LEXIA
AVVOCATI

THE RELATIONSHIP BETWEEN BIG DATA AND THE RIGHT TO THE PROTECTION OF PERSONAL DATA

AUTHORS

Aurora Agostini, *Counsel*

Jessica Giussani, *Associate*

Data & Technology Innovation Team

SUMMARY

THE RELATIONSHIP BETWEEN BIG DATA AND THE RIGHT TO THE PROTECTION OF PERSONAL DATA

In recent times, the so-called Big Data collection and analytics has profoundly changed the way of doing business, making their use almost essential for companies that wish to gain a competitive advantage. Indeed, Big Data provides companies with important information on the consumption habits of their customers, that can be used to guide their own business. The development of an economy that becomes more and more digital, based on the collection and analysis of a huge ever-growing amount of data, makes Big Data one of the most relevant and discussed sectors of Internet law.

Although the concept of Big Data is relatively new, the origins of large data systems date back to the middle of the last century: as early as 1950, an article written by researcher Alan Turing opened with the very strong provocation: "Can machines think?".

Five years earlier, during the Second World War, thanks to Turing's theories one of the first computers (the Colossus) had been developed with the intent to identify and decrypt messages that were being intercepted at an unimaginable speed for that era, by reducing the time spent from weeks to a few hours.

Thanks to the development of the first data centers, for the first time in the 1960s it was possible to digitize and archive large amounts of data, even though in servers that were very different in size and power compared to the current ones.

Eventually, in the 90s, with the birth of the web, the foundations for the real Big Data revolution were laid. However, it was only around 2005, with the advent of the first social networks, that the real potential of the huge amount of user-generated data through Facebook, YouTube and other online services was understood:

in the same year the first software created specifically to store and analyze large data sets was developed - Apache Hadoop - whose updated version is still used today by companies such as Ebay, Facebook, LinkedIn, Twitter and Spotify to process Big Data.

What are the features of big data?

Big Data means a data set of significant dimensions from various sources, including smartphones, social media, wearables, etc. This data can be presented both in a structured form, which is more easily analyzable, and in an unstructured or semi-structured form (for which more complex data processing systems are required).

Big Data can also be defined through the so-called. three “V”s:

- volume of data generated and collected;
- variety of types of data available (the more the production sources vary, the more the available data vary);
- velocity of processing operations.

In addition to these words, further Vs have been identified, including the value that the data have once they have been processed and analysed.

The knowledge extraction process

The ultimate purpose of the complex processes behind the use of Big Data is to improve the efficiency of production processes and guide the strategic choices of those who manage a business. This occurs thanks to a more accurate identification of market trends, and therefore of the ideal consumer target for a specific product or service, as well as a more targeted advertising and commercial proposals. For this purpose, the process of "extraction of knowledge" from Big Data is essential and it is developed in 3 phases:

1. Collection
2. Processing
3. Interpretation and use

1. BIG DATA COLLECTION

Data origins come from multiple sources. Especially nowadays, where all content is made available in digital format and most of the activities are carried out online, it is easy to obtain large amounts of data from users. Consider, for example, personal devices such as smartphones, tablets and computers, satellite navigation, social networks, in which users publish their own content (photos, videos, texts), apps and websites.

Additionally, there is the Internet of Things, which apply both in the industrial field (for example in the so-called predictive maintenance), and also with regard to the life of individuals, from home automation to wearable devices (for example, wearable devices that collect data relating sports activities and/or biological parameters).

2. DATA PROCESSING

Stand-alone data have a low value, but they acquire it when they are organized: for this reason, the processing phase plays a central role in the entire Big Data supply chain, because it allows the organization of unstructured raw data into information that can be used for economic purposes.

Information is the result of the data analysis process.

3. DATA INTERPRETATION AND USE

Once the information has been processed, it shall be properly treated so that it can be used in practice. This is where Big Data Analytics technologies step in, allowing data transformation into useful and valuable information for the business (for example, in making better decisions, improving performance and/or productivity, increasing profitability and competitiveness, etc.).

To simplify, Big Data Analytics are extraction logics, analysis methodologies and mathematical models of prediction and optimization - they can be divided into four categories:

- Descriptive Analitycs: they are tools that describe current processes or past performances. For example, they allow to view (Visual Analytics) the main performance indicators.
- Predictive Analytics: these are tools that analyze data to understand what could happen in the future. They often resort to techniques such as regression, forecasting, predictive models and are based on Machine Learning.
- Prescriptive Analytics: they are tools capable of proposing operational/strategic solutions that are useful to the decision maker in order to make their own choices.

- Automated Analytics: these are tools capable of autonomously implementing the choices that, based on the analyzes carried out, they consider more valid than the pre-set objectives (for example, if a customer is identified by the analysis as "at risk of abandonment", the A.A. can choose to initiate a loyalty action, such as sending a reserved promotion).

The last three categories belong to the so-called Advanced Analytics - an extremely advanced analysis tools that have a great impact on all business processes.

The daily use of Big Data

For the first time in the history of mankind we are able to keep track of what millions of people do every day. Here are some practical examples of how Big Data is a phenomenon that closely affects our daily lives:

- Smartwatches keep track of the various activities carried out during the day through the analysis of the steps taken and heartbeat. These tools are able, for example, to identify when the individual wearing them is under stress.
- Spotify doesn't limit itself to keeping track of users' musical tastes. Indeed, the App is able to recognize when users are happy, when they are traveling or are having a party, by analyzing the playlists that he or she plays during the day.
- Delivery apps collect and sell data relating to users' consumption habits and restaurants within the user's area of interest. The restaurants will then be able to use this information for commercial and marketing activities such as, for example, sending discount codes or promotions.
- Car/bike/scooter sharing platforms identify the places where the users hang out thanks to geolocation.
- Electronic payment tools allow to capture information about the user buying behavior and preferences. They are useful to verify the effectiveness of personalized advertising campaigns, as well as to further profile their users.

Is it possible to protect user privacy?

Ultimately, the tool that most favors the tracking of user data is the smartphone. It plays a central role in data acquisition, as it is connected to the Internet, accompanies the user in all his daily activities and has numerous input features (such as motion, brightness, location sensors, keyboard and the touch screen).

This large amount of data is created by the use of smartphones thanks to two main tools:

1. the geolocation system which, even if not activated, still tracks the user's position.
2. Applications that often require access to contacts, microphone, photos and other features capable of taking track of the user, even if they are not necessarily pertinent to the use of the application itself, and that may be subject to transfer to third parties.

The value of big data for business

The value of Big Data can be easily understood by referring to the so-called Data Economy: today, indeed, data constitutes an inestimable value for those who are able to extract, use and monetize it.

There are mainly two business models to create profit from the Big Data exploitation:

1. by collecting, processing and interpreting data relating to its users in order to improve its service and/or for specific purposes of the company and then use them internally.
2. By selling data to third parties. This implies the establishment of the role of Data Broker, i.e. professionals (often companies) who deal with recovering data and information and then

processing, interpreting and analyzing them in order to create a profile. This profile can be sold to third-party companies interested in profiling users to better identify who falls within the target of their business and customize advertising campaigns accordingly. It should be noted that the data is not only sold to companies that sell consumer products but also to financial companies that use it to identify risk profiles and interest rates to apply to a potential customer.

GDPR and Big Data

So how does the use of Big Data coexist with the rules on the protection of personal data?

The problem arises from the moment of the collection of Big Data. In the huge variety and quantity of information found, it may happen that data of a personal nature is collected. In this case, the processing cannot exist - at least in Europe - without compliance with the GDPR. However, it is often problematic (i) on the one hand, to define the demarcation line between data of personal nature, especially due to the possibility of reconnecting apparently anonymous information to single individual thanks, for example, to increased computing capacities and the plurality of archives that can be used in hypotheses. Consider pseudonymized data - for example IP address, which is only partially hidden; (ii) on the other hand, the massive acquisition of data makes it difficult to identify the specific ex ante of the purposes of the related processing. The GDPR requires that the collection and use of data can occur upon request for the consent of the interested party or upon the occurrence of one of the conditions provided for by art. 6. It is also established that personal data are processed in a lawful, correct and transparent manner, they are collected and processed for specific, explicit and legitimate purposes and are adequate, pertinent and limited to what is necessary with respect to the purposes for which they are processed (the data minimization principle). These data shall also be accurate and, if necessary, updated, as well as stored in a manner suitable to identify the interested parties for the time necessary to achieve the purposes for which they are processed. Lastly, they must be treated in order to ensure adequate safety. However, principles such as minimisation, limitation of purpose and retention for the time necessary for carrying out the processing are not suitable for massive collections. The data is indeed collected not for current needs but in view of future and possible needs and reused for purposes that are not always compatible with the original ones.

The advent of the GDPR has certainly made the transfer of data more transparent, however it is still not enough. If it is true that the user shall give his consent and has the right to know to whom this data is sold, very few people have actually the time or skills to read and verify the information that is offered to them every time they access to a site. The average user accepts what is offered to him in order to access the content of his interest as quickly as possible: this is the so-called privacy paradox.

A further problem concerns the fact that users often seems to have no alternatives: although they are informed or partially informed, the refusal to provide certain data can compromise the use of the complete experience of the website, application, etc.

In any case, it is important to bear in mind that if a company has collected data using the data subject's consent or a legal obligation as a legal basis, no further processing is permitted outside the areas covered by the original consent or by the legal provision.

For this reason, instead of collecting the express consent of their users/customers, the data controllers of personal data often prefer to use the legal basis of legitimate interest, whose demarcation limits are not yet clearly defined.

For the moment, a possible solution could be the dynamic consent: according to this model, the individual initially gives a broad consent in the face of a general information about the possible purposes of the processing and, subsequently (once specifically identifying the purpose of use of the data) receives a more

precise information with the request for a new and more specific consent to the treatment. However, it is a model that for now has only found application for bio-banks and for consensus on bio-medical research.

Are the interested parties owners of their data?

The interested parties always have the right to access, rectify and delete personal data and to restrict its processing under the GDPR: this means that companies must be able to dig into the vast amount of data stored in different systems to locate and/or delete the data belonging to the interested party. Naturally, the interested party does not have the possibility to verify that the owner has actually deleted such data, nor is he/she able to know if the deletion also concerned those data that may have been transferred to third parties.

Furthermore, many companies claim to operate in compliance with the GDPR and to use personal data in a lawful and correct manner, providing the interested party with timely and complete information, but they often forget to delete unused data or whose retention period it is over. This conduct also constitutes a violation of the GDPR: indeed Article 5 establishes that personal data should not be kept longer than necessary. The companies cannot therefore limit themselves to establish retention periods, but they must also ensure that the data is effectively deleted, regardless of an explicit request from the interested party.

Can anonymization be the solution?

A possible solution to the problem of using personal data contained in Big Data (i.e. when they are transferred to third parties or stored longer than expected) could be to make them anonymous. With this regard, it should be noted that there is a profound difference between pseudonymisation and anonymisation.

Pseudonymisation consists of the replacement of direct identifiers in a way that the data can no longer be assigned to a specific individual without the use of additional information. This means that pseudonymised data remains personal data; and therefore, subject to the application of the GDPR.

Anonymisation, on the other hand, refers to the practice of rendering data unidentifiable in a way that it is impossible to reconstruct the identity of the data subject.

In theory, anonymisation could be a solution if companies are certain that it is impossible to re-identify the individual. In the reality, this is hardly possible; indeed, the algorithms used by Big Data technologies are able to identify and re-identify, by comparing various databases, a user. Therefore, even anonymization was considered a reliable technique until recently, nowadays in the era of Big Data no longer seems to be sufficient to guarantee the effective confidentiality of data.

The new Digital Service Package

In the near future the scenario could change. The implementation of the Digital Service Package, made up of two regulations, the Digital Services Act and the Digital Markets Acts, introduces a special regime for the so-called Gatekeepers with reference, among other matters, to the protection of personal data.

The Gatekeepers of the digital market are the service providers of basic platforms, such as social networks, browsers, search engines and messaging services and can be identified on the basis of three different parameters:

- company size: annual turnover of €7.5 billion or more in the last three financial years or total market share value of at least €7.5 billion in the last year and supply of platform services to at least three states of the EU;
- control of the user data access gateway: It is valuated the registration of at least 10,000 active European users during the last year and more than 45 million active European users per month;

- lasting and stable position on the market, if the thresholds identified in the previous principle have been reached in each of the last three financial years.

These regulations are fundamental in terms of Big Data, as the Gatekeepers identify themselves in that handful of companies that manage data from almost all over the globe, among others, Meta, Google, Apple, Amazon, Microsoft.

Some of the main innovations introduced by the two regulations aim to place restrictions on advertising based on user monitoring, targeting and profiling. In particular, the content of the legislation prohibits Gatekeepers from:

- profile minors for commercial purposes;
- process the data collected through third parties who use the Gatekeeper platform to offer advertising services;
- combine the personal data collected on the platform with those collected on any other Gatekeeper or third-party platform or with those coming from other services offered separately by Gatekeeper itself;
- automatically subscribe the user to other Gatekeeper services to combine personal data.

In conclusion, we believe that recent regulatory developments has finally put a spotlight on the protagonists of this market who until now had been acted according to the motto of "too big to care". The message that the competent authorities are trying to convey is that Big Data and Privacy can and must coexist, while always respecting the explicit will of a consumer who is adequately trained in understanding the importance of their data and their use.

LEXIA
AVVOCATI

Avv. Aurora Agostini
Contacts

Dott.ssa Jessica Giussani
Contacts

www.lexia.it
Via del Lauro 9,
20131 Milano (MI)
Follow us on LinkedIn