years should be approached.

One of them would be to redefine the concept of privacy. One of the keys for users to keep their privacy is to redefine what privacy is. The legal definition of this concept served before the revolution of the Big Data however, at this moment it is necessary that it is coupled to the new technologies.

However on many occasions the study with the Big Data of personal data can be very useful. In medicine, it is anticipated that in the near future studies can be carried out that minimize the risks of contracting diseases to patients with prediction, prevention and participation.

The **prediction** with the Big Data is used today in a small part of hospitals around the world. Only those with the right technology are able to make accurate predictions. Sensors are used in patients and measuring behavior patterns are able to control when something is going to happen before it happens.

An example of prediction is used at the Toronto Children's Hospital, where sensors exist to process the heartbeat of newborns. With the help of the MBM supercomputer, these beats were analyzed in real time and found patterns that revealed signs of infection 24 hours before they appeared. [55]

Prevention is able to inform beforehand of heart attacks, diabetic comas or even falls of the elderly while they sleep, etc. Obviously in order to find all these patterns it is necessary to have the appropriate sensors and a computer that analyzes the information in real time.

The **participation** depends on the user and Big Data has much to contribute. All the 'wereable' devices that came out on the market and the applications that monitor our data both in sports and when, are helping a great participation by the users.

In the future, all that is now happening in a small part of the population will be common to most of the population. This will bring great benefits in the field of medicine.

Another question we must answer for the future is who owns the data being analyzed. For example, in a future where cars are connected to the Internet with a large number of sensors and all this information is analyzed, this information may belong to drivers, owners or manufacturers of the vehicle.

In addition to all this, in the future it is very likely that techniques will emerge to deanonymize or reestablish the identity of the data. One of these techniques was already used in early 2013 where several computer security experts were able to identify surnames corresponding to 50 supposedly anonymous human genomes. These techniques are very rare and require experts in the subjects, however in the future it is very likely that the future will be used more than it should. [55]

If we talk about information collected in the future of Big Data we find really large numbers. The information has been kept for some years. There are companies dedicated to storing customer data like the fingerprint without their prior permission. In addition to large companies such as Facebook in which customer data are the business model.

"*73% of banks in Spain already have 'big data' initiatives, according to figures from KPMG and Funcas*" [56]

In the field of banking, data is extremely important because it is one of the main tools to facilitate the work. The volume of data of any banking system is a challenge both to store them and to analyze them, in addition to being a critical system which can not have errors.

Beracoechea said: "*Banking data is extraordinarily rich and varied, but the real challenge is how to tame that diversity of information in ways that add value to our customers. This means having the right talent, the necessary technology platforms, and An organizational design that facilitates the interactions between the different actors*" [56]

The growth of these data is surpassing any forecast and it is believed that in the year 2025, the information created each year will exceed 180 Zettabytes.[57] Few people can understand the information that can be stored in this figure. In the figure 16 we can see which part of the data scale is located.

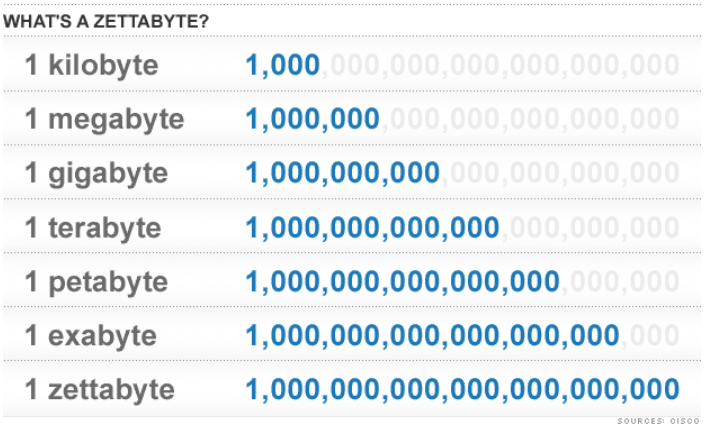


Figure 16 Visual schema of the information contained in a zettabyte. [58]

This amount of data should not only be stored but also analyzed. All these data come from users and contain from anonymous data to personal information. Information that the user must have given their consent for it to be stored and analyzed.

In addition, what they call the fourth industrial revolution will destroy some 7 million jobs, most of them automatic actions, and create 2.1 million jobs. Most of these posts related to the Big Data among which are Computer scientists, engineers and mathematicians most in demand for these positions due to the ability to analysis.

Top 10 Most In-Demand Skills

Skill	No. of Big Data Jobs Mentioning this Skill Set	% Growth In Demand For This Skill Set Over The Previous Year
Big Data	112,469	118%
Java	35,700	106%
Hadoop	31,274	118%
Python	31,100	231%
Structured query language	28,037	76%
Software development	27,990	128%
VMware	27,249	1269%
Application development	27,202	396%
Data warehousing	26,418	272%
Open source technology	23,666	387%

Figure 17 Top 10 Most in-demand skills according to the study '6 Reasons Why Big Data Career Is A Smart Choice' [59]

As you can see in the figure 17, the most demanded position will be Big Data. However all other skills shown are also related to Big Data. This data in the future will skyrocket even more when the percentage of companies that use this technology grow as much as predicted. Both the high growth in demand and the number of positions in the Big Data that will need the skills exposed in the image makes it possible to see the importance of the matter. We are facing a very important change where Big Data is going to be one of the great protagonists.
All these data show us how clearly, this Big Data technology is not a fad and in the future will go more. It has had an exponential growth in the last years and although it does not continue with the same intensity we must begin to face all the legal problems that this constitutes.
The society is in a problem of difficult solution in which today its privacy is affected by the new technologies to analyze large amounts of data. This problem will increase in the future, giving rise to algorithms and data that would never have been thought to exist. That is why we need a legal solution now. In addition, the entire legal sector should be informed of the new risks and dangers of this technology if we do not want any privacy in the future.

CAPITOLO 10
POSSIBLE SOLUTIONS FOR BIG DATA AND THE FUTURE

A new scenario, independent of the analyzed legal, to find solutions is to implement an ethics that protects privacy and a legal use of our data.

Indeed, in many of the decisions that concern us, they are performed by algorithms and mathematical models that sometimes make their own business decisions. It creates a problem here to identify the "responsible author", since on many occasions, we will find ourselves before a machine. [60]

Rosa Colmenarejo Fernández, in his work Ethics and Big Data, and following the lines also of Beck V in his work Risk Society, states that "The social structures that begin to manifest themselves from the management, availability and use of management, availability And the use of the Big Data bring ethical conflicts in many cases inherited from business ethics, although now it is outdated by the circumstances in which they operate the generation and massive management of data". She considers that the problems of data storage, management and use may affect the privacy, ownership, identity, trust or reputation of the interested parties, a differentiated ethics of business ethics should be applied. It could be called cybernetics.

It is very interesting what Bauman (2006) poses as the "technological cover-up of the moral self"[61] which understands that the fragmentation of life, of the ego, into a group of facets, which requires different application techniques Conclude in a "moral innocence", since the subject never acts as a total person, is what Beck called the "society of risk." [62]

For Rosa Colmenarejo, understands that what must be done is the good, the just and the right, within the application of some professional principles, which divides into:

Principle of beneficence, where the "good" identified is maximization.

Principle of autonomy, which analyzes the instrumentalization of people in the fulfillment of this objective.

Principle of justice, which prioritizes actions with a criterion of social justice.

Principle of non-maleficence, in which, if the above principles can not be applied, actions must be performed that do not harm the individual.

I believe that a solution to the possible violations of the use of Big Data is a joint application of business ethics and on the other of the strict compliance with legal regulations. This must be accompanied by a set of high penalties that reasonably prevent, infringement of fundamental rights and strict control by the Administrations.

CAPITOLO 11
POSSIBLE SOLUTIONS IN THE LEGAL ASPECT

In Spain data protection is also included within the protection of the fundamental rights of article 18 of the Spanish Constitution as a right to honor and personal and family privacy.
That at European level, data protection is nothing less than a fundamental right, enshrined in Article 16 of the Treaty on the Functioning of the European Union and Article 8 of the Charter of Fundamental Rights of the European Union.
That at European level, data protection is nothing less than a fundamental right, enshrined in Article 16 of the Treaty on the Functioning of the European Union and Article 8 of the Charter of Fundamental Rights of the European Union.
For this reason, the Spanish legislation on protection will apply in the following cases:

a) When the processing of the data is carried out in Spanish territory within the framework of the activities of an establishment owned by the data controller.
b) When the data controller is not established within Spanish territory, but Spanish law is applicable according to the rules of Public International Law.
c) When the data controller is not established in any country of the European Union, but in the processing of the data uses means located in Spanish territory, unless such means are used only for transit purposes.

Excluded from this legislation are:

a) Matters that have specific regulations such as the public statistical function or data from images and sounds obtained by the Security Forces and Bodies through video cameras.

b) Nor shall it apply to files of natural persons in the exercise of their exclusively personal or domestic activities.

The big problem of Big data is that there is no specific regulation, but it applies, Law 15/1999 of data protection that has become obsolete in the analysis of large data using this technology. The impact of this technology on issues such as civil liability, competition law, intellectual property rights, etc., will be an important milestone in the search for that weak balance between personal privacy and data protection.

It is necessary to consider that to the exposed thing previous, are united other important legal problems as: [8]

1. That the legislation is not adapted to the new technological environment, so changing and that the legislator can not even predict.
2. The principle of data minimization, ie that the data collected are not excessive, is not met in practice and is also countered against the same logic of the Big Data. The new analytical models are based precisely on massive amounts of data.
3. The informed consent of the person is not the solution to the use, modification and transformation of your personal data, as observed in daily practice.
4. The anonymous personal data have deep limitations and do not prevent reidentificar the subjects.

Aware of this technological and legal reality, the European Union has issued, as explained above, Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural Concerning the processing of personal data and the free movement of such data. At the same time it repeals the previous Directive 95/46 / EC (General Data Protection Regulation - RGPD).

This is of esencial importance since both the Regulation and the Directive are binding rules, which affect all Member States but with a clear difference. While the Directive is a rule binding on every Member State as to the result to be achieved, allowing national countries the freedom of the means to achieve that result (which in practice implies that the Directive needs to be transposed into national legislation) The Regulation, on the other hand, has direct applicability, in an attempt to preserve the privacy of European citizens, with a wider content such as the right to be forgotten and data portability.

The problem of Big Data and its legal effect is complex as it affects a multitude of contents such as discrimination, civil liability, competencies, intellectual property rights, etc., which require an interdisciplinary solution.

In addition to the former, new issues are included within the right to privacy such as the right to forget and data portability.

Solutions how has already indicated through a tougher regulation regarding the obligation on the part of the companies of a transparency in the use of data and informed consent of these by the citizens.

In this sense, the European regulations are intended to regulate the different stages in which the Big Data is developed.

Thus in a first phase, which affects the collection of information and use of algorithms or automated procedures. It is possible to know that all the data that we have of a person, belong to that same person, being able to analyze their activities, tastes, etc., but it is impossible or difficult to be able to determine of which person is.

The excessive protection of these data can directly affect the quality or usefulness of the results that can be extracted. The purpose is to obtain relationships between the data for study, research or common use without prejudicing the holder of such data. The regulations allow the interested party specific procedures to express their consent or not, and to be able to leave the data collection whenever they want.

In a second phase, where the data are applied by a particular person, they require their specific consent. And this, because its use has a greater privacy risk.

It has already been pointed out above that Directive 95/46 / EC required specific requirements to give effect to such consent (which has to be unequivocal, necessary for the fulfillment of a contract, obligation, vital interest of the interested party or a public interest, etc). This Directive was transferred to our legislation with the Organic Law of Data Protection, which is more complete and demanding than the Directive, requiring it to be free, unambiguous, specific and informed [39] and also in writing [41]. As can happen in medical records, economic capacity, etc.

Special mention must be made in those online contracts where the unilateral clauses of the contract prevent knowing the content of the contract and its obligations. It can be greatly improved with a less technical, simpler wording, with a real notification of the terms of privacy. To this must be included the exit procedures of the interested party.

Likewise social networks can raise new problems consisting of a person who does not want to contribute any data of his, can be obtained indirectly, by being included in a Facebook group. An example is the knowledge of his homosexual status, because he belongs to a group of contacts of said social network. Thus, although this person does not want to know relevant personal data (ideology, religion, etc.), that information can be public or deduced by being in their contacts.

The development of opt-in procedures (need to express consent) and opt-out (indicating their refusal to collect their data) is necessary for a greater control of that consent, seconds.

The pass of time will tell us if this Regulation has allowed the protection of data as a critical factor to achieve a strengthening of the information society, determining to provide products and services respectful and quality, generating confidence in users of them.

CAPITOLO 12
CONCLUSION

The Big Data is a technology that has been consolidated between the year 2011 and 2013. Due to a number of factors is increasingly used in companies around the world and the data of all users are stored in department stores. There have been cases such as global espionage where it has been proven that the privacy of users has been violated on several occasions.
There are several factors that do not penalize these privacy violation behavior in the right way. The most important are:

1. General lack of knowledge: This is one of the key points. There is widespread ignorance on the part of the user who, as has been seen, is left with a great responsibility and yet does not know the risks involved in each decision made.
2. There is no suitable regulation: The current regulation is characterized by its absence. Norms have been created over the last few years but do not fit the scale of the problem. The legal aspect of this issue is far behind in most countries.
3. The advancement of technology is faster than can be regulated: not only must regulate all current privacy issues but all new techniques that may emerge in the future. More and more techniques are being created to deanonymize the data or techniques seen as the fingerprint that are hidden in a legal vacuum.

In order to solve this complex problem, which does not have large sources of information, all users should be informed of the way in which they analyze their data and renew the regulations in a regular way according to all that this matter advances. The problem of privacy is a problem today and if not regulated in an effective way in the near future will be one of the biggests problems in society.

Bibliografical References

[1] C. Jones, “Historia Cronológica del Big Data,” 2015. [Online]. Available: http://www.winshuttle.es/big-data-historia-cronologica/.

[2] C. Alonso, “Big Data y Privacidad por Chema Alonso,” 2014. [Online]. Available: https://www.youtube.com/watch?v=_oeqjepFkEY.

[3] C. Alonso, “¿Qué es el Big Data?” [Online]. Available: https://www.youtube.com/watch?v=24AbvDimPkQ.

[4] “Google Trends ‘Big Data,’” 2017. [Online]. Available: https://trends.google.es/trends/explore?q=big data.

[5] M. Mulcahy, “Big Data Statistics & Facts for 2017,” 2017. [Online]. Available: https://www.waterfordtechnologies.com/big-data-interesting-facts/.

[6] ComunicacionFHF, “Utilidades del Big data,” 2016. [Online]. Available: http://www.factorhumanoformacion.com/utilidades-del-big-data/.

[7] R. Caballero, “Con V de Big Data,” 2017. [Online]. Available: http://tecnologia.elpais.com/tecnologia/2017/04/26/actualidad/1493195037_932452.html.

[8] E. Gil, *Big data , privacidad y protección de datos*. 2015.

[9] Logicalis, “Las cinco principales aplicaciones del Big Data,” 2014. [Online]. Available: https://blog.es.logicalis.com/analytics/las-cinco-principales-aplicaciones-de-big-data.

[10] I. para tu Negocio, “Beneficios del Big Data para tu empresa,” 2016. [Online]. Available: https://www.informaticaparatunegocio.com/blog/beneficios-del-big-data-empresa/.

[11] L. Aguilar, *Big Data, Análisis de grandes volúmenes de datos en organizaciones*. 2016.

[12] Adexus, “Multiples beneficios del Big Data para tu empresa,” 2016. [Online]. Available: https://www.adexus.com/noticias/los-multiples-beneficios-de-big-data-para-las-empresas/.

[13] E. Quintana, “El Big Data moverá más de 187.000 millones de dólares en 2019,” 2016. [Online]. Available: http://www.muycomputerpro.com/2016/05/29/idc-el-big-data-movera-mas-de-187-000-millones-de-dolares-en-2019.

[14] J. Serrano-Cobos, “Big data y analítica web. Estudiar las corrientes y pescar en un océano de datos,” pp. 67–79, 2011.

[15] L. de la UE, “Protección de datos y privacidad,” 2017. [Online]. Available: http://europa.eu/youreurope/citizens/consumers/telecoms-internet/data-protection-privacy/index_es.htm.

[16] M. Rouse, “entity tag (ETag),” 2014. [Online]. Available: http://whatis.techtarget.com/definition/entity-tag-Etag.

[17] Lucb1e, “Cookieless cookies,” 2013.

[18] E. F. T. Madrid, “Vivir en un mar de datos: Big Data, privacidad y seguridad,” 2014. [Online]. Available: https://www.youtube.com/watch?v=zQqc05EqrjM.

[19] C. Europeo, “Reforma de la protección de datos,” 2017. [Online]. Available: http://www.consilium.europa.eu/es/policies/data-protection-reform/.

[20] F. T. C. (FTC), “Protecting Consumer Privacy in an Era of Rapid Change. Recommendations for Businesses and Policymakers,” 2012.

[21] P. Europeo, *Considerando 26 de la Directiva 95/46/EC*. 1995.

[22] Mazikeen, “Un señor lee en voz alta los Términos y condiciones del Kindle de Amazon y fija el récord el 8 horas y 59 minutos,” 2017. [Online]. Available: https://www.meneame.net/m/ocio/senor-lee-voz-alta-terminos-condiciones-kindle-amazon-fija-8-59.

[23] OSI, “¡Lee antes de aceptar! Lo que no leemos de las Condiciones y Términos de uso.,” 2015. [Online]. Available: https://www.osi.es/es/actualidad/blog/2015/02/16/lee-antes-de-aceptar-lo-que-no-leemos-de-las-condiciones-y-terminos-de-uso.

[24] C. Alonso, “You Are: Where You Are por Chema Alonso,” 2016. [Online]. Available: https://www.youtube.com/watch?v=mMI_rYKPapU.

[25] I. de I. del conociciemnto (IIC), “Seguridad en Big Data, privacidad y protección de datos,” 2016. [Online]. Available: http://www.iic.uam.es/innovacion/seguridad-big-data/.

[26] F. Andrades, “Big Data y la privacidad: Cuando el negocio eres tú,” 2013. [Online]. Available: http://www.eldiario.es/turing/BigData_0_120038458.html.

[27] P. Fy and F. Percent, “TOP SECRET // SI / TK // NOFORN Resource Exhibit No . 1A National Intelligence,” no. 1. 2013.

[28] Wikipedia, “Global surveillance disclosures (2013–present),” 2017.

[29] J. Stobart, “Britain denies using PRISM to get around domestic spying laws,” 2015. [Online]. Available: http://articles.latimes.com/2013/jun/10/world/la-fg-wn-britain-nsa-prism-surveillance-program-20130610.

[30] C. Española, *Boletín Oficial del Estado 311 de 29 de diciembre de 1978*. 1978.

[31] B. O. del Estado, *Ley Orgánica 15/1999 de 13 de diciembre de Protección de Datos de*

Carácter Personal. 1999, p. núm. 298 de 14 de diciembre (pág. 43088 a 43.099).

[32] B. O. del Estado, *Real Decreto 1720/2007 de 21 de diciembre por el que se aprueba el Reglamento de desarrollo de la Ley Orgánica 15/1999 de 13 de diciembre de protección de datos de carácter personal*. 2007, p. (pág. 4103 a 4136).

[33] I. T. Consulting and S. Tel, "Análisis de la legislación sobre protección de datos personales," 2016.

[34] R. S. González, "Breve comentario a la Ley de Protección de datos de Carácter Personal," 2007. [Online]. Available: http://noticias.juridicas.com/conocimiento/articulos-doctrinales/4295-breve-comentario-a-la-ley-de-proteccion-de-datos-de-caracter-personal/.

[35] C. L. Serrano, *La Ley de Protección de Datos . Análisis y comentario de su jurisprudencia*. 2008.

[36] B. Gisbert, "La Unión Europea quiere acabar con el vacío legal del Big Data," 2017. [Online]. Available: http://www.lavanguardia.com/economia/20170109/413193723623/reglamento-ue-big-data-proteccion-consumidor-consentimiento.html.

[37] A. T. Reigada, *Comentario a la Ley Orgánica de Protección de Datos de Carácter Personal*. 2010.

[38] P. Europeo, *Reglamento (UE) 2016/679*. 2016.

[39] U. Europea, "Protección de los datos personales," 2014. [Online]. Available: http://eur-lex.europa.eu/legal-content/ES/LSU/?uri=CELEX:31995L0046.

[40] B. van Der, Schendel, and S. Van, "Ten Questions for Future Regulation of Big Data: A Comparative and Empirical Legal Study," 2016. [Online]. Available: https://www.jipitec.eu/issues/jipitec-7-2-2016/4438.

[41] C. Duhigg, "How Companies Learn Your Secrets," 2012. [Online]. Available: http://www.nytimes.com/2012/02/19/magazine/shopping-habits.html.

[42] C. CARABAÑA, "CUATRO CASOS EN LOS QUE EL 'BIG DATA' PASÓ DE ÚTIL A ESCALOFRIANTE," 2015. [Online]. Available: http://elpais.com/elpais/2015/08/11/icon/1439304143_858615.html.

[43] Sisense, "Sisense web site," 2017.

[44] "Sisense capture," 2017.

[45] Kumar, "Supervised Machine Learning," 2017. [Online]. Available: http://feynmand.com/ml-vader/.

[46] A. Monleon-Getino, "El impacto del Big-data en la Sociedad de la Información. Significado y utilidad"," *Hist. y Comun. Soc.*, vol. 20, no. 2, pp. 427–445, 2016.

[47] "The non-negative matrix factorization toolbox for biological data mining," 2013.

[48] U. Europea, "Legislación sobre las estadísticas de la UE," 2017. [Online]. Available: http://eur-lex.europa.eu/content/legis/legis-statistiques.html?locale=es.

[49] D. 95/46/EC of the E. Parliamen, *Opinion 8/2014 on the on Recent Developments on the Internet of Things*. 2014.

[50] U. de Valencia, "Seguridad física," 2015. [Online]. Available: http://www.uv.es/~sto/cursos/icssu/html/ar01s04.html.

[51] B. Angel, "Seguridad, Mecanismos de Seguridad, Concurrencia, Metodos de Control," 2009. [Online]. Available: http://unefabasededatos2009.blogspot.it/2009/04/seguridad-mecanismos-de-seguridad.html.

[52] V. Martínez, "Big Data para medir la salud al minuto," 2016. [Online]. Available: http://www.innovacionensalud.elmundo.es/tecnologia-de-la-salud/big-data-para-medir-la-salud-al-minuto.

[53] R. Riddle, "13 Predictions About The Future Of Big Data." [Online]. Available: http://www.dataarchitect.cloud/13-predictions-about-the-future-of-big-data/.

[54] B. Open4U, "Infografía: Big Data, presente y futuro," 2015. [Online]. Available: https://bbvaopen4u.com/es/actualidad/infografia-big-data-presente-y-futuro.

[55] E. Pulido Cañabate, "Big data : ¿ solución o problema ?," p. 48, 2014.

[56] J. G. FERNÁNDEZ, "La banca española confía su futuro al 'big data,'" 2017. [Online]. Available: http://www.expansion.com/economia-digital/companias/2017/04/12/58ecf5eb468aeb0c7d8b45b5.html.

[57] M. Kanellos, "Amount of Data Created Annually to Reach 180 Zettabytes in 2025," 2016. [Online]. Available: https://whatsthebigdata.com/2016/03/07/amount-of-data-created-annually-to-reach-180-zettabytes-in-2025/.

[58] P. Gonzalez, "¿QUE ES UN ZETTABYTE?," 2013. [Online]. Available: http://queesunzettabyte.blogspot.it/2013_09_01_archive.html.

[59] A. Bansal, "6 Reasons Why Big Data Career Is A Smart Choice," 2015.

[60] R. Colmenarejo, "Ética y Big Data," 2015. [Online]. Available: http://www.loyolaandnews.es/etica-y-big-data.

[61] Z. Bauman, *Etica Postmoderna*. 2005.

[62] U. Beck, *Risk Society*. 1992.