GCPMED 2020
Global Challenges and Prospects of the Modern Economic Development

BIG DATA: LEGAL REGULATION AND ROLE IN THE DIGITALIZATION OF THE ECONOMY

E. K. Gubaydullina (a)*
*Corresponding author

(a) Samara State University of Economics, Soviet Army Str., 141, Samara, Russia, elmira_zaripova@mail.ru

Abstract

The daily growing volume of electronic data poses challenges to traditional ways of organizing data storage, processing, and analysis. The feasibility of further research in this area is confirmed by the high demand for data storage and analytical data processing services. The article is devoted to the study of big data security issues. Despite earlier technological capabilities, the term "Big data", or Big data, has only been actively used in the last decade. This term, originally coming from the exact sciences, programming, has caused and continues to cause a lot of controversy: what exactly is meant by Big data. In the course of the work, the specifics of Big data were described, the main characteristics of Big data were highlighted, and modern approaches to big data protection were analyzed. The role of Big data in the development of the digital economy and the need to address issues of their legal regulation are outlined. The following problems of Big data protection were formulated: limiting the speed of data access, organizing access via network protocols through General-purpose networks. A review of known developers of protection methods in these areas and recommendations for their application is made. Problems of legal regulation of Big data are revealed. The complexity of the problem, the need for multi-level protection and the use of modern methodological, theoretical and software developments from the first steps of working with Big data are shown.

Keywords: Big data, digitalization of the economy, legal regulation

1.   Introduction

One of the features of the modern stage of information processing is the continuous increase in the volume of processed and accumulated data. As well as permanent and dynamic complication of methods for their analysis, visualization, and protection. Despite a significant number of scientific papers on the formation and development of Big data, the problem of Big data security is unsolved and most relevant in the modern information technology industry. The situation is compounded by the fact that there are no currently standards for Big data protecting. Existing antivirus systems are not designed to provide the necessary level of security for such data. In addition, research in this area is not sufficiently systematic. In this regard, the issue of protecting large volumes of data requires further development and research and development in this direction are relevant.

2.   Problem Statement

The main task of the article is to identify the problems of Big data protection in modern conditions of globalization and informatization of socio-economic processes. Determining the level of impact of BigData protection issues on the development of the digital economy. Foreign researchers associate the definition of Big Date (Big data) with structured and unstructured data of large volumes and a significant variety of formats. At the same time, the processes of their formation, accumulation and processing are usually performed in on-line mode, thanks to the work of distributed systems and the use of cloud services (Johnson et al., 2019).

Examples of what may be the source of large amounts of data, include:

GPS signals for a transport company;

data from sensors of an industrial enterprise;

digital books in the electronic library;

Bank customer transactions;

information about products or purchases of a large retail chain, etc. (Shilina, 2018).

The main task of Big data methods, from the positions of domestic authors Poltavtseva et al. (2019), is to process huge amounts of data and build predictive models based on them, identify hidden connections and interactions. This is an urgent task that is paid attention to, for example, by such companies as Informatica (ETL-technologies for developing storage and data management); Hewlett Packard Enterprise; Imperva (development and production of products for protecting database management systems and web applications); Apache Software Foundation (Apache-open source software development technologies).

Big data security issues arise at all stages of working with data-during generation, transmission, accumulation, storage, analysis, and visualization. Well-known IT companies on the market pay significant attention to the development of specialized Big data security solutions. The largest contribution to the development of Big Data security technologies was made by the international alliance Cloud Security Alliance (CSA), the US National Institute of Standards and Technology (NIST) and the European Union Agency for Network and Information Security (European Union Agency for Network and Information Security, ENISA).

3. Research Questions

Along with the opportunities and prospects that are opening up when Big data is used, a number of issues of legal regulation of Such data in Russia arise. To date, Russian legislation does not have a clear, legally defined term Big data (Big Data), and there are no rules governing the rights and obligations of participants in the collection, storage and use of information. When working with Big data, it is impossible to determine whether there is a duty to notify the information subject of all the purposes, methods and types of information that will be collected and processed. In the Russian Federation, a draft law has been introduced to amend the Federal Law "On Information, Information Technologies and Information Protection," which proposes to regulate large amounts of data similar to personal data (Draft bill of 14 February 2020). In particular, it was proposed to supplement this law with Article 12.2 The operator of large user data, prior to the start of processing of large user data for the purposes of third parties, must obtain informed consent in electronic form from the user of the subscriber terminal. In turn, with Big Data, these requirements are virtually impossible to fulfil.

On February 14, 2020, the Ministry of Digital Development, Communications and Mass Media of the Russian Federation on the official electronic website "regulation.gov.ru" published a draft bill on the regulation of Big Data, which was an amendment to the Law "On Information, Information Technologies and Information Protection" (Draft bill of 14.02.2020). This draft bill was prepared in pursuance of the President's instructions following the message to the Federal Assembly on January 15, 2020.

The main task of the changes being made is to ensure regulation of the Big Data turnover while preserving human rights and freedoms when processing personal data. These amendments to the legislation introduced such terms as" Big data"," Big data operator "and "Big data processing", and also stipulated the principle of creating a register of Big data operators. However, this draft bill was rejected by the Russian Government on March 5, 2020. The reasons for the rejection of the draft bill were the fact that the amendments broadly define the concept of "Big Data", introduce excessive regulation for operators of such data, and also contain corruption risks.

Big data is usually stored and processed on distributed file systems. A distributed system is a set of interconnected autonomous computers or processors. It is a group of independent processes that interact by sending messages to exchange data and coordinate their actions. Distributed networks have recently reached a new stage of development-cloud networks that provide cloud computing. The cloud can mean both software and infrastructure. This can be an application that is accessible from the Internet or a server. In a distributed system, each process has its own state, represented by a data set that contains the current values of command counters, registers, and variables that the process has access to and can change. The state of each individual process is completely closed to other processes. The operation speeds of various processes in a distributed system are different and unknown in advance, and the delivery of sent messages may take an unpredictable period of time. In most cases, a distributed system contains several processors that are interconnected by means of communication (Tan et al., 2020).

4. Purpose of the Study

The purpose of the research is to identify problems in the legal regulation of Big Data. The essence of the development of the digital economy is not only in the emergence of new technologies, but also in the integration of existing ones into one system. So, in cloud computing, in IoT, in VR, in the field of information security, new technologies have appeared, which just made it possible to bring developments in 20-30 years into real production, to make them available for widespread use. Digitalization of the economy is impossible without the rapid processing of a huge amount of data supplied by thousands of sensors and smart devices. For example, only one Maersk Group vessel equipped with smart sensors transmits about 2 TB of data every day. The smart ship monitors weather conditions, engine operation, routes of neighboring vessels, and many other factors. Such careful monitoring can significantly increase the safety of navigation, automate some of the processes (for example, the declaration of goods in ports), and optimize business processes. But the collected amounts of information must be constantly processed and analyzed in real time. Big Data Analytics is effective in a wide variety of industries. That is why the issue of legal regulation of Big Data protection is currently the most relevant and requires an immediate solution.

5. Research Methods

The methodological basis of the research is primarily general scientific methods. As such, you can specify: dialectics, analysis, synthesis, induction, deduction, comparative legal and system-structural methods of analysis. Along with them, formal-legal and formal-logical methods were also widely used. In particular, the formal legal method was used to analyze the legal basis for the use of Big Data in Russia, research their development and prospects for improvement. The formal-logical approach helped to identify the link between the development of legal regulation of Big Data protection and the development of digital technologies in the economy. Comparative legal was applied in the study of foreign experience in data regulation, its comparison with Russian experience and identification of possible areas of exchange of experience. The work is generally carried out on the basis of a systematic approach to the study of the essence of legal categories, based on the principles of objectivity, comprehensiveness and complexity.

6. Findings

Approaches to organizing data protection depend on the architecture of computer systems and the capabilities of computer networks. In General, the main problems of big data protection are:

1. Insufficient performance and capacity of the storage methods and technologies used, partly due to the use of outdated equipment.

2. Low-speed tools, media, and technologies.

3. Lack of standardization of Big Data protection technologies and techniques (Priyanshu & Khan, 2020).

These problematic aspects lie on the surface and are well known to company managers in the field of it. However, many organizations still cannot effectively use Big Data because their outdated it infrastructure cannot provide the necessary storage capacity, data exchange processes, utilities, and

applications needed to protect the large fields of unstructured data that can be extracted from them. In addition, the increased processor performance required to protect ever-growing volumes of data may require significant investment in the organization's legacy it infrastructure, as well as additional support resources that can be used to develop new applications and services. Identifying Big Data threats and making recommendations to prevent them is one of the priorities of the European Union's network and information security Agency (Pearce, 2017).

The second important development is the review of the International Cloud Security Alliance (CSA) "The big data security and privacy handbook: 100 best practices in big data security and privacy". A review published in 2016 outlined 100 recommendations for securing Big Data (Cloud Security Alliance, 2016). This document describes the best practices tested in the Big Data market. The list includes both standard cybersecurity measures (such as authentication and access control) and modern cryptographic technologies. Each item in the document has two sections that specify why these security measures are necessary and how they can be implemented.

The third significant research concept in the field of big data protection is presented by the US National Institute of Standards and Technology (NIST) - experts have proposed the Interoperability Framework V1.0 (NIST, 2015) specification, which includes documents describing all aspects of working with big data: Big Data Definitions; Big Data Taxonomies; Big Data Use Cases and Requirements; Big Data Security and Privacy; Big Data Architecture White Paper Survey; Big Data Reference Architecture; Big Data Standards Road map. The document introduces the conceptual model of the Big Data architecture (NIST Big Data Reference Architecture, NBDRA), which is a Big Data system. The system includes five logical functional components. The unifying part is the Security and Privacy Managament, which focuses on issues of identity, authorization, auditing, and security of network devices.

In the most General form, foreign experts recommend focusing on four areas of big data protection:

1. The security of the infrastructure.

2. Ensuring data confidentiality.

3. Data management procedures.

4. Continuous security monitoring.

The main tools and technologies for working with Big Data are libraries for massively parallel processing of vaguely structured data, SQL database management systems, including NoSQL, MapReduce algorithms, Hadoop and MPI projects (Johnson et al., 2019). These tools are useful for processing Big Data, but NIST recommends adding several layers of protection to ensure security. One is for program protection, and the other is for data protection. At the same time, it is possible to use, for example, a special Kerberos Protocol that controls access to hadoop resources or other software products that implement role-based access functionality, for example, Apache Accumulo, Sentry, etc. (Johnson et al., 2019). The IT market offers fairly complete solutions for deployment, maintenance, or administration of large storage facilities and hardware and software solutions for them. A special feature of the best is the presence of built-in systems and security tools, technological solutions and a harmonious combination of software products used.

In the Russian market, services are offered by a branch of Dell EMC Corporation. The company's website contains a significant list of products that provide integration, storage, administration, and data

protection, as well as allow you to improve applications or develop your own programs for Big Data Analytics and protection. These include Isillon storage, a platform for deploying the ECS cloud with Hadoop support, high-performance servers, and integration solutions. Avamar 4.1, Cisco solution, InfoMover, and VNX are offered as security tools. The site also has an offer "Try", which is very appropriate at the first stages of working with Big Data.

IBM solution which consists of three products (How big data can enhance business strategies: Effective knowledge management is key, 2020) should also be mentioned:

Netezza: a specialized hardware and software package for building analytical applications and data warehouses;

Info Sphere Big Insights: a solution for analyzing and processing unstructured data based on Hadoop technologies;

Info Sphere Streams and Vivisimo: tools for analyzing current information and complex processing of large volumes of unstructured data.

Hitachi Data Systems offers are in great demand (Hitachi Data Systems, 2020), which include two specialized hardware and software complexes: a platform for storing and managing large volumes of unstructured data (Hitachi Content Platform, HCP) and solutions for providing file access to data, with saving and managing a large number of files (Hitachi Network Attached Storage (HNAS). To protect data storages, the authors of the projects propose the use of techniques for creating a SUNDR repository, the development of digests of certified messages; rotation of keys; or creating your own cloud storage.

One of the significant limitations of Big Data projects is the risks of information security, and especially the violation of the confidentiality and integrity of Big Data (Gurcan & Berigel, 2018). The data used for analysis often contains personal or up-to-date commercial information. And issues of ensuring integrity may concern both the analyzed data and the results obtained during their processing. Protecting and analyzing of Big Data requires the use of new computer graphics technologies, virtual environments, and augmented reality environments. Comprehensive research is necessary not only from the point of view of computer science and mathematics, but also from the point of view of cognitive psychology. Today's Big Data is unlikely to be a problem tomorrow, but in the future there is a possibility of a problem with more advanced information. This problem will remain unsolved until automated security systems are available that can adapt to the ever-growing volume of Big Data and current issues of their legal regulation are resolved.

7. Conclusion

The Big Data phenomenon has made many companies realize the need to collect, analyze, store and protect structured and unstructured data. However, implementing these processes requires an action plan and the right tools to optimize the processes. A lot of small and medium-sized businesses can't really get a tangible effect from Big Data. This problem is associated with the use of old databases and DBMS, which lack functionality and scalability. And Big Data for active use, to this day, does not have the proper level of legal protection of stored and processed information. In fairness, we recall that (as described above) the Russian legislator has repeatedly made attempts to solve the problem of legal regulation of the use of Big Data, but all the draft bills were not implemented and for a number of these reasons were rejected by the

Government of the Russian Federation. Thus, the issue of legal regulation of the use and protection of big data remains open and represents a prospect for further research in the field of protection of Big Data within the framework of the functioning of business entities.

References

Cloud Security Alliance (2016). The big data security and privacy handbook: 100 best practices in big data security and privacy. https://cloudsecurityalliance.org/artifacts/big-data-security-and-privacy-handbook/

Draft bill of 14.02.2020 «On amendments to the Federal Law «On information, information technologies and Information protection». https://regulation.gov.ru/projects#npa=99581

Gurcan, F., & Berigel, M. (2018). Real-time processing of big data streams: Lifecycle, tools, tasks, and challenges. In R. Polat & R. Goren, (Eds.), *Proceedings of the 2nd International Symposium on Multidisciplinary Studies and Innovative Technologies, Proceedings* (pp. 284-290). IEEE.

Hitachi Content Platform (2020). Modern approach to object storage. https://www.hitachivantara.com/ru-ru/products/storage/object-storage/content-platform.html

How big data can enhance business strategies: Effective knowledge management is key. (2020). *Strategic Direction, 36*(9), 41-44. https://doi.org/10.1108/SD-07-2020-0133

Johnson, S., Gray, P., & Sarker, S. (2019). Revisiting IS research practice in the era of Big Data. *Information and Organization, 29*(1), 41-56. https://doi.org/10.1016/j.infoandorg.2019.01.001

NIST (2015). NIST Special Publication 1500-1. NIST Big Data Interoperability Framework. https://bigdatawg.nist.gov/_uploadfiles/NIST.SP.1500-1.pdf

Pearce, H. (2017). Big data and the reform of the European data protection framework: An overview of potential concerns associated with proposals for risk management-based approaches to the concept of personal data. *Information and Communications Technology Law, 26*(3), 312-335. https://doi.org/10.1080/13600834.2017.1375237

Poltavtseva, M. A., Zegzhda, D. P., & Kalinin, M. O. (2019). Model of security threats to big data management systems. Problems of information security. *Computer Systems, 2*, 16-28.

Priyanshu, D., & Khan, R. L. (2020). Big data analytics - Pertaining technology 'vocative' to big data enhancing organisational capabilities. *SSRG International Journal of Engineering Trends and Technology, 68*(8), 46-52. https://doi.org/10.14445/22315381/IJETT-V68I8P209S

Shilina, N. V. (2018) What is Big Data? Main problems: Data storage and management. *Young Scientist, 28*(214), 1-3.

Tan, X., Xing, L., Cai, Z., & Wang, G. (2020). Analysis of production cycle-time distribution with a big-data approach. *Journal of Intelligent Manufacturing, 31*(8), 1889-1897. https://doi.org/10.1007/s10845-020-01544-7