PROBLEMS OF PERSONAL DATA PROTECTION WHEN USING BIG DATA TECHNOLOGIES

Varvara Vladimirovna Bogdan

Southwest State University, Department of civil law, Kursk, Russian Federation

Elena Anatolyevna Kirillova

Southwest State University, Department of civil law, Kursk, Russian Federation

Cite article:

Bogdan V. V., & Kirillova, A. E. [2020]. Problems of personal data protection when using Big Data technologies. *Journal of Applied Engineering Science*, 18(3) 438 - 442.

Online aceess *of full paper is available at: www.engineeringscience.rs/browse-issues*

PROBLEMS OF PERSONAL DATA PROTECTION WHEN USING BIG DATA TECHNOLOGIES

Varvara Vladimirovna Bogdan*, Elena Anatolyevna Kirillova
Southwest State University, Department of civil law, Kursk, Russian Federation

The study analyzes Big Data technologies and considers the role, features and significance of these technologies. The relevance of the research is important because of the large-scale use of Big Data technologies in many areas and the weak legal regulation of the employment of Big Data when using personal data. The purpose of this study is to determine the legal status of Big Data technology, differentiate the concepts of "personal data" and "Big Data technology" and offer practical recommendations on the use of this technology. The methodological basis of the research is legal, statistical, private scientific methods, as well as the methods of case analysis and "snowball". Based on the results of the study, the authors define the "Big Data" technology and "personal data in electronic form". The principles for using Big Data technologies that can be applied in the international legal acts and judicial practice were developed. They are: the legality of collecting personal data in electronic form; electronic data processing in limited ways; prohibition of combining various databases with personal information in electronic form; limited period of storage of personal data in electronic form; free use of information; commercial usage. It is proposed to classify personal information provided electronically on the Internet by citizens into the following categories: personal information open to public; special information on actions; biometric data; confidential personal information.

Key words: Big Data technologies, information, personal data, electronic form, privacy

INTRODUCTION

In modern society, information is the main strategic resource and has the same value as material objects. Some experts rightly remark that the value of information can be compared to raw materials, which bring huge profits to the budgets of many countries [1]. At this stage of the society's development processing a large amount of information is an urgent problem, so it is interesting to analyze"Big Data" technologies that are successfully used in insurance, banking, marketing and other areas.

The use of Big Data technologies allows all parties of legal relationship to engage substantively, since any subject receives information automatically and individually, in addition, it is necessary to take into account the reduction in costs when applying this technology.

The term Big Data was first used in 2008 in an article in the journal Nature, which noted the need to use special technologies to work with a large amount of data. The subject under consideration was relevant in connection with the development of technologies that could solve this problem [2]. Two years later, the first programs were released that worked with a large amount of data and at the same time could minimize risks. Such companies as Microsoft, Oracle, EMC, IBM began to make use of the first developments in the field of big data and take this fact into account in their strategies.

Many researchers considered big data technologies. Thus Song Ying, Stavros Sindakis, Sakshi Aggarwal, Charles Chen, Jiafu Su (2020) devoted their work to the problems of big data management in Singapore retail paying attention to the legal regulation of Big Data technologies; Tahiru Azaaviele Liedong, Tazeeb Rajwani,

Thomas C. Lawton (2020) investigated the conceptual relationships between big data and corporate political activism; Nathaniel D. Line, Tarik Dogru, Dahlia El-Manstrly, Alex Buoye, Jay Kandampully (2020) analyzed the control, usage and ownership of big data in the tourism industry; Vagelis Papakonstantinou, Paul de Hert (2020) dedicated their work to big data analytics in electronic communications, at the same time they noted the need for legislative regulation of the use of Big Data. Many other scientists, experts and researchers have considered big data technologies, but the legal status of Big Data using personal data has not yet been determined. In addition, the legislators raise questions about the need for an effective protection of personal data if they are used in this technology; there are other conflicts that need to be considered. The purpose of this study is to determine the legal position of Big Data technology, differentiate the concepts of "personal data" and "Big Data technologies" and offer practical recommendations for the use of this technology.

METHODOLOGY

The bulk of the study is based on the use of comparative legal means due to the diversity and inconsistency of the use of Big Data technologies in different countries of Western Europe and the USA. The authors defined "personal data" and "Big Data technologies" using formal legal tools and it was identified that the legislator had been indifferent to this issue. That was expressed in inefficient legal regulation of the application and use of Big Data technologies.

When writing the article, more than twelve regulatory le-

gal acts in the field of personal data protection and a number of advisory documents were analyzed. The legal doctrine was also investigated which reflects the scientists and lawyers' opinions on the regularities of the application conditions and areas of activity in which Big Data technologies are used. The authors also studied the fundamental guidelines that were in operation in this area, which should serve as a guide for experts and lawyers in law enforcement practice related to the use of Big Data technologies.

The dialectical method of cognition was used while conducting research, which made it possible to formulate the principles of the employment of Big Data technologies and determine the prospects for the legal regulation of these technologies. The theological method allowed us to consider the application of the confidentiality regime to both Big Data technologies and personal data and differentiate these concepts making use of the identified features.

The logical method made it possible to systematize the methods of using Big Data technologies. This method was supplemented by a formally legal method, which allowed to divide personal information into categories.

The implementation of the system and structural approach made it possible not to equate Big Data technologies with personal data because when using Big Data, a large amount of information is used that cannot be defined as personal information about a subject.

RESULTS

There is no generally accepted definition of Big Data. This technology is disclosed through the definition of its main features, which include: processing a large amount of data; the variety of data processed; constantly changing the data being processed. Given these characteristics, Big Data definition can be defined as follows: it is a technology for processing, storing, and integrating structured and unstructured ever-changing large amounts of data. Big Data technology allows to structure diverse information from different sources and create new information and deliver it to the necessary subjects.

Big Data technology is fundamentally different from block-chain technology and the main difference is that in the distributed block-chain technologies information is not integrated, but it is stored in blocks in the information chain. Thus ensures an effective protection of information, since each copy is stored on the different computers; such technology is decentralized and characterized by the absence of intermediaries. Big Data technology, on the contrary, integrates information from different sources. It also has advantages in the speed of processing information. As for the distributed block-chain technology it cannot provide such fast data processing [3].

To create Big Data technologies, various techniques are used. We can distinguish the main ones among them and represent them in the table (Figure 1. Techniques for creating Big Data technologies):
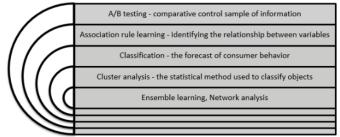

Figure 1: Techniques for creating Big Data technologies

Information sources can be classified into two types: external, when information is obtained directly from the Internet sources, and internal, which include various classifiers: ERP, CRM

Big Data technologies can be used in various areas of activity. The most promising of them are sales offices. With the help of this technology they can predict demand which depends on many factors, such as seasons, interests, fashion trends, price policy, etc. [4].

Online stores often increase sales by expanding the volume of their buyers using data about potential clients from social networks, from browser requests, from mobile phones. Then the data is processed. The technology of Big Data allows to investigate the behavior of the buyer on the store website, taking into account such parameters as time spent on the site, frequency of visits, types of goods which have attracted the attention of potential buyers. Based on the collected data a strategy for attracting buyers is developed [5]. In particular, the advertising companies are created that can use retargeting when a buyer searches for a specific product on a smartphone. After that the product information appears on all devices or IP address data recorded in certain places, for example, on sports grounds, can be used to sell sports equipment to a specific person whose address was in the Big Data database and was used in personalized advertising [6].

Big Data technologies are widely used in the banking segment in order to assess the solvency of the client, the presence of poor credit histories, existing debts, and the presence of a criminal record. In Western Europe for a long time, specialists have been used to collect this information using all available resources, including the Internet. Then the data were manually classified and provided to banks. Now Big Data technologies allow to do this work without unnecessary costs in a short time [7]. Thus, banks using Big Data technologies reduce the risk of providing a loan to insolvent customers, reduce the collection of necessary documents and simplify the procedure for concluding a loan agreement.

Big Data technologies can solve the problem of actuarial calculations in insurance companies. The actuarial calculations are used for calculating insurance payments. For this purpose a complex system of mathematical methods is used which take into account the monetary resources of insurance companies and the degree of risks under which insurers will pay premiums to custom
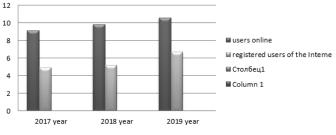
ers. The use of Big Data technologies in the insurance industry is quite promising, since only this technology is able to analyze a large amount of information and give a forecast for the degree of risk [8]. In such cases, various circumstances can be taken into account, for example, when insuring a car, the risks of theft or accident can be calculated, and Big Data technology takes into account many factors, from the identity of the owner to the quality of roads on which the car owner is likely to drive. Such insurance is already used in Western Europe and the United States and is defined as " usage-based insurance" [9].

When insuring life and health insurance companies can use data from "smart wristbands" that take into account motor activity, sleep quality, measure heart rate, pressure, number of steps. The processing of this information using Big Data technologies can predict the quality of life and health of a potential client. Some insurers can offer the discounts and special insurance proposal based on positive data about the client's health.

The employment of Big Data technologies is unlimited. These technologies can be used in almost any field of activity. And the companies, firms, and enterprises which use Big Data technologies will have advantages over other organizations.

DISCUSSION

At the present stage of the social development, the amount of information used by both citizens and organizations on the Internet has increased a hundred fold. At the same time, the personal information often gets into the public access, where can be found not only the surname, name, patronymic, but the information about the place of residence, phone number, bank card, and other personal data. The volume of personal information that gets into the Internet according to experts has recently tripled. On the example of Russia, we can say that there is 5% growth of the Internet users in 2019 compared to 2018 [10], the data are presented in the graph (Fig 2. Growth of the Internet users in Russia):

Figure 2: Growth of the Internet users in Russia

Many disputes among experts cause problems of personal data protection when using Big Data technologies [11], since the employment of this technology affects the personal data of hundreds of thousands of people and can violate the right to privacy of personal information.

When using personal data offline, such data are guaranteed to be protected, but it is problematic to guarantee the protection of personal information on the Internet. The experts notify that the use of Big Data technologies is incompatible with the protection of personal information [12]. However, the information with which Big Data technologies work can be divided into personalized and impersonal, when the information is generalized and is not tied to personal data.

Let's look at how the concept of "personal data" is defined, which is enshrined in the international legal acts and is subject to protection. Personal data is personalized information about a particular subject that can be used to identify it. Such information may contain facts about a physical, genetic, social and other identity of the subject. When forming a definition of personal data online, you can add an IP address to these facts, provided that there is a clear time link to one particular person.

A prerequisite for assigning (or not assigning) an IP address to personal data is the ability to bind a specific IP address to a specific individual or legal entity for a specific extended time period. A citizen can get a personal IP address to access the Internet, but he can also get an IP address that will be used by another person at the same time and it will no longer be personalized, therefore it cannot be attributed to personal information.

International legislation contains norms in accordance with which administrative liability is provided for violation of the use of client personal data. In some countries of Western Europe the amount of fines is calculated in large amounts [13]. However, the client data in electronic format which was obtained by the related parties on the Internet is often used but the confidentiality is not ensured. The clients themselves providing their data do not think about the consequences. The principle of informing the client about the use of their personal data is observed nominally. The site contains an agreement on a privacy policy, which is signed by customers, as a rule, without even reading, thus they consent to the use, exchange, sale of personal information. In addition, citizens purchase SIM cards, receive bonus cards in stores, sign up for an account on social networks, and in such a way the personal information goes to interested persons who use them for their business purposes. People themselves sometimes are not aware of how they leave "digital traces" that is how many experts began to call personal information in electronic form. Except personal information the related companies can also receive information about movements, transactions and other actions in which personal data is not used but other identifiers are used in which business services are interested.

 Cliqz and Ghostery provided data according to which more than 80% of sites are collecting information about user behavior on their and third-party resources in order to provide data about potentially interested persons, companies, organizations that will use them to attract new customers [14]. For example, mobile operators are involved in the distribution of targeted advertising, that is, such advertising that takes into account the interests,

needs of a certain contingent and, therefore, they are more likely to purchase certain categories of goods or services. It is worth noting that information about the user behavior on the Internet at the legislative level is not protected by legal norms and is not subject to the personal data regime. Thus, personal information in electronic form cannot be protected at the proper level. Even a technical solution to depersonalise personal data presented on the Internet cannot guarantee the complete confidentiality of such information, since the depersonalization process can be reversed. Therefore, personal data in electronic form obtained on the Internet and included in the information stream of Big Data technologies can be freely used.

Personal data can be divided into the following categories (Figure 3. Types of personal data):
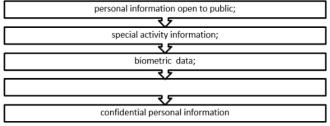

personal information open to public;

special activity information;

biometric data;

confidential personal information

Figure 3: Types of personal data

The legal regulation of the employment of Big Data technologies is based on the main provisions of the Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data and the 2002/58/ EU Directive of 31.06.2002 in respect of personal data processing and privacy protection in the electronic communication sector. But in practice, Big Data technologies should not be equated with personal data, since when using Big Data a large amount of information is used that cannot be defined as personal information about a subject. In general it is depersonalized information. It should be borne in mind that Big Data is a digital asset, since the received and processed data is sold to the relevant companies and the sale of impersonal data is a significant source of profit. But the legislation does not take into account this fact, so it is advisable to consolidate the possibility of commercial use of impersonal data obtained using Big Data technologies.

Thus, the use of Big Data technologies is based on certain principles, which are presented in Figure 4. Principles of using Big Data):
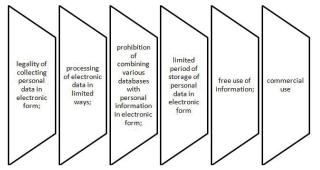

legality of collecting personal data in electronic form;

processing of electronic data in limited ways;

prohibition of combining various databases with personal information in electronic form;

limited period of storage of personal data in electronic form

free use of information;

commercial use

Figure 4: Principles of using Big Data

To enshrine the principles of using Big Data using personal data in the legislation will allow to introduce these technologies into the legal framework. At the same time it is necessary to pay attention to the fact that the data is provided by the subjects and not only their subsequent use. According to statistics, 65% of the people surveyed are ready to share anonymized data, while more than 70% are ready to provide minimal personal information on the Internet [15]. Therefore, the legislative regulation of Big Data technologies can help companies to comply with legal norms in exchange for a more liberal procedure for collecting and storing personal information in electronic form.

CONCLUSION

Big Data is a technology to process, store and integrate structured and unstructured, ever-changing, high-volume data. Big Data technology allows you to structure diverse information from different sources. It also permits to create new information and convey it to the necessary subjects.

Personal data in electronic form is personalized information about a particular subject which can be used to identify this subject. Such information may contain facts about the physical, genetic, social and other identity of the subject, as well as an IP address, provided that there is a clear time link to one particular person.

The study developed the usage principles of Big Data technologies, which can be applied in the international legal acts and judicial practice, these include: the legality of collecting personal data in electronic form; electronic data processing in limited ways; prohibition of combining various databases with personal information in electronic form; limited period of storage of personal data in electronic form; free use of information; commercial use.

It is proposed to classify personal information provided electronically on the Internet by citizens into the following categories: personal information open to public; special information concerned client behavior; biometric data; confidential personal information.

For further studies the consideration should be given to legal consolidation of "Big Data technologies" definition in international instruments and investigate legal conflicts in existing legal norms.

REFERENCES

1. Song Ying, Stavros Sindakis, Sakshi Aggarwal, Charles Chen, Jiafu Su (2020) Managing big data in the retail industry of Singapore: Examining the impact on customer satisfaction and organizational performance. European Management JournalIn press, corrected proof Available online 4 April 2020.

2. Tahiru Azaaviele Liedong, Tazeeb Rajwani, Thomas C. Lawton (2020) Information and nonmarket strategy: Conceptualizing the interrelationship between big data and corporate political activity. Technological Forecasting and Social Change Volume 157. August 2020.

3. Vichugova A. Big Data and blockchain: combo or confrontation? April 7, 2019 https://www.bigdataschool.ru/bigdata/big-data-%D0%B8-%D0%B1%D0%BB%D0%BE%D0%BA%D1%87%D0%B5%D0%B9%D0%BD.html

4. Longfei He, Mei Xue, Bin Gu (2020) Internet-of-things enabled supply chain planning and coordination with big data services: Certain theoretic implications. Journal of Management Science and EngineeringIn press, corrected proof Available online 29 March 2020.

5. Nathaniel D. Line, Tarik Dogru, Dahlia El-Manstrly, Alex Buoye, Jay Kandampully (2020) Control, use and ownership of big data: A reciprocal view of customer big data value in the hospitality and tourism industry. Tourism Management. Volume 80. October 2020.

6. Hengyun Li, Mingming Hu, Gang Li (2020) Forecasting tourism demand with multisource big data. Annals of Tourism Research. Volume 83.July 2020.

7. Anne Gregory, Gregor Halff (2020) The damage done by big data-driven public relations. Public Relations Review Volume 46. Issue 2. June 2020.

8. Heli Hallikainen, Emma Savimäki, Tommi Laukkanen (2020) Fostering B2B sales with customer big data analytics. Industrial Marketing Management. Volume 86. April 2020. Pages 90-98.

9. Karim Moharm (2019) State of the art in big data applications in microgrid: A review Advanced Engineering Informatics. Volume 42. October 2019.

10. Nelson Lozada, Jose Arias-Pérez, Geovanny Perdomo-Charry (2019) Big data analytics capability and co-innovation: An empirical study Heliyon. Volume 5. Issue 10. October 2019.

11. Kirillova, E.A., Pavlyuk, A.V., Mikheyev, A.A. Online contractual process: Status and technology. 2019 International Journal of Recent Technology and Engineering. 8(1) 2234-2240.

12. Xin Li, Rob Law (2020) Network analysis of big data research in tourism. Tourism Management Perspectives. Volume 33. January 2020.

13. Vagelis Papakonstantinou, Paul de Hert (2020) Big data analytics in electronic communications: A reality in need of granular regulation (even if this includes an interim period of no regulation at all). Computer Law & Security Review. Volume 36. April 2020.

14. Ajay Major, Suzanne M. Cox, Samuel L. Volchenboum (2020) Using big data in pediatric oncology: Current applications and future directions. Seminars in Oncology. Volume 47. Issue 1 February 2020. Pages 56-64.

15. Robert Northcott (2019) Big data and prediction: Four case studies. Studies in History and Philosophy of Science Part AIn press, corrected proofAvailable online 12 September 2019.