CONTEMPORARY COHERENCE OF BIG DATA INTO I.P.

H N Shree Harini[1] & Dr. Ranjit Oommen Abraham[2]

ABSTRACT

As the evolution of man took place, his intellect started to develop eventually. The concept of 'Intellectual Property' emerged in the 17th century itself but the term began to be used only since the 19th century. As technological development took place, the intellectual property started its march towards reaching the zenith since the 20th century. In the present 21st century, it can be observed that the intellectual property has ventured in almost all the fields, and the digital world is no exception. The businesses today are no longer advocating the storage of data manually. The data is being stored and processed digitally in huge quantities and this kind of a huge collection of data is known to be the 'Big Data'. The legislators felt the need for protecting the digital data, they amended the various Intellectual Property Rights. As the usage of big data also started increasing, the need for protecting these big data as an intellectual property emerged. This paper dwells into the big data and various aspects of intellectual property like the copyrights, patents and trade secrets. The paper will also elaborate about the benefits of protecting the big data as an IP and also, about the hindrances that come up with it.

Keywords: Intellectual Property, Big data, Copyrights, Patents, Trade Secrets.

Introduction:

With the today's world becoming a knowledge- based society, every new idea is playing an essential role in the business scenario. In an era where an idea is generating capital, along with the idea comes the concept of 'Intellectual Property'. Intellectual Property can be precisely defined as the rights that are granted legally for the creators and innovators to gain the economic benefits out of their creations and innovations respectively. This ensures that their works are not misused by anyone else maliciously.

Intellectual property is defined by the *Oxford English Dictionary* as "*Intangible property that is the result of creativity*".

---

[1] B.B.A.LL.B. (Hons.) 2nd Year, School of Excellence in Law, Dr. Ambedkar Law University,Chennai.
[2] Registrar (resp.) School of Excellence in Law, Dr. Ambedkar Law University, Chennai

Intellectual property rights are the rights that adhere to such creations and that grant the holder(s) thereof a monopoly on the use of that creation for a specified period and subject to certain exceptions.[3] The underlying aim of granting such (temporary) monopoly, which – admittedly – entails a certain social cost, is to incentivize creators to share their creation with the public, and to achieve the social benefits of increased creative activity.[4]

What is big data?

Big data refers to the huge and complex data that cannot be processed by traditional methods. Even before the usage big data gained momentum, there were a lot of efforts made to store and access large information for analytics. But none of such efforts succeeded until the concept of 'Big data' came into the practice in the early 2000s.

In simpler terms, it can be said that big data is a huge collection of data that keeps growing exponentially with time. It can be structured like that of an employee table in a company's data base or unstructured like that of the results of the "Google Search" or even semi-structured like that of a personal data stored in an XML file.

Characteristics of Big data:

The characteristics of the big data can be described by using the 5V's which are:

- Volume
- Variety
- Velocity
- Variability
- Veracity

The name itself states that the size of the data is huge. Only such enormous collection of data can be put under the umbrella of 'Big data'. Whether a particular collection of data is a big data or not, depends on its size.

Secondly, the variety of data being stored in big data is varied i.e., heterogeneous data is only stored in big data. Thirdly, the speed at which the data traverses from sources including such business processes, application logs, networks, social media websites, sensors, mobile devices, etc. is referred to as big data velocity. There is an immense and continual influx of data.

---

[3] R. S. Khemani and D. M. Shapiro, 'Glossary of Industrial Organisation Economics and Competition Law' (OECD 1993) <http://www.oecd.org/regreform/sectors/2376087.pdf> Accessed on 27.06.2022.
[4] Id.

Fourthly, since big data consists of a huge collection of data, each data might be in different formats and thus, maybe be collected from different sources. This makes it seem to varied in nature.

And finally, the enormous data stored in the form of big data is not in consistency with each other. They are all inconsistent and it is difficult to predict their certainty.

Interaction between IP and Big data:

An insight into the world of big data tells that it necessitates access to massive amounts of information for its growth and dominance in analytics. With more than three quarters of the world involved in being online, businesses are able to draw conclusions and make decisions mostly through the data available online. And now, this places a demand from the creators' side to deserve and withhold the originality of their works.

And this is where the interface between the big data and the Intellectual Property protection happens. Even though big data cannot generate any meaningful insights by itself, it can be analyzed, which is where intellectual property enters the equation. The patented hardware used to access, gather, and store data, as well as the copyrighted software that aids in deduction, are all examples of intellectual property. Moreover, once the deduction is made, this may yield protectable subject matter or assets which can be useful to the company and, as a result, be secured as a trade secret.

When the big data is being processed for analytics, there are huge chances that there might be an infringement of the Intellectual Property Rights like copyrights, patent, trade secrets, etc. The freely available information on the big data can cause severe losses to the businesses if it infringes the IPRs. This happens in a way that the businesses may use up the creator's original work and build their big data. Thus, ensuring moral usage of big data analytics assists in ensuring maintenance of IPRs.

Big data and Copyright:

According to World Intellectual Property Organization (WIPO), *"Copyright is a legal right that is provided to the creators of literary, dramatic, musical and artistic works and producers of cinematograph films and sound recordings."*

It can be said as an accumulation of rights of adaptation, reproduction and translation of another's work. By means of copyright, one's creativity is protected from stealing and rewarded. This encourages the creators to come up with more and more creative works. In India, the copyright laws are enshrined in the *Indian Copyright Act, 1957* which is protecting literary, dramatic, musical and

artistic works and producers of cinematograph films and soundrecordings from unauthorized usage. When a creator of a tangible work is given a privilege like that of a copyright, then what will be the stature of a creator of an intangible asset? Well, in order to provide the same kind of dealings with the digital creators, copyrights have been introduced into the digital field. The emergence of online platforms and Internet have enhanced the significance of using the copyright laws in big data. Copyright comes into the picture since the law safeguards the computer software and programs that are used to collect and analyze big data. These tools are useful in data analytics which involve protecting the data that can be searched, segregated, transformed or deleted. Notwithstanding the growth of powerful datasets on the utilization of such works and the technological resources available to content and technology companies, there are still serious accountability and accounting issues related to the use and remuneration of such works in copyright.

Criteria for qualification of being copyrighted:

For a work to be protected by copyright, it must only have the personal touch of the creator. It must be the author's original creation. A software like big data will be protected only if it is original, which is one of the basic tests to qualify for a copyright. Most of the information collected and processed in the form of big data will not be original and thus, cannot be protected. So, the law been tailor made that the software need not be mandatorily hundred percent original. A certain degree of originality will fulfill the criteria of originality in order to be copyrighted. But this degree of originality varies from each and every country. Some countries follow lower degree and while others follow a higher degree of originality.

Similarly, for a software to be protected under copyrights, it must be fixed upon a tangible medium. When a work is stored on a medium that allows it to be perceived, reproduced, or otherwise communicated, it is considered to be fixed. The way of keeping data can range from handwritten notes (files) to photographic documentation (images) or recorded testimonies (sound) to digitized archives (digital files), as long as it is concrete, easily identified, and described. For the time being, results that have not yet been produced (future data) or results that cannot yet be described (e.g., because there are no means to express them) are not protected by copyright.[5]

Is copyright an absolute right?

The creator once he copyrights his works, is conferred upon with numerous exclusive rights with respect to the facilitation of usage of the work by a third party with or without

---

[5] Big data & Issues & Opportunities: Intellectual Property Rights - Bird & Bird (twobirds.com) Accessed on29.6.2022.

authentication. The rights of copying, deleting, translation of another's works is balanced by some copyright exceptions which indirectly allows an optimal protection of the creator's interests. These exceptions are elaborately explained in the catalogue of the rights related to copyrights. These rights and exceptions deal with the scope of using another's works but at the same time, protect the legitimate interests of the creator.

Now a question arises as to how these rights and exceptions are incorporated in a data environment. Many big data analytics feel that the significant hindrance caused by copyright protection that they face is the requirement to obtain permission from the copyright holder for each individual piece of data. Usually, big data deals with huge amounts of data. Now, copyrighting each and every data will be a great deal of work for them. Another major problem is that of identifying the creator or owner of the copyright of every data. Each data collected will have its distinct creator. Tracking these creators and then using their data is a time-consuming as well as a tedious process. Moreover, there might be some anonymous creators who do not prefer to reveal their identity. Such kind of creators pose a bigger task to the big data analytics to figure out their rights.

Another major threat these big data analytics face is that a particular data cannot have more than one copyright. For example, if a particular creator already owns the copyright over a specific data, then another creator with the same kind of data cannot claim the copyright, irrespective of whether they have collected from the same or different sources. Each copyrighted data must be distinct and unique. But at the same time, the law does not prohibit the co ownership of two or more creators on a particular data. Say, if a certain data has been collected and processed by two creators jointly, then both of them can own a copyright over the data worked upon by them.Electronic Copyright Management Systems:

The IPR Acts does not provide with legal protection through the copyright laws. It also provides technological protection through methods like Electronic Copyright Management System (ECMS) and prohibiting the acts that are overriding the copyright laws. The ECMS, also known as DRMS (abbreviated as Digital Rights Management System) makes the copyrighting protocol tougher but the licensing protocol easier. The enormous collection of data constituted in the big data can be protected through the technological ways. Some of these are through watermarking, finger printing, tamper proof hardware and software; access control by user ID and password; content use through disabling printing and downloading, copying specified number of times only and restricting copying through originals (masters) only. The DRMS was, thus created with the sole purpose of protecting the digital product from unauthorized use and distribution.

In a big data environment, copyrighting has become simpler than before. The DRMS will itself scrutinize whether a particular data will be qualified to be copyrighted or not and whether its origin is from the creator himself or stolen from other's. It provides a continual protection of the works and rights, no matter where the works are located and who possesses the rights. It is going to ensure that the digital products are transferred to authenticated persons and devices. Some techniques of the DRM like watermarking, encryption, finger printing, etc. are going to block all the unauthenticated access of the data, if found any.

Complications in copyrighting big data:

The complication from a practical point of view arises as it is difficult to select or arrange big data primarily because of its 5V's and also big data is almost always automatically generated in a segment from varying sources, which introduces the need of case-to-case analysis of whether an invention about the said subject-matter shall attract copyright protection or not. This brings with the complication of tracking the legal and illegal usage of such data.

The process of copyrighting the big data is still in its amateur stage. An in-depth analysis is required to find out about how online infringement happen while processing the big data and to identify the part that the internet has to play in blocking the unauthenticated data. Moreover, the rights that are entitled to a digital creator has to be jotted down clearly because only then copyrighting can be utilized to the fullest. By doing so, digital creators will be motivated to come up with new innovations and creations in the field of big data.

Big data and Patents:

When an inventor uses his intellect and invents a new product or process which might be the probable solution to an existing problem, then such an invention should be given some sort of protection from others copying it. It is moral to protect one's invention from the unauthorized use and distribution by the others. This kind of protection given to an invention is called as 'patent'. The world patent has been coined from a Latin word *patent-em* meaning open. A patent is a document issued by government to the inventor granting permission to exclusively make, use and sell on disclosure of the invention for a definite period of time. According to the WIPO, *"A patent is an exclusive right granted for an invention, which is a product or a process that provides, in general, a new way of doing something, or offers a new technical solution to a problem."*

In principle, the patent owner has the exclusive right to prevent or stop others from commercially exploiting the patented invention. In other words, patent protection means that

the invention cannot be commercially made, used, distributed, imported or sold by others without the patent owner's consent.[6]

In India, the laws related to the patency are enshrined in the Patents Act, 1970. According to *Section 3, 5* of the Act, inventions that are frivolous, obvious, exploiting commercially to public, immoral, prejudice to human, animal, plant life or health or to the environment, scientific principles, abstract theories, identified to possess new use for a known substance, known process, known machine or known apparatus, aggregation of the properties by admixture and process for production of such substances, arrangement or re-arrangement or duplication of known devices, methods of agriculture, horticulture, processes for the medicinal, surgical, curative, prophylactic or other treatment of human beings, animals to render them free of disease or to increase their economic value or that of their products, a mathematical or business method or a computer programme *per se* or algorithms, literary, dramatic, musical, artistic work, cinematographic works, television productions, rule or method of performing mental act, method of playing game, presentation of information, topography of integrated circuits, aggregation or duplication of known properties of a traditional knowledge, atomic energy, claiming substance intended for use, or capable of being used, as food or as medicine or drug, or substances prepared or produced by chemical processes are not patentable.

The Patents Act, 1970 was then amended by the Patents (Amendment) Act, 2005, wherein the product patent was extended to all fields of technology. Now a question arises as to whether the big data is included in this ambit? It can be said that big data is itself not patentable; but its algorithm and software programme may be protected by patents. Big data is so large that it cannot fit into the main memory of a single machine. This brings in the necessity to process big data by efficient algorithms which arises in Internet search, network traffic monitoring, machine learning, scientific computing, signal processing, and several other areas.[7] This algorithm can be protected by means of patent.

This concept was later discussed by the Australia's Full Federal Court in the case of *Commissioner of Patents vs. RPL Central Pty Ltd*.[8] It was held that the RPL's invention is a "business method or scheme" and is not patentable subject matter. This decision highlights how the Court differentiates between business and technical innovations, and sets out Australia's position on software patenting. The court reiterated the key requirements of patentable inventions – novelty, an inventive step, and manner of manufacture.

Big data does not entirely fulfil any of these criteria. The radical novelty of big data techniques

---

[6] Patents (wipo.int) Accessed on 30.6.2022.

[7] Algorithms for Big Data (CS 229r) (harvard.edu) Accessed on 30.06.2022.

[8] [2015] FCAFC 177.

would result in relatively isolated technological groups, which share a low similarity with each other. Big data being a heterogeneous mix, when separated will definitely produce only heterogeneous individual products. Moreover, big data is created on its own by means of artificial intelligence without any human intervention. So, it is definitely not an inventive step and the manner of manufacture of big data does not fall under the purview of patency.

Challenges in granting patents to big data:

As mentioned earlier, individual components of data that perform specified functionalities are patentable. Now, when such individual components are bound together as a single collection, then the notion of granting patent to this huge collection becomes a difficult and complex formula.

In accordance to the judgement in the case of *Commissioner of Patents vs. RPL Central Pty Ltd.[9]* wherein the Court held that putting a business method or scheme into a computer is not patentable unless there is an invention in the way in which the computer carries out the scheme or method.

It can be said that acquiring patents for these inventions inferred from computer-generated works can be a complicated equation since works produced by unchaperoned artificial intelligence are not patentable subject matter.

Thus, it is hard to protect big data through patency and if the system begins to issue patents to big data, then the rate at which the patency is issued will also be slower and this will make the entire process complex and time consuming. So, it can be concluded that patent and big data do not go hand in hand.

Big data and Trade secrets:

According to WIPO, trade secrets are intellectual property (IP) rights on confidential information which may be sold or licensed. In general, to qualify as a trade secret, the information must be:

- commercially valuable because it is secret,
- be known only to a limited group of persons, and
- be subject to reasonable steps taken by the rightful holder of the information to keep it secret, including the use of confidentiality agreements for business partners and employees.

---

[9] Id.

The unauthorized acquisition, use or disclosure of such secret information in a manner contrary to honest commercial practices by others is regarded as an unfair practice and a violation of the trade secret protection.

In general, trade secrets can be any confidential business information that gives an enterprise a competitive advantage and is unknown to others. Trade secrets include both technical information, such as manufacturing processes, experimental research data, and software algorithms, and commercial information, such as distribution methods, supplier and client lists, and advertising strategies. A trade secret can also be a combination of elements, each of which is in the public domain on its own, but the combination, which is kept secret, provides a competitive advantage. Other types of information that may be protected by trade secrets include financial information, formulas and recipes, and source codes.

When the domain expertise of data scientists and business knowledge is used to enrich the data collected by machines, very specific performance-enhancing actions can be taken. The additional knowledge and human expertise of this enrichment step is what allows the creation of ratios of variables, trends or derivatives (e.g., changes in customer behavior), and categories from numerical variables (e.g., low, medium, high income instead of the actual number) that then enable powerful action.[10] Companies like Amazon[11] can be taken as an example of how companies can use Big Data to further improve customer preference and develop their business. Now this naturally requires IP protection since IP deals with creation or innovation in fields of new technologies, designs and promotes economic growth and the only way to reward any person/organization that helps in this aspect is by rewarding their effort with complete and legal control over their creation so that they may exploit its use in any way they seem fit.[12]

In the case of *Palantir Technologies, Inc. v. Marc Abramowitz,[13] major* complains as to the increasing cases of theft of trade secrets where the trade secret is in form of data or a business model and is found in places where start-ups are densely present and the competition is very high. It was also felt that as access to internet and more sophisticated technology increases there will be more cases of theft of trade secrets hence it will be important for us to have Legislature to provide remedies for such theft and even act as a deterrent.

Trade secret laws protect sensitive business information that acquires value from not revealing

---

[10] Mckinsey Achieving business impact with data A comprehensive perspective on the insights value chain Authors: Niko, Holger Hürtgen, Achieving business impact with data | McKinsey Accessed on 30.6.2022.

[11] ICAS article "10 companies that are using big data" (https://www.icas.com/ca-today-news/10- companies-using-big-data) Accessed on 30.6.2022.

[12] Intellectual Property (wipo.int) Accessed on 30.6.2022.

[13] Complain Filled at US court California (http://tsi.brooklaw.edu/cases/palantir-technologies-inc-v-marcabramowitz/filings/former-palantir-employee-sued-misappropria) Accessed on 30.06.2022.

them to the public. These trade secrets remain "reasonably confidential" to the business. In the present digital era, businesses no longer believe in storing manual or tangible data. The world has changed, so as the manner of data storage and processing. Companies are now storing huge amounts of data as big data and these larger sets of data are becoming a critical asset in the economy, fostering new industries, processes, and products and generating significant competitive advantages, and one method for retaining this data as a competitive advantage is to protect it as a trade secret. At the same time, protecting big data in the form of a trade secret is not just an easy process. It comes with both advantages and disadvantages.

Advantage of big data being protected as a trade secret:

Many companies like the KFC, Google, Coco Cola, etc. protect their big data in the form of a trade secret. These companies usually deal with huge amounts of data because they have a global acclamation, larger set of audience and use varied marketing strategies, which all requires a lot of data to be worked upon.

Now, say there is a company producing a particular product which is unique in the market. If the said company cannot acquire patent over this product, the alternative remedy available for the company is to get the product protected as a trade secret. So, this implies that what cannot be protected as a patent, maybe protected as a trade secret, i.e., to say that the individual data which cannot be protected as a patent, can be protected by the business as a trade secret.

Disadvantage of big data being protected as a trade secret:

Defining big data in terms of trade secret comes with certain limitations. A well-known case in this regard in the Canada's Federal Court is the case of *Doshi v. Canada.*[14] This case concerned section 21.1(3)(c)[15] of the *Food and Drugs Act* enacted pursuant to the *Protecting Canadians from Unsafe Drugs Act (Vanessa's Law).* In this context, *Dr. Peter Doshi*, an assistant professor at the University of Maryland applied to Health Canada to obtain unpublished information, which included clinical trial reports, related to Gardasil, Gardasil 9, and Cervarix in one application and Tamiflu and Relenza in another. Health Canada replied that in order to release this information to Dr. Doshi, he would be required to sign a confidentiality agreement, and a declaration of conflict of interest. Dr. Doshi refused to sign

---

[14] (Attorney General), 2018 FC 710.

[15] *21.1 (3) The Minister may disclose confidential business information about a therapeutic product without notifying the person to whose business or affairs the information relates or obtaining their consent, if the purpose of the disclosure is related to the protection or promotion of human health or the safety of the public and the disclosure is to: (c) a person who carries out functions relating to the protection or promotion of human health or the safety of the public.*

the confidentiality agreement. Consequently, his request for the clinical trial reports was rejected. Health Canada stated in its reasons for rejecting the application that it is Health Canada's policy to treat all regulatory data as (CBI). The Court drawing a distinction between language in *TRIPS* and *NAFTA* that differentiated between "undisclosed information," "trade secrets," and "data," *Justice Grammond* found that the data should be shared in light of the public interest exception. The dispute presented an interesting example of how the defense of trade secrets does not hold up against the needs and demands of the public interest.

Thus, although a business is protecting its big data in the form of a trade secret, it has been disclosed if it demands some sort of a public interest.

Moreover, even the big data is in the form of a trade secret, it cannot escape the escape the access of competitors by illegitimate means like bribery of the company's men itself, spying, hacking or even data theft.

Conclusion

Both IP and big data are emerging fields. The legislature is trying its level best to bring the protection of big data under the umbrella of IP but this process does not seem to be as simple as that. Protecting big data as an IP is bringing up with a lot of limitations and concerns which require immediate attention. The biggest responsibility is now in the hands of the lawmakers to encourage more innovations in the big data analytics but at the same time, protect a creator's work. A new outlook with respect to the IP and big data has to be given for a better implementation of the rights available to a creator from any unauthorized use and distribution of his data.

****