Organization of Big Data in The Structure of The Digitalization Ecosystem of a Globalized Society

The notions of “globalization” and “digitalization” are discussed. The main tendencies of globalization processes, their positive and negative implications for the society are determined. The modules of digitalized eco-systems are described. It is determined that the data sets that currently exist are the engine of innovation and are new alternative sources of statistics for official statistics. An approach to organization of big data is elaborated for demonstrating the data hierarchy.

In spite of all the risks globalization opens up new opportunities and eliminates the borders first and foremost for education, R&D, medical services, and manufacturing. Developing in the conditions laid by globalization, countries need to consolidate the effort, because global problems imply global approaches to their solutions.

Keywords: globalization, Industrial Revolution 4.0, digitalization, eco-system of digitalization, big data.

O. O. HOROBETS,
Postgraduate student,
National Academy of Statistics, Accounting and Audit,
e-mail: babutska@ukr.net,
ORCID: 0000-0003-1762-2140,
Researcher ID: G-7664-2018

Introduction. We are living through the era of information technology, with gadgets and devices being basic necessities and information flows permeating the human life. There is no human life activity not changed by technological innovations related with Internet. Now we buy goods and services, engage in banking, find information and talk with friends in Internet, often create special cloud groups by interests. The speed and scales of Internet and the more and more powerful technologies born by it had great significance for the human society [11]. Such rapid growth and ramification of technologies and related processes pose new research and technical problems related with development, improvement, modernization of production processes and training of future specialists in keeping with the times.

© O. Horobets, 2020

UDC: 330.341.1
JEL Classification: O3; O1; M 15; F01; C80
doi: 10.31767/nasoa.3-2020.10

Organization of Big Data in The Structure of The Digitalization Ecosystem of a Globalized Society

Організація великих даних у структурі цифровізаційної екосистеми глобалізованого суспільства

У статті зосереджено увагу на поняттях «цифровізація» та «глобалізація». Визначено основні тенденції глобалізаційних процесів, їх позитивні та негативні наслідки для суспільства. Охарактеризовано блоки цифровізаційної екосистеми. Визначено, що ті масиви даних, які наразі існують є двигуном інновацій та виступають новими, альтернативними джерелами статистичних даних для офіційної статистики. З метою усвідомлення ієрархії даних, розроблено організацію великих даних. З’ясовано, що незважаючи на усі ризики, глобалізація відкриває нові можливості для розвитку освіти, науки, медицини. Розвиваючись в глобалізаційних умовах, країнам необхідно діяти консолідовано, з метою подолання глобальних проблем. Таким чином, наукова розвідка сучасного стану цифровізаційних процесів в умовах глобалізації 4.0, стали основою темою написання цієї роботи.

Ключові слова: глобалізація, Четверта промислова революція, цифровізація, екосистема цифровізації, великі дані.

Introduction. We are living through the era of information technology, with gadgets and devices being basic necessities and information flows permeating the human life. There is no human life activity not changed by technological innovations related with Internet. Now we buy goods and services, engage in banking, find information and talk with friends in Internet, often create special cloud groups by interests. The speed and scales of Internet and the more and more powerful technologies born by it had great significance for the human society [11]. Such rapid growth and ramification of technologies and related processes pose new research and technical problems related with development, improvement, modernization of production processes and training of future specialists in keeping with the times.
The topic “Globalization 4.0” was the central one at the World Economic Forum in Davos in 2019. Schwab K. said that “Globalization 4.0 has only just begun, but we are already vastly underprepared for it. Clinging to an outdated mindset and tinkering with our existing processes and institutions will not do. Rather, we need to redesign them from the ground up, so that we can capitalize on the new opportunities that await us, while avoiding the kind of disruptions that we are witnessing today” [42]. Describing Industrial Revolution 4.0 in his book, Schwab K. puts emphasis on its specific features: much more disseminated and rapid Internet, smaller, more powerful and cheaper sensor devices, the developed artificial intellect and machines capable to self-learning [37].

The above said raises the importance of a study of tendencies underlying the globalized world, with analysis and rethinking of the current trends in digitalized processes in the context of globalization 4.0, which determines this article’s objective.

**Theoretical Frame.** Globalization is seen as an all-permeating phenomenon taking root in infinite numbers of factors and events, rapidly transforming our society and encompassing not only economic, political and technological domains, but socio-cultural and environmental ones. The enhanced economic integration, global forms of management and globally linked social and environmental events are often referred to as globalization [23].

According to Kreshnik A., globalization is a reality of modern economies. However, given the irreversibility of globalization process, it still remains a subject of debate and critique [22].

Reznikova N. argues that “at the beginning of the 21th century, the global economy entered a complicated turbulent period of the evolution… The balance between traditional state institutions of decision-making and new centers controlling resources and economic processes required to implement the decisions was broken” [34].

No doubt that globalization and technologization are powerful and mainly positive forces. But there is one indisputable fact: they trigger complex problems, in particularly ones related with inviolability of private life and national security [1].

In 2002, the Swiss Economic Institute, in collaboration with the Swiss Technological Institute, constructed KOF Globalization Index, which includes three groups of global integration indicators, namely: level of economic globalization (indicators of international trade, business activity, trade flows, tariff policies etc.); level of social globalization (indicators of cultural integration, percentage of foreign residents, development of tourism, information flows etc.); level of political globalization (membership of countries in international organizations, number of embassies etc.) [16]. These groups of indicators enable to monitor the penetration of globalization processes. Useful for comparisons are illustrative graphs of global processes penetration, built on the basis of Globalization Index.

Figure 1 shows the world map as it was in 1070, with dark shades marking the countries that had already underwent considerable globalization-driven transformations in course of their development. As can be seen from Figure 1, the overall KOF Globalization Index was equal to 40.29 de facto and 36.35 de jure.

Figure 2 illustrates the globalization-based integration in 2016, with the overall KOF Globalization Index equaling 58.35 de facto and 64.2 de jure.

By now, with estimates of Globalization Index available for 2017, the top three leaders of globalization are the Netherlands, Ireland, and Belgium.

In 2017, three globalization-driven megatrends were emphasized at the session of the UN General Assembly: change in the production process and labor markets; rapid development of technologies; climate change [40].

Our society witnesses the birth and gradual expansion of digitalization that has direct impact on the above mentioned risks. Change in daily life due to the rise of Internet is obvious for each one. We are on the midway to the digital economy. In fact, this process started nearly 50 years ago, but the pace of digital penetration has become much more rapid in the latest years [29].

Pinker S. when discussing future scenarios of social development, emphasizes that the contemporary world has plunged in the digital megalomania. This is, first and foremost, related with the rapid advance in information technologies. Lacking good awareness of this phenomenon, the society is prone to idealize its technologies and overestimate its perspectives [31].
When investigating digitalization concepts, Jandrić P. divides them into three ages: the analog world and its digital perceptions, information revolution and its consequences, and post-digitalization challenge [21].

Digitizing opened up incredible opportunities for information collection. New types of technologies appeared and penetrated in the variety of life spheres: business, trade, travels, medical care, leisure activities, daily communications etc. It made the people start generating extremely large quantities of information. According to Ross A., while in 2000 the share of data stored in digital form was 25%, in 2007 it grew up to 95% [36]. The cumulative amount of digital data was 5.6 zetabytes in 2015¹. Jain V. K. argues that information generated by the people is expected to grow up to 44 zetabytes by 2020 [17]. But these amounts of data are of no value without appropriate treatment: quality assurance, timely and logical processing and interpretation.

Technological achievements in the several latest years allowed for wide-scale exploitation of high speed broadband connection providing for easy access to information any time, and wide-scale applications of mobile devices. The massive dissemination of mobile devices pushed companies and government offices to introduce online and mobile supply of information and services. Transition from physical assets to digital ones is but the starting point. Thus, by taking a paper form and transferring it to Internet, we do more than just create an electronic version of data that were recorded before on a paper sheet by means of writing. The access to online form can be opened to a wider audience, the

¹ One zetabyte equals to sextillion (1021) bytes or trillion gigabytes.
data can be analyzed in the mode approximated to real time, in a common template, with responses standardized to ensure a broader exchange, and user feedback being quicker and more effective [4].

Each minute people send 204 million electronic messages, make 2.4 million entries in Facebook, and display 216 thousand new photos in Instagram, thus laying the conditions for destruction of privacy [36]. Most part of Internet users do not understand that the data reported by them in Internet will remain in the digital environment for ever. These include geo-locations, messages, video-translations, comments, photos, data on jobs, attendance of events etc. Although most part of this information is unlikely to be used against an Internet user, but the full protection is not guaranteed.

However, the so called “digital apocalypse” should not be talked about, as the real world makes a global barrier for it [31]. Kelly K. agrees that now all the digitalized processes are socially integrated in spite of their power [19].

Due to the global change, a range of rapidly advancing technologies promoting the next phase of the digital transformation has been already created. It can be noted that each digitalized process consists of the data that form the modules creating the digitalize eco-system. According to OECD report, the eco-system of digital technologies consists of seven modules (Figure 3), with each containing its own arrays of unique data [28].

For better understanding the phase at which digitalized process penetrates in national economies and human life activities, these modules should be briefly described in this paper (Figure 3):
Organization of Big Data in the Structure of the Digitalization Ecosystem of a Globalized Society

Figure 3: An ecosystem of interdependent digital technologies

Sources: [28].

*Internet of Things* (IoT) is a sensor network incorporating billions of gadgets intellectually linked via PI-connection, which contains non-structured data generated in the real time regime. IoT (Internet of Things) is an emerging paradigm of communication, where devices act as objects or “things” capable to feel their environment and connect with each other and exchange data via Internet [20]. It should be understood that IoT is actually a part of software package SCADA (Supervisory Control And Data Acquisition), which purpose is to ensure collection, processing, reproduction and archiving of data about an object. Therefore, data analysis and IoT is used in a variety of development scenarios: in health care, education agriculture, transport, in particular for testing of technical services [9]. Data used by IoT grow exponentially due to connection of many billions of devices in a relatively short period time [26]. One trillion of IP-addresses or objects will be connected to Internet via IoT by 2022 [8; 14]. The totality of opinions about IoT is divided by into two classes: the first one is the reactive base of ideas and opinions, which treats IoT as a layer of digital connection above the existing infrastructure and things, i.e. the controlled set of convergent developments of infrastructures, services and applications. According to Rob van Kranenburg & Bassi A., IoT is a new technology which is yet to gain the public awareness. Currently we can discern two main blocks of thought on IoT. The first is a reactive framework of ideas and thought that sees IoT as a layer of digital connectivity on top of existing infrastructure and things. This position sees IoT as a manageable set of convergent developments on infrastructure, services, applications and governance tools. It is assumed that, as in the transition from mainframe to Internet some business will fail and new ones will emerge, this will happen within the current governance, currency end business models. The second is a proactive framework of ideas and thought that sees IoT as a severely disruptive convergence that is unmanageable with current tools, as it will change the notion of what data and what noise is from the supply chain on to ‘apps’. In both these approaches we find the same challenges. The difference will be in the solutions and approaches. But an important issue remains to be security of IoT, because the integrity, confidentiality and accessibility are three significant factors in IoT systems. The applications using IoT model, such as industrial or medical ones, are considered as vitally important in most cases. The powerful security measures in IoT networks are required. Such security mechanism must protect IoT network and its resources, not affecting the system productivity and the confidentiality of users [35].
The 21st century is obviously a “century of speed” in all the possible spheres, especially in communications: a great variety of services, software, equipment, opportunities etc. In spite of this, the global supply makes human life more complicated and causes loss of time. In order to integrate the time, a new super-powerful technology started to be sought, which would ensure quick data processing and, consequently, time saving [3]. 5G is the speed of Internet of the next generation. 5G is an integrated network with various services and technologies corresponding to the future needs of work with a broad range of big data and rapid development of great many businesses, and enhancing the user experiences. The mobile system 5G requires great achievements in design and development of the system architecture, to have the mobile traffic greatly increased [43].

The creation of Internet laid the background for gaining this grandiose vision of “computer utilities” of the 21st century through building the global system of computer networks, by which individual computers can communicate with any other computers located in other parts of the world. Kelly K., who compares clouds with colonies of millions of computers, argues that clouds run our digital life [19]. He stresses that the greatest advantage of a cloud is that the higher it is, the smaller and the thinner is a gadget. A cloud today is computations with an impressive degree of reliability, high speed, capacities for detailed data processing, with no need for technical support. It follows that cloud computations are a new and advanced paradigm providing IT services as computing software [6]. Cloud computing platforms are interpreted by us as an amorphous environment with infinite numbers of elements characterized by scalability and high computing capacities given their low cost. Being directly in the cloud environment, a platform is a service. According to Zhang Q., cloud computations have recently become a persuasive paradigm of management and service provision via Internet [42]. The growing scales of cloud computations have been rapidly changing the landscape of information technologies, with eventually turning the long-standing promise of launching massive useful computations into reality. However, in spite of considerable advantages offered by cloud computations, the existing technologies have not been sufficiently mature to implement all their potentials. Many key problems in this field, computerized resource supply, energy management and security management in particular, have just begun to draw the research community’s attention. It follows that cloud software requires a high level of infrastructure computerization and competencies in specialized operations, which is not common for corporate IT organizations [7]. All the components of the digital technology eco-system have a common core: the data that are integrating and integral component of their existence. When it comes of the data, it should be understood that in the present-day context it refers to big data. Technology advancement led to the rapid growth in the data scopes in the latest years. Scopes, diversity (structured, non-structured and semi-structured data) and speed of big data also have changed the paradigm of systems’ computing capacities. Non-structured data imply heterogeneity of data types and contents, and sophisticated semantic interpretations. These implicit problems of non-structured data raise problems for big data analysis. To make non-structured data accessible in a form fit for analysis, they need to be transformed in a structured context and prepared for analysis [2]. In 2016, the 126th meeting of the Competition Committee held a hearing where seven key points of big data were outlined, by which big data need to be considered as a digital capacity generating innovations, for data collection, processing and analysis. Thus, control of big data sets will not necessarily build the market capacity, because some digital markets feature tough and dynamic competition. But the capacities to generate and process big data sets may be associated with effects for market capacities, resulting from economies of scale, network effects and cycles of data feed back in the real time. If even these effects do not necessarily lead to domination on the market, they should be considered as part of competition analysis [27]. Schwab K. emphasizes that the so called turning point in understanding, acceptance and use of big data will occur by 2025, when governments replace a conventional population census by bases containing big data sets [37].

This data are a clear indication of the progress triggered by digitalization processes. Considering the conventional sources of big data [30], granulation of big data reported by representatives of Eurostat [11], and big data types developed by OECD [26] (Figure 4), one needs to have at least an elementary idea on how big data are organized. Organization of big data will enable for more detailed granulation of the data.
ORGANIZATION OF BIG DATA IN THE STRUCTURE OF THE DIGITALIZATION ECOSYSTEM OF A GLOBALIZED SOCIETY

Figure 4: Organization of big data
Source: author’s development.

It is well known that all the data appear constantly, i.e. once an information process is completed, and create a single array of raw data. The latter can be divided into three modules of data, with each one broken into several groups.

The first module contains the most general groups of data. They are five in number, from broader data (administrative or official data) to data about individual or public opinion (photo-data, video-data etc.). But all of them are interlinked, meaning that sometimes similar data can occur in these groups.

The second module is more granulated and contains seven groups of data. This module’s specific is that each group has its individual profile. Thus, the group “personal data” often contains confidential information. Therefore, the data mining process requires much caution for (i) avoiding violation of confidentiality, (ii) extracting high quality statistical information from the data included in this group. The groups “Data on private sector” and “Individual data” are more often overlap, creating a mix of duplicated data. The group “Research data” contains results of multidisciplinary research activities in form of research conclusions contained in traditional and digital sources of information. A difficulty with this group is that not all the data are digitalized and harmonized with each other.

The third module in this raw data array is designed for data screening and extracting reliable, precise and harmonized data, to form more aggregate groups of data, make them internally hierarchical and prepare them for repeated use.

Accordingly, each of the modules will have dark data remaining after handling data of any business process. Usually these data are not intended for repeated use, because their keeping is rather costly, which raises the question about the feasibility of these data storage [13].

Such organization of big data cannot be regarded as complete, but it can lay the ground for further elaborations and improvements.

Artificial intellect (AI) is intellect that can be displayed by machines, unlike the natural type of intellect shown by humans, or a way by which we can create smart machines that work and react like humans [32].

When discussing artificial intellect, the novel of D. Brown “Origin” comes to mind, where the quantum computer nicknamed Winston was the best and the most intellectual
helper, disciple and friend of its “creator” [5]. But D. Brown did not finish it with a happy end: finally Winston killed its “creator”, because it had learned to take own decisions that had been beyond its creator’s control. Using such stereotype of artificial intellect, D. Brown could scare the public and reinforce the negative attitude to robotics and any form of artificial intellect. But Naam R., a computer scientist, reassures that “…Imagine that you are a superintelligent AI running on some sort of microprocessor (or perhaps, millions of such microprocessors). In an instant, you come up with a design for an even faster, more powerful microprocessor you can run on. Now…drat! You have to actually manufacture those microprocessors. And those fabs take tremendous energy, they take the input of materials imported from all around the world, they take highly controlled internal environments which require airlocks, filters, and all sorts of specialized equipment to maintain, and so on. All of this takes time and energy to acquire, transport, integrate, build housing for, build power plants for, test, and manufacture. The real world has gotten in the way of your upward spiral of self-transcendence” [25]. This opinion about digitalization was confirmed by Kelly K. in 2017 and by Pinker S. in 2018 (already mentioned above). It should be noted that it was in 2016 when Microsoft, Amazon, Google, Facebook, and IBM companies announced the partnership in the work on artificial intellect for the public benefit. A telling example of employing artificial intellect for the public benefit is applications of Watson system created by IBM in the health care. This system is capable of analyzing and screening extremely large scopes of information (books, results of research, medical records, medical histories etc.), for making correct diagnosis by finding minute causalities traced according to the patient’s anamnesis and the analyzed reference base [12].

Blockchain is the increasing digital register for storage of entries in the distributed form, which cannot be changed [32]. By now Blockchain has reached the fifth level of development, which implies all-purpose applications of smart contracts and products developed by telecommunication technologies, biotechnologies or chemical technologies (Pro blockchain media). This technology was the first step in creating digital currencies and absolutely new systems for storage and exchange of valuable assets both in digital and real economies. Blockchain is regarded as a revolutionary technology that helps one control both positive and negative features of the digital revolution, with assuring transparency and invariability of data. It is also important that human interference in the processes is minimized, and digital registers can be inclusive. The blockchain of today can be compared with Internet of early 90s. According to Primavera de Filippi, the revolutionary role of blockchain is in the set of tools preventing from exploitation, with the possibility of impact on the society and economy that are increasingly dependent on technologies [37].

From the report it follows that nearly three of the ten managers consider blockchain rules as the largest barrier for exploitation of the technology on the full capacity [39]. About one quarter of respondents reported lack of user trust as the largest barrier for the blockchain adoption. A case of successful application of blockchain technologies is the “smart city” strategy developed in the city of Seoul (South Korea): “Blockchain Urban Plan” worth 105 million USD for the period of 2018–2022 [24].

All the above mentioned components of the digital eco-system are impossible without computing power providing for uninterrupted operation of technologies.

When discussing computing facilities, the law of Moore should be emphasized, by which the number of transistors per square inch was being doubled each 1.52 year beginning with the middle of 1960s. Accordingly, the computer size was becoming smaller, with the computing capacity growing and the computer price reducing each year.

The quantum computations are predicted to gain the global leadership in the era of Industrial Revolution 4.0. When it comes to quantum computers, it should be noted that these computers use quantum qubits for information storage. The advantage of these computers is their capability to process information hundreds or thousands times quicker than classical computers. Advanced research institutes and companies like IBM, Oxford, Stanford, Google and others have practiced quantum computations since long, which is not surprising, because, as already mentioned, Google and IBM companies have been developing artificial intellect for the public benefit, which requires powerful computations [38].

Concluding this part of the article, the worlds of Enno de Boer and Subu Narayanan should be mentioned that the effort of many companies to apply decisions of Industry 4.0
Organization of Big Data in The Structure of The Digitalization Ecosystem of A Globalized Society

has stuck in “pilot purgatory” [10]. Successful cases of digitalizing projects can by no means be ignored, but, as mentioned above, there are many barriers blocking quick promotion of innovations, and their removal still may take a lot of time.

While the society needs time for the awareness, acceptance and regular use of technologies, businesses need time for technology testing, modernization of production processes, staff learning and technology implementation. The greatest challenge for governments is to have these new expectations implemented. Governments adapt approaches to providing public services and policy setting to the era of digital technologies. But new digital approaches supported by necessary change in the culture of government sector must be introduced, so that the governments could successfully meet the needs of citizens of business [26].

Conclusion. It can be concluded that the globalization context inevitably change both technologies and social behavior of human. Globalization-driven transformations are an integral part of any national economy, but a persisting problem is uneven development of countries (or even regions within a country)².

The humanity generating and accumulating the extremely large scopes of information has not a clear idea of its further processing or storage. But such information growth triggers performance of new research, writing of new algorithms, creation of new mechanic organisms, initiation of new trends in business management and project development.

By exploiting technologies, humans can make the life easier (washing, cleaning etc.) and save the time, but in exchange for voluntarily given personal data.

However, in spite of all the above mentioned risks globalization opens up new opportunities and eliminate the borders first and foremost for education, R&D, medical services, and manufacturing. Developing in the conditions laid by globalization, countries need to consolidate the effort, because global problems imply global approaches to their solutions.

References

² Given that 17% of the global community has not yet approached the threshold of the second global revolution, and 1.3 billion of people still have no access to electricity, the statement about unevenness of penetration is valid [37].


Bibliographic description for quoting: