

AMURCON 2021
AmurCon 2021: International Scientific Conference**BIG DATA: THEORETICAL AND PRACTICAL
UNDERSTANDING**Daria A. Petrova (a)*, Galina M. Radyanskaya (b)
*Corresponding author(a) Far Eastern Federal University, FEFU Campus, 10 Ajax Bay, Russky Island, Vladivostok, Russia,
petrova.dan@dvfu.ru(b) Far Eastern Federal University, FEFU Campus, 10 Ajax Bay, Russky Island, Vladivostok, Russia,
radianskaia.gm@students.dvfu.ru**Abstract**

Big Data is a new unique and promising object of legal relations, which has significant advantages. The developed mechanisms for processing Big Data are essential for healthcare, insurance, law enforcement, public administration, economic development, and many other areas. However, the practical application of Big Data is fraught with some difficulties. In this article, the authors conclude that there are at least three such difficulties. The first problem is the lack of a unified approach to the definition of Big Data. The number of opinions on this topic is increasing every day; scholars are attempting to eliminate the existing theoretical gap and form doctrinal answers to many questions. Therefore, the authors set the task to highlight the doctrinal approaches to the definition of BD. The second problem is that there is no legal approach to Big Data in any country yet. Regarding Russia, the authors express the point of view that, in addition to working out some standards, the Russian legislature will have to carry out large-scale work in two directions: streamlining the application of current legislation and creating specialized norms for relations connected to Big Data.

2357-1330 © 2022 Published by European Publisher.

Keywords: Big data, database, array of information, personal data

1. Introduction

It is difficult not to agree with the idea that “we are already living in a new era — the era of big data” (Mayer-Schönberger & Cukier, 2014). In the XXI century, the age of information technology, the Internet has become an integral part of society. People’s interest in the available information and the desire to share the accumulated knowledge led to the creation of an enormous array of information. Today, a person can find an answer to any of his requests. At the same time, specialists can track his search queries and use this information against him: actions in social networks, online stores, or even the order of viewing sites are analyzed by specific algorithms, resulting in an ideal offer for a potential client of the organization, a conclusion about the location, a person’s preferences and much more. This phenomenon is called Big Data.

The history of the appearance of the term itself is intriguing. There is no exact date of the database “birthday,” and the author of the name has not been determined. Some believe that the term “Big Data” was introduced on September 3, 2008, by the editor of the scientific journal “Nature” Clifford Lynch in a special issue (Guseva, 2016). However, there are some alternative opinions regarding this fact. It was indicated in 2005 when this term was first introduced into scientific circulation (Picciotto, 2019). Also, Korneev (2018), describing the history of Big Data development, claims that earlier specialists were developing this topic. For example, John Mashey, a computer scientist, popularized the term Big Data back in the 1990s. The author also found evidence that the notion similar to Big Data in meaning was mentioned in the Oxford English Dictionary of 1941.

The possibilities of Big Data are great. For example, Google created a system that analyzed the most popular search queries to find new outbreaks of H1N1 influenza in 2009 (Guseva, 2016). This system was based on the data processing of the epidemiological situation in the world for 2007-2008 in conjunction with Google’s data. In this way, specialists could create a new tool for detecting outbreaks of epidemiological diseases. Its results coincided by 97% with the actual detection. The application areas of Big Data technologies are not limited to the medical field. Savelyev (2015) identifies many other spheres where Big Data is used:

- 1) Banking sector (the solvency analysis of a potential borrower);
- 2) Insurance activity (insurance risks assessment identified by insurance companies by analyzing the probability of an insured event);
- 3) E-commerce (through the study of the users’ behaviour in online stores, goods that can interest a potential buyer are identified);
- 4) Prevention of offences (the Blue CRUSH system, based on crime statistics, can identify areas of the possible threat of committing an illegal act and provide this information to the police), and the like.

Bulgakova et al. (2015) separately examined the application of Big Data technologies in the public administration system. In their opinion, the potential of their application has not yet been revealed in the Russian Federation fully, which is a fault since BD technologies have impressive advantages. Firstly, the system can analyze the prospects for social services development in advance when providing state and municipal services to citizens and solve problems that arise in the shortest possible time. Secondly, the transition of public administration to an automated mode will help eliminate problems associated with

violating rules of services provided to citizens, corruption, and the like. Thirdly, it is assumed that the introduction of Big Data technologies will favorably contribute to manufacturing, trade, healthcare, public administration development, and many other areas.

2. Problem Statement

The topic of “Big Data” is currently debatable and open. Nowadays, we can talk about at least three unresolved problems in the functioning of Big Data. The first problem is the lack of a generally accepted definition at the legislative and doctrinal levels. The second problem is expressed in the fact that Big Data has no special normative regulation. (Therefore, self-regulation of legal relations is often present in the industrial sphere. For example, in the United States, the Federal Trade Commission stimulated companies to work out privacy rules independently (Sosnin, 2019)). And the third one can be defined as the problem of separating Big Data from personal data and protecting the latter.

3. Research Questions

This article examines the following issues:

1. Definition of Big Data in domestic and foreign doctrine;
2. Coverage of approaches to the regulatory regulation of the phenomenon of Big Data;
3. Correlation problem research of Big Data and personal information in the legislation of Russia.

4. Purpose of the Study

The purpose of the work is to highlight the theoretical and practical aspects of the functioning of Big Data.

5. Research Methods

The authors used a set of methods and information processing analytical techniques. The main research methods are analysis, synthesis, induction, and comparative legal analysis.

6. Findings

The topic of “Big Data” is currently debatable and open. At the moment, we can talk about at least three unresolved problems in the functioning of Big Data. The first problem is the lack of a generally accepted definition at the legislative and doctrinal levels. The second problem is expressed in the fact that Big Data has no specific normative regulation. (Therefore, self-regulation of legal relations is often present in the industrial sphere. For example, in the United States, the Federal Trade Commission stimulated companies to work out privacy rules independently (Sosnin, 2019)). And the third one can be defined as the problem of separating Big Data from personal data and protecting the latter. Next, we will consider each thesis in more detail.

If we turn to the first thesis on the definition of “Big Data,” we can see three dominant approaches in the scientific literature to understanding the object under study. 1) enormous array of information; 2) methods and techniques of information processing; 3) a set of massive information and its processing methods.

The first approach is that Big Data is a data array that represents substantial amounts of information with a complex heterogeneous or indefinite structure, characterized by broad fields of application and a decentralized way of storing data. The authors of this approach include Volkova (2016), Gorodov and Egorova (2018).

Several authors Savelyev (2015), Tolstova (2015), Bulgakova et al. (2015), and others adhere to the definition of Big Data as a method of enormous information processing. A. I. Savelyev writes about Big Data: “a set of tools and methods for processing structured and unstructured data of huge volumes from various sources, subject to constant updates, to improve the quality of managerial decision-making, create new products and increase competitiveness” (Savelyev, 2015, p. 60). The author also quotes a brief formulation created by the consulting company Forrester: “Big Data combines techniques and technologies that extract meaning from data at the extreme limit of practicality” (Savelyev, 2015, p. 61).

Tolstova (2015) believes that Big Data is a technology that allows searching, forming, and analyzing arrays big data arrays, both structured and unstructured. Protasov (2015) shares a similar opinion. He points out that big data is a set of technologies designed to perform three operations: process more data than usual; work with rapidly arriving data in large volumes, and work with structured and poorly structured data parallelly in different aspects.

In her work, Guseva (2016) reflected the third approach to the definition. Based on the analysis of domestic and foreign scientists’ works, derived the following concept: “Big Data is a complex concept that combines: 1) the data itself (massive encoded information); 2) a set of technologies for working with this data; 3) a new look, a new paradigm in data science” (p. 321)

The latter approach, in our opinion, seems to be more accurate because its consolidates the previous methods and most fully reflects the content of Big Data.

So, in his work “The History of the concept of Big Data: dictionaries, scientific and business periodicals,” Korneev (2018), referring to the research of Doug Laney for Meta Group made in 2001, mentions the principal parameters of the three “V”: Volume, Velocity, Variety. In some works (Tsarkova & Smolyanov, 2016; Xu et al., 2015), scholars add additional criteria: Veracity and Value. According to these parameters, Big Data is characterized by the massive size of information, high speeds of data collection and analysis, heterogeneity, and lack of structuring of accepted materials, together with the specific value of reliable information and the cost justification of its processing. These are widely used in foreign authors’ works, for example, Aryabhata University Professor Kumar (2015), Yuan University researchers (Chen et al., 2006), as well as other research works (Kasemsap, 2016; Picciotto, 2019).

One of the most significant gaps in Big Data is the lack of specific legislative norms. These apply to both Russian and international legislation. The legislation of the USA, one of the most advanced states in the field of digital technologies in the world, does not provide for the regulatory consolidation of the term “Big Data” at the federal level. There is no single centralized approach to the protection of Big Data, and there is no regulatory authority in this area.

For the countries of the European Union, the Pan-European Regulation on Personal Data Protection has been in effect since May 25, 2018. There is no separate regulation of Big Data in this document. This category is not provided separately, but it is regulated by the norms intended for personal data due to a specific approach. They are included in personal data due to the broad interpretation of databases. The regulation also provides for the European Data Protection Council creation, including the Personal Data Protection Inspector and the heads of the supervisory authorities of each EU Member State. The norms of this document are Extraterritorial in nature, which allows other states to use these norms when interacting with EU members and rely on them, forming their legislative regulation of this area.

In Russian legislation, the term “Big Data” first appeared as a type of “end-to-end digital technology” in the program “Digital Economy of the Russian Federation” (Rasporiazhenie..., 2017), approved by the decree of the Government of the Russian Federation dated July 28, 2017. In the Order of the Ministry of Education and Science of the Russian Federation dated 20.09.2017 No. R-603 “On approval of the rules for organizing access to scientific and scientific-technical information in the Russian Federation” Big Data is found in the title of section 4: “Access to databases of primary (factual) scientific data, including “Big Data,” their processing and obtaining new knowledge” (Rasporiazhenie..., 2017).

The mentioned above documents only consider Big Data. Normative documents do not contain definitions or any regulations governing them. The legislature sees Big Data as a new object of the legal relation. Detailed legislation of this area will begin shortly. Now there are no distinctions between databases and similar concepts of “big data arrays,” “large amounts of data,” “large user data,” and the like in the subordinate regulations (Belaya et al., 2018).

Since at the moment, one of the system-forming laws for regulating areas related to Big Data is Federal Law No. 152-FZ of 27.07.2006 “On Personal Data,” the question arises about the differentiation of Big Data and Personal Data in Russian legislation. Since there is no legal definition of Big Data in Russian legislation, several approaches to these concepts correlation have appeared in legal theory.

According to the first approach, Big Data are components of personal data (and there is no boundary between them). During the accumulation of information from social sources, it is likely to collect private information about a person that may constitute personal data. Rozhkova (2019) follows this approach “Quite often in the literature, there are statements about the fine line or insufficient clarity of the distinction between *big data* and *personal data*. There is no such boundary/distinction at all.”

The second approach consists of the significant differences between Big and personal Data, so these are different concepts. Savelyev (2015) adheres to this approach. He claims that the capabilities of BD technologies directly contradict the principles of personal data legislation, and we cannot apply these norms to this area. After analyzing the principles of personal data protection and processing, he concluded that Big Data is incompatible with:

- The principle of limiting the personal data processing to predetermined purposes (since it is impossible to predict the purpose of using the information);
- The concept of informed, definite and conscious consent as the principal basis for legitimizing the processing of personal data (the database cannot provide an exhaustive scope of methods

and purposes of information processing, its will not be possible to interact individually with organizations processing information, and the like);

- Depersonalization of information cannot guarantee anonymity. (There is too much volume for depersonalization, and, in most cases, companies do not need a name to personalize a person).

In theory, there may be other logical and reasoned opinions. We can form the most suitable approach after the issue of the ratio of Big and personal Data legislative settlement.

In addition to these significant gaps, there are several unresolved issues. So Big Data also comes with some risks. As a result of using the database, information leaks are possible. Specialists cannot guarantee the use of this information for other purposes and the security of user data. There is a risk of processing inaccurate or erroneous information, which can lead to various consequences: starting from annoying advertising, ending with incorrect diagnoses of a medical analytical program, or refusal to grant a loan, and a citizen will not be able to find out about it on his own.

7. Conclusion

Thus, Big Data is a new unique, and promising object of legal relations, which has a significant number of advantages. The developed mechanisms for processing Big Data are essential for healthcare, insurance, law enforcement, public administration, economic development, and many other areas. As you can see, there is no single approach to BD issues in science. However, the number of opinions on this topic is increasing every day; scholars are attempting to eliminate the existing theoretical gap and form doctrinal answers to many questions.

The tasks assigned to the States to settle existing BD relations require an innovative approach to the interpretation and application of legislation. There is no legal approach to Big Data in any country yet, but work is underway on this. And we can assume that specialists will settle all unresolved issues soon. As for the Russian legislation, it will have to carry out large-scale work in two directions. They are streamlining the current legislation application and the creation of specialized norms for relations connected to Big Data.

Acknowledgments

This work was financially supported by the Grant of the President of the Russian Federation No. NSh-2668-2020.6 “National-cultural and digital trends in the socio-economic and political-legal development of the Russian Federation in the XXI century.

References

- Belaya, O. V., Kononenko, D. B., Semchenkova, M. N. (2018). Pravovoe regulirovanie deiatelnosti startupov v oblasti Big Data [Legal regulation of startups in the field of Big Data]. *Business. Education. Right*, 1(42), 174-179.
- Bulgakova, E. V., Bulgakov, V. G., & Akimov, V. S. (2015). Ispolzovanie bolshikh dannykh v sisteme gosudarstvennogo upravleniia usloviia vozmozhnosti perspektivy [The use of "Big data" in the system of public administration: conditions, opportunities, prospects]. *Legal science and practice: Bulletin of the Nizhny Novgorod Academy of the Ministry of Internal Affairs of Russia*, 3, 10-14.

- Chen, J., Chen, Y., Du, H. Li, S., Lu, J., Zhao, S., & Zhou, H. (2006). The Big Data problem: a data management perspective. *Frontiers of Computer Science*, 7(2), 157-164. <https://doi.org/10.1007/s11704-013-3903-7>
- Gorodov, O. A., & Egorova, M. A. (2018). Osnovnye napravleniia sovershenstvovaniia pravovogo regulirovaniia v sfere tsifrovoi ekonomiki v Rossii [The main directions of improving legal regulation in the field of digital economy in Russia]. *Law and digital Economy*, 1, 6-12.
- Guseva, A. A. (2016). Bolshie: dannye poniatie, istochniki, vozmozhnosti [Big data: concept, sources, opportunities]. *Perm: master magazine*, 1, 320-324.
- Kasemsap, K. (2016). Mastering large amounts of data in the digital age. Effective management of big data and opportunities for implementation. *Book series "Advances in Data Mining and Database Management"*, 104-129. <https://doi.org/10.4018/978-1-5225-0182-4.ch008>
- Korneev, M. S. (2018). Istoriia poniatii «Bolshie dannye»: slovari, nauchnaia i delovaia periodika [The history of the concept of "Big Data": dictionaries, scientific and business periodicals]. *Bulletin of the Russian State University for the Humanities*, 1(34), 81-85.
- Kumar, N. (2015). The Development of New Industrialization in the Context of Big Data. *Annual Conference on the Development of New Industrialization and Urbanization: International forum on the development of new Industrialization in the era of Big Data*. <https://www.webofscience.com/wos/woscc/full-record/WOS:000380582500001>
- Mayer-Schönberger, V., & Cukier, K. (2014). *Big data: a revolution that will transform how we live, work, and think*. Harper Business.
- Picciotto, R. (2019). Evaluation and the big data problem. *American Journal of Evaluation*, 41(2), 166-18, <https://doi.org/10.1177/1098214019850334>
- Protasov, S. (2015). Chto takoe Big Data [What is Big Data?]. *PostNauka website*. <https://postnauka.ru/faq/46974>
- Rasporiazhenie Pravitelstva RF ot 28 07 2017 N 1632-r Ob utverzhdenii programmy TSifrovaia ekonomika Rossiiskoi Federatsii [Order of the Government of the Russian Federation dated 28.07.2017 N 1632-p «On approval of the program "Digital Economy of the Russian Federation"». (2017). http://www.consultant.ru/document/cons_doc_LAW_221756/f62ee45faefd8e2a11d6d88941ac66824f848bc2/
- Rozhkova, M. A. (2019). Chto takoe bolshie dannye, chem oni otlichaiutsia ot obychnykh dannykh i v chem sostoit problema pravovogo regulirovaniia Big Data [What is big data, how do they differ from ordinary data and what is the problem of legal regulation of big data]. (2019). *Zakon.ru*. https://zakon.ru/blog/2019/04/22/chto_takoe_bolshie_dannye_big_data_chem_oni_otlichayutsya_ot_obychnyh_dannyh_i_v_chem_sostoit_proble
- Savelyev, A. I. (2015). Problemy primeneniya zakonodatel'stva o personal'nykh dannykh v epokhu "bol'shikh dannykh" [Problems of application of legislation on personal data in the era of "big data"]. *Pravo. Journal of the Higher School of Economics*, 1, 43-66.
- Sosnin, K. A. (2019). Pravovoe regulirovanie Bolshikh dannykh zarubezhnyi i otechestvennyi opyt [Legal regulation of Big data: foreign and domestic experience]. *Journal of the Court of Intellectual Rights: network journal*, (25), 30-42. <http://ipcmagazine.ru/legal-issues/legal-regulation-of-big-data-foreign-and-domestic-experience>
- Tolstova, Yu. N. (2015). Sotsiologiia i kompiuternye tekhnologii [Sociology and computer technologies]. *Sociological research*, (8), 3-13.
- Tsarkova, N. I., & Smolyanov, A. S. (2016). Big Data Razvitie analiz i tekhnologiiia [Big data. Development, analysis and technology]. *Actual problems of humanities and natural sciences*, 7(1), 86-95.
- Volkova, Yu. S. (2016). Bolshie dannye v sovremennom mire [Big data in the modern world]. *Scientific and methodological electronic journal "Concept"*, (11), 1176-1180.
- Xu, H., Sheng, K., Zhang, L., Fan, Y.S., Dastdar, S. (2015). From Big Data to Big Service. *A computer*, 48(7), 80-83. <https://doi.org/10.1109/MC.2015.182>