IMPLEMENTATION OF BIG DATA
IN OFFICIAL STATISTICS

Aliaksei Zhurauliou[1]

National Academy of Statistics, Accounting and Audit, Kyiv, Ukraine

Oleksii Simachov

National Academy of Statistics, Accounting and Audit. Kyiv, Ukraine

ABSTRACT

Big data is a huge resource that needs to be used at all levels of economic planning. The article is devoted to the study of the development of the concept of "Big Data" in the world and its impact on the transformation of statistical simulation of economic processes.

Statistics at the current stage should take into account the complex system of international economic relations, which functions in the conditions of globalization and brings new forms of economic development in small open economies. Statistical science should take into account such phenomena as gig-economy, common economy, institutional factors, etc. The concepts of "Big Data" and open data are analyzed, the problems of implementing "Big Data" in the official statistics are shown. The ways of the implementation of "Big Data" in the official statistics of Ukraine through the active use of technological opportunities of mobile operators, navigation systems, surveillance cameras, social networks, etc. are presented. The possibilities of using "Big Data" in different sectors of the economy, also on the level of companies, are shown. The problems of storage of large volumes of data are highlighted.

The study shows that "Big Data" is a huge resource that should be used across the Ukrainian economy.

---

1 Address correspondence to Dr. Aliaksei Zhurauliou, Associate Professor, National Academy of Statistics, Accounting and Audit. vyl. Pidhirna 1, Kyiv, 04116 Ukraine E-mail: iro@nasoa.edu.ua

Key words: Big Data, official statistics, institutional environment, open data, Internet, information.

JEL classification: B23, B52, C15

INTRODUCTION

The lack of correspondence between perceptions and official data is one of the issues that prevent researchers from deep and comprehensive modeling of economic systems in post-socialist countries.

The implementation of the concept of Big Data in official statistics may be the basis for a thorough diagnosis of the transformations that take place in the Ukrainian economy (and other countries of Central and Eastern Europe) under the influence of globalization. It is these data that provide the opportunity to form an institutional environment conducive to economic change and further development. They determine the direction of adjustment of the institutional environment when developing social, economic, demographic policies and other important components of economic development. In addition, the use of Big Data emphasizes the importance of new technologies, which allow to transform statistics to a radically new level, in accordance with the current technological layout.

Along with a large number of examples of the successful use of Big Data, there is the problem of developing a mechanism for their implementation in the system of state statistics, as well as the coverage of Big Data of an important sphere such as the institutional environment. Therefore, the introduction of new indicators through Big Data is an important scientific issue.

The aim of the study is to develop ways of implementing Big Data into state statistics.

LITERATURE OVERVIEW

According to O. Vasyechko (2014), "the use of large data for statistical purposes is simultaneously both very promising and revolutionary in terms of the organization

of the production process and methodology. A very attractive point is that, unlike traditional statistical data, mega data can be a fast barometer of social sentiment and well-being. But at the same time they have a significant drawback - the complexity of managing the information flow, since the data is unstructured or poorly structured (they do not have a predetermined relational description model) and come online (data flow)".

O. Ershova and T. Tomashevskaya (2016) emphasize that "to a large extent, the effect of Big Data will depend on the efficiency of management and use of resources. Countries with large enterprises, global communications, with a well-developed infrastructure of information and communication technologies will be able to receive significantly more benefits than those that are lagging behind. Thus, the consequences of the use of large data will be more pronounced in Northern Europe, while in most of the new European and Southern European countries the result will be much lesssignificant. In Ukraine, the use of large data is only heard of in the narrow professional circles of the IT industry".

We agree with the results of the scientific research of these authors and will continue to explore ways of implementing the concept of Big Data into state statistics. In addition, given the growing interest of economists in institutes as determinants of economic development and the importance of Big Data in the sense of formal and informal institutions, we have proposed such a category as "institutional statistics" - an official statistics branch that explores the "rules of the game."
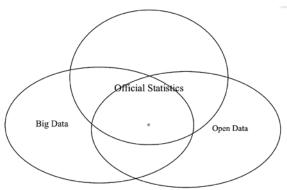
METHODOLOGY

The methodology for Big Data analysis varies greatly from the traditional statistical approach. Usually modelling the data is to explain a response. Big Data analysis presents the researchers with the data. We cannot design an experiment that fulfills the models, thus statistical modeling and regression problems are keystones of the Big Data methodology.

RESULTS AND DISCUSSION

The concept of open data, which reflects the idea that certain data should be freely available for machine use and subsequent reprinting without limitation of copyright, patents and other control mechanisms (Figure 2), is a special place in the Big Data.

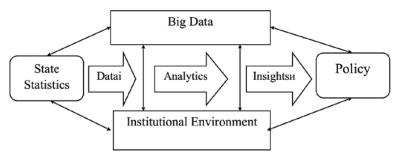*Figure 1 Big Data and official statistics*

You can release data from copyright restrictions using free licenses such as Creative Commons licenses. If any dataset is not a public domain or is not linked to a license granting free re-use, such a dataset is not considered open, even if it is presented in a machine-readable form on the Internet.

An important place in the analysis of data is the theoretical definition of "triads of data". In English, there is a triad of words - data, analytics & insights (data, analysis, and understanding). Since there are no questions with the first two terms in translation into Ukrainian, the term "insights" should be given an additional definition.

Insights are the "added value" of data obtained through analytics. The permeability gained through analytics is incredibly powerful and can be used for a variety of purposes, from exploring business perspectives to fundamental research in many fields of knowledge. The relationship between the data of institutional statics and economic policy is presented in Fig. 2.

Figure 2 Block diagram of the interaction of components of state statistics, institutional environment and economic policies

Working with Big Data is, first and foremost, an effective tool for an effective redistribution of resources. That tool, which will allow you to get fast and effective "victories" here and now. This is an analytics that will promote an entirely different culture of political creation - when politics is formed on the basis of facts, not theoretical considerations. Access to open data and the widespread use of data analysis will improve the decision-making process in the public sector, including legislation and public administration at national and local levels.

The use of Big Data in state statistics is related to solving the problems of securing confidentiality, financing of the relevant work, as well as the availability of specialists for large data and analytics.

The real value of Big Data and their analytics is their ability to provide interesting ideas. You can have many data points, but you must work with them so that these data can provide valuable information.

In the modern world, open data is not only a tool for increasing the transparency and efficiency of public authorities, but also a powerful source for stimulating the development of the country's economy. Unfortunately, most Ukrainian government top managers do not yet fully comprehend the potential and opportunities of open data in the economic aspect. Experts from the EGAP Program have prepared another piece of material on the impact of open data on the EU economy. We hope that this material will be useful for enhancing political will and public activity in Ukraine for further disclosure.

In 2015, the EU conducted one of the most powerful studies to assess the impact and potential of open data on the economic development of the countries and the Eurozone as a whole. As a result, the main indicators of the growth of the open data market were calculated. In particular, the size of the direct market size and indirect market sizes are determined, which together form the general market of open data (total market). The projected size of the open market for open Eurozone data for 2016 will be 55.3 billion euros. Between 2016 and 2020, the market is expected to grow by 36.9% to € 75.7 billion euros. The total market value of open data is estimated at 193-209 billion euros for 2016 and 265-286 billion euros by 2020, adjusted for inflation (Ershova, O. and Tomashvekaya, 2016).

The impact of using open data in terms of economic sectors looks like this. According to experts, the public sector will most benefit from the development of open data (€ 22 billion in 2020). This confirms that the public sector is the main user of their own data. For sectors such as agriculture, arts and entertainment, the expected benefits are much lower and amount to € 379 million for each. Open data in these sectors has significant potential, but it will take time to open it.

As far as the labor market is concerned, the number of jobs that will be directly linked to the reuse of open data in the private sector of the EU will be determined. Due to this, employees working in statistical and other government organizations were not taken into account. Based on forecasts from World Economic Forum the maximum number of jobs directly related to open data in 2016 will be 75,000. By 2020, this figure should increase to 100,000, ie 32% (7.3% per year).

Governmental bodies are striving to more accurately assess the cost savings that they can derive from open data. The economy can be calculated on the basis of projected GDP of the countries of the Eurozone until 2020. The total savings for EU countries in 2020 is projected at 1.7 billion euros. Efficiency is an important factor for both the public and private sectors in all sectors. Improved efficiency aims to improve resource allocation in a way that minimizes costs and results, using the same amount of resources. Based on Deloitte insights three key indicators were used to calculate the efficiency: saving life, saving time and environmental benefits from using open data.

"Great data" not only revolutionizes marketing and sales, they also have a significant potential for use in the public sector: from city planning to coordination

in emergencies, from logistics of medicines in health care facilities to investigative journalism, and projections of election results. If we have access to them, we can use and share them - these are already open data.
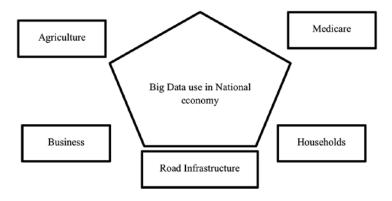
Big Data can be used not only in the system of state statistics of Ukraine, but also domestic enterprises.

An example of the effective use of Big Data in business is Dunkin Donuts. Together with cash checks, the company receives data on the most popular offers from customers who are then processed. This approach allowed the company to increase profits and turnover of goods in stock. The impetus for the implementation of Big Data has been an attempt by the company's management to see what satisfied or dissatisfied customers are. Finding that customer expectations are the fastest service, the company has decided to replace the standard loyalty system (coupons or regular customer cards) with the new smartphone application.

Due to the possibility to analyze such information, the company can decide whether there are enough local workers, or shortage of products during peak times, to make promotional offers more individual, etc. Thus, using "great data", the company can predict consumer behavior and develop a development strategy based on their analysis.

The possibilities of using Big Data in various sectors of the economy are presented in Fig. 3.

Figure 3 Possibilities for using Big Data in different sectors of the economy

In energy, the use of Big Data provides an increase in the accuracy of the distribution of power generators. In the mining and oil industry - the efficiency of the development of the deposit, tracking the schedule of overhaul and the state of equipment, forecasting demand for products and prices.

For law enforcement agencies, Big Data is the information of the system ALPR (automatic identification of license plates), crime maps, etc.

In the health system, Big Data has information on:

- susceptibility to diseases by analysis of blood of patients and donors;
- reducing the cost of the state and improving the effectiveness of treatment;
- fighting diseases and monitoring the growth of epidemics;
- potentially possible epidemics.

According to Deloitte, "great data" is "fuel" for modern business, and humanity creates data at an impressive rate. Every 2 days in the world, the volume of new data produced equals what has been produced since the beginning of civilization until 2003.

It is clear that organizations are trying to reveal the potential of Big Data. The result of the analysis of Big Data is accelerating the pace of innovation and the destruction of traditional business models.

The main problems of storing large volumes of data:

- computing security in distributed software systems;
- security of non-relational databases;
- safe storage of data;
- verification of reliability;
- real-time security monitoring;
- encryption of access control and security of connections;
- partial access control;
- the origin of the data.

CONCLUSION

The concept of Big Data and open data are analyzed, problems of implementation of Big Data in the state statistics are shown. The ways of implementation of Big Data are presented in the state statistics of Ukraine through active use of technological possibilities of mobile operators, navigation systems, surveillance cameras, social networks, etc. The possibilities of using Big Data in different sectors of the economy, also on the level of companies are shown. The problems of storage of large volumes of data are highlighted.

However, gathering information for the formation of Big Data and incorporating it into the statistical production process is not an easy task. For example, traffic data consists of measurements of traffic intensity. Each cycle calculates the number of cars per minute that pass a certain location, as well as measures such as speed and length. Such data is of interest to traffic and transport statistics, as well as to other economic phenomena associated with transport.

REFERENCES:

Vasechko, O. (2014) Suchasni vyklyky statystychnoi vyschoi osvity i nayki. *Statistics of Ukraine*, 4(67), 4-10.

Ershova, O. & Tomashvekaya (2016), N. Velyki Dani (Big Data) i modernisaciya nacionalnyh system oficiynoi statystyki. Applied Statistics: problems of theory and practice. 2, 32-38.

World Economic Forum [Electronic resource] URL: https://www.weforum.org/agenda/2016/05/are-economic-statistics-in-danger-of-becoming-irrelevant

Statistical Database. Eurostat. European Commission [Electronic resource]. URL: http://ec.europa.eu/eurostat/data/database.

Statistical Year book [Electronic resource]. URL: http://unstats.un.org.

Deloitte insights [Electronic resource]. URL: https://www2.deloitte.com/ca/en/pages/deloitte-analytics/articles/data-analytics-and-privacy.html