

# **Big Data Usage by Private and Public Sectors: reality and prospects**

**Ilya Lavrov,**

School of Governance and Politics, MGIMO University

**Viktor Koveshnikov,**

Faculty of Economic Sciences, Higher School of Economics

## **Abstract**

This article presents a broad definition of big data, specifies its main characteristics and lists the steps required for its accurate usage. The primary focus of the paper is on the application of big data both in the private and public sector, specifying what has been achieved already and what may be implemented in the future. The paper also highlights the urgent need for the development of a methodological framework within which big data will have to be confined in the future.

**Keywords:** big data, private sector, public sector, smart city, policy analysis, public policy, transformation

## **Introduction**

Nowadays people consistently encounter enormous amounts of information and not only are these people receivers of data, they are also its producers, thus, massive flows of incoming data are generated at each second. In order to deal with these flows, the field of 'big data' emerged at the end of 2010s. Currently, the majority of companies as well as governments use big data methods to achieve their goals.

### **I. What is big data?**

Big data usually stands for a range of instruments, approaches and methods for working with both structured and unstructured data in order to use it for specific tasks and goals [2].

There are three main characteristics of big data which are traditionally distinguished [7]:

- **volume** - conventional data with which we work every day is calculated in gigabytes and terabytes, big data is stored in petabytes and zettabytes; as indicated by a report from Dell, the size of the digital universe doubles every two years and is projected to reach 44 zettabytes by the end of 2020;

- **velocity** – every process related to big data is conducted with extraordinary speed. Businesses, government and various organizations should be able to interpret data in real time, because this will provide an opportunity for decision-makers to act faster, thus, giving them an edge over competitors or in the case of government winning favor with citizens;

- **variety** - big data, as it is distributed in all sorts of formats (from structured numerical data to unstructured texts, emails, video and audio), is difficult to fit into a traditional framework. This diverse data is collected from a variety of sources such as business transactions, smart devices, industrial equipment, social media, and more.

SAS Institute, an American multinational developer of analytics software, looks at big data from two more dimensions [8]:

- **variability** - data streams are volatile, information changes frequently and varies significantly. When working with big data, its users should quickly detect what trends are developing in social networks and distinguish between daily, seasonal, and event-driven changes;

- **veracity** – this attribute refers to the quality of data. Due to information coming from so many sources, it is complicated to make connections, clean and transform it across different systems. When analyzing big data, the focus should be on identifying the dependencies and hierarchies which may be hidden in it.

To extract as much value from big data as possible, one should follow certain steps when operating with it. Five key steps may be suggested [2, 8].

**1) Devising a strategy.** A concrete scheme is designed to monitor and optimize the way data is acquired, stored, managed, exchanged, and used. What is crucial is that big data should be perceived as any other valuable beneficial commodity, not just a result of a company's operations.

**2) Identification of sources of big data.** As mentioned above, data is derived from numerous sources and the decision-maker should first determine what the main sources of information are in their organization. For instance, streaming data may come from devices people wear (smart watch, smart clothes, smart glasses) as well as medical equipment, industrial machinery and smart vehicles. The latter is becoming especially famous: for instance, data from cars is expected to be monetized at a global scale and the overall revenue from selling car data might reach 750 billion USD by 2030 [4]. Nowadays a lot of data is also

obtained from social media networks (“likes”, reposts, tweets, comments) on Facebook, Twitter, YouTube, and Instagram. This data can be especially useful for marketing, advertising and technical support. Moreover, there is a considerable amount of open data sources and databases of governments and international organizations. Finally, the data may come from the client themselves.

**3) Accessibility, management and storage of data.** The user should consider all three of these necessary conditions. Infrastructure is needed to access huge volumes and types of big data instantly. When reliable access is acquired, companies should look for proper methods for data integration and data quality assurance. Furthermore, any data should be stored somewhere: either locally in conventional data storage or maybe with the help of cloud systems.

**4) Analysis of big data.** Two approaches are possible regarding the analysis. Either an organization attempts to utilize all the data in its disposal and analyze every piece of it or, first, the relevant information is determined and then it undergoes the analysis. Every user decides for themselves. The use of mathematical statistics, machine learning, deep learning, visualization, predictive analytics are all effective methods of coping with big data and identifying patterns.

**5) Decision-making based on data.** Well-structured, accurate information leads to reliable analytics and effective solutions. Both companies and government need to form their judgements with the help of evidence presented by big data, not only with the help of their instinct. The benefits of data management are obvious. Data-driven enterprises are sure to be more profitable in addition to all of its activities becoming more predictable, meaning no unexpected issues will disrupt the production process.

To summarize, big data has the following benefits for its users:

- Information comes from multiple sources, thus, it is more trustworthy;
- Information is being received constantly and almost instantly, thus, it is more relevant;
- Data is not analyzed manually and that, in turn, reduces the number of errors and increases the amount of information that may be processed;
- The owner of information has centralized access to it.

Thus, big data may serve as a helpful tool to various entities. Firstly, the application of big data in business will be outlined and then the use of big data in governance and public policy will be examined in detail.

## **II. How can big data be used by businesses?**

If collected and analyzed properly, big data may help any business to 1) identify patterns and trends on the market; 2) understand customer behavior better; 3) provide advanced accurate forecasts; 4) reduce costs with the help of improved research; 5) optimize the existing production process and reduce time needed for certain parts of production; 6) launch a new product or service; and just in general 7) make more rational and feasible decisions.

Over the past 10 years, there has been a continuous growth in the number of companies utilizing big data. In 2011 big data was already used by business giants such as Hewlett-Packard, IBM, and Microsoft. In 2015, the share of massive corporations using big data was 17% in the world. Today, the share of such companies is 50% and this share significantly varies across industries. For instance, in telecommunications this share amounts to 87%, in financial services – 76% and in healthcare 60% of enterprises apply big data in their research [1].

In finance big data is used for fraud prevention, evaluation of risks and calculation of credit ratings for each client. In order to predict the markets processes better, financial market models are built. In addition, banks and other financial institutions use big data to improve their cybersecurity and personalize financial solutions for customers [1].

In the entertainment industry media companies build massive recommendation systems which successfully determine the “likes” and “dislikes” of viewers. These recommendation systems carefully analyze viewers’ watching habits, star ratings, reviews to create a personalized experience.

In the agricultural industry big data is used for numerous purposes: from seed development to fighting weed to crop yield forecasting with amazing accuracy. The increase of the amount of data has led researchers to use big data to fight hunger and malnutrition in underdeveloped countries. Such organisations as Global Open Data for Agriculture & Nutrition promote free and unlimited access to global data on nutrition and agriculture, and, perhaps, that is why progress has been made in the fight against world hunger.

## **III. How can big data be used by government?**

Business has always been quite receptive to any innovations and that is why no wonder that, as soon as the advantages of big data were found out, enterprises started to implement new methods connected with big data analytics. However, government structures due to their bureaucratic nature are quite conservative and rather often new methods must work perfectly for them to be implemented into the political process.

The potential value of big data in the public sector is enormous. Governments generate and collect vast amounts of data in their daily activities, such as managing social benefits, collecting taxes, monitoring national health and education systems, recording traffic data, issuing official documents. Access to this information in real time allows governments to identify areas that require attention, make better and faster decisions, and make the necessary changes.

Fields of public administration where big data is applied frequently nowadays [5]:

- Transportation

Road safety is vital when it comes to building a safe environment in cities and towns. Many variables affect how safe different roads are, among them are road conditions, behavior of police officers, vehicle safety, and weather conditions, therefore, it is almost impossible to control anything that might lead to an accident. But big data is the tool that allows governments to guarantee safer and better new roads. Different analytical models help obtain traffic data in real time from CCTV cameras or GPS devices and traffic managers may identify potential threats to road safety and somehow correct the problem.

- Health care

Many health systems nowadays depend on government funding and help. Therefore, there is a possibility of waste of resources or unequal distribution of government subsidies. Big data gives governments a clear idea of where money is going and what the reasons for allocating it are. This means that government agencies will be completely in charge of their resources.

- Education

Big data allows the government to understand educational needs better at both the state and municipal levels. This ensures access to the highest quality of education for young

people, who will be developing the country in the future. In addition, big data may help evaluate students better and track their individual performance.

- Taxation

Tax authorities may apply automated algorithms for analyzing large amounts of data and for integrating structured and unstructured data from social networks and other sources will help them verify the accuracy of information or identify potential fraud.

- Open government

In accordance with open data initiatives, the unlimited exchange of information from institutions to users enhances trust, transparency and accountability between citizens and the government.

Other possible applications correlate a lot with what was said about business applications, among them are [5]: 1) citizen sentiment analysis for politicians to be able to prioritize services and be aware of citizens' interests and viewpoints; 2) segmentation of citizens to adapt public services to specific individuals; 3) combining multiple data sources to assist government economists to make accurate financial forecasts; 4) improving cybersecurity to detect and counter hacking attempts; 5) understanding the electorate's preferences better to conduct more efficient political campaigns.

- Smart cities

The concept of "smart cities" also changes management systems and decision-making algorithms. Increasingly, local governments in different countries are using big data technologies to automate current processes and provide services to citizens and consumers. In smart cities big data is actually applied in the same areas which are described above. One can distinguish main pillars of a smart city, those are: smart mobility, smart economy, smart governance, smart environment and smart people. Additionally, in smart cities big data may be utilized to increase public participation, work on economic growth, energy efficiency, decrease greenhouse gas emissions, develop security and emergency services, manage waste and water pollution [6].

#### **IV. What is the difference in the application of big data between business and public sectors?**

Business and the state have different goals and value systems. Many business decisions are short-term, focused on the competitive environment and are limited in the scale of decision-making groups [3]. However, the purpose of using new technologies in public administration concerns, first and foremost, ensuring security at all levels and improving the quality of life. Decision-making in the public environment usually requires much more coordination, which is hierarchical in nature, while allowing a much lower level of risk, they affect a wider range of people and are often designed for a long-term perspective and joint action [5].

### **Risks and limitations**

Undoubtedly, there are many advantages of using "big data", which will simplify the lives of citizens. However, there are also significant risks and limitations associated with its use.

First of all, that concerns issues of confidentiality and security of personal information. Secondly, due to the penetration of high technology into many processes, employees of various enterprises will also face the problem of job losses, the need for continuous professional development, and businesses, in turn, will have to think about preventing negative consequences, changing management mentality and developing new guidelines for the use of personal information. Finally, in the modern public sector a lack of political willingness to force the public sector to take advantage of these technologies is clearly observed: a change in the mindset of senior officials in the public sector is required to make it possible for big data to be used widely in public administration [5].

### **V. The future of the development of big data**

Gartner, a leading research and advisory company in the field of high technology, has developed a visual interpretation of the development life cycle of any new technology (hype cycle) [3]. As they estimate, big data has already passed the peak beyond expectations on this curve and that means that big data is used universally nowadays, it is no longer an innovation. However, the problem is that, despite big data technologies flourishing, there is still not too much understanding of how to use it properly especially from the point of view of the interaction with society. Having a set of IT tools for big data analysis is only the first step and it is definitely not enough. The most concerning issue is that society does not fully comprehend what the consequences of using smart devices and big data so extensively will be. Not only does big data alter the world completely but also humans in general. It is highly likely that in the future there will not be any things left which will not be collecting data at all

times, therefore, serious research needs to be conducted to define certain limits within which big data will be developing.

## Literature

1. 40 Stats and Real-Life Examples of How Companies Use Big Data // Science Soft  
URL: <https://www.scnsoft.com/blog/big-data-use-cases-stats-and-examples> (дата обращения: 25.11.2020).
2. Gandomi A., Haider M Beyond the hype: Big data concepts, methods, and analytics // International Journal of Information Management. 2015. Vol. 35, Issue 2. P. 137-144.
3. Gartner Hype Cycle // Gartner URL: <https://www.gartner.com/en/research/methodologies/gartner-hype-cycle> (дата обращения: 25.11.2020).
4. Monetizing car data // McKinsey & Company URL: <https://www.mckinsey.com/~/media/mckinsey/industries/automotive%20and%20assembly/our%20insights/monetizing%20car%20data/monetizing-car-data.ashx> (дата обращения: 25.11.2020).
5. Munné R. Big Data in the Public Sector // New Horizons for a Data-Driven Economy. Springer, Cham, 2016. P. 195-208.
6. Nuaimi E., Neyadi H., Mohamed N., Al-Jaroodi J. Applications of big data to smart cities // Journal of Internet Services and Applications. 2015.
7. Patgiri R., Ahmed A. Big Data: The V's of the Game Changer Paradigm. // Conference: 18<sup>th</sup> IEEE High Performance Computing and Communications, 2016.
8. What is Big Data // SAS URL: [https://www.sas.com/en\\_us/insights/big-data/what-is-big-data.html](https://www.sas.com/en_us/insights/big-data/what-is-big-data.html) (дата обращения: 25.11.2020).