

УДК 316.334.2:004.65

**BIG DATA: CONCEPT, TRENDS AND SECURITY ISSUES' RELEVANCES**

**Valeriy Nikolayevskyy** – PhD (Philosophy), Professor of the Department of Sociology, V. N. Karazin Kharkiv National University, 4 Svobody Sq., Kharkiv, 61022, Ukraine, e-mail: vnikolayevskyy@gmail.com, ORCID ID <http://orcid.org/0000-0003-1840-2932>

**Viktoriia Omelchenko** – PhD (Sociology), associate Professor of the Department of Philosophy, Kharkiv National University of Radio Electronics, 14 Nauky Ave., Kharkiv, 61166, Ukraine, e-mail: viomelchenko@gmail.com ORCID ID <https://orcid.org/0000-0002-1095-5244>

**Ніколаєвський Валерій Миколайович** – кандидат філософських наук, професор кафедри соціології Харківського національного університету імені В. Н. Каразіна, майдан Свободи, 4, Харків, 61022, Україна, e-mail: vnikolayevskyy@gmail.com, ORCID ID <http://orcid.org/0000-0003-1840-2932>

**Омельченко Вікторія Вікторівна** – кандидат соціологічних наук, доцент кафедри філософії Харківського національного університету радіоелектроніки, пр. Науки, 14, Харків, 61166, Україна, e-mail: viomelchenko@gmail.com, ORCID ID <https://orcid.org/0000-0002-1095-5244>

*Authors of the article focus on some theoretical and practical issues of Big Data. They present a brief overview of theoretical approaches in studying of Big Data concept in up-to-date English literature in Social sciences and Humanities, as well as computer and information studies; and give various definitions of this concept. The authors note the interest to this concept in Ukrainian sociology discourse and mark a certain mythologization of the reasons that caused that interest. The article illustrates the practice of business usage of Big Data in Ukraine. 2017-2018 top Big Data trends which were highlighted by experts of Dataversity, the Economist, Gartner, Tableau are presented in the article also. The authors point out in the article the following trends: Big Data becomes fast and approachable, Big Data grows up; the use of business intelligence, based on Big Data, becomes more important to small and medium-sized businesses, and even start-ups; variety, not volume or velocity, drives Big Data investments; the convergence of Internet of Things, cloud and Big Data creates new opportunities for self-service analytics; changing security challenges are among them. Authors focus on some research perspectives, and security issues, in the context of the analysis of Big Data for Social Sciences and Humanities.*

**Keywords:** Big Data, data, Big Data trends, Big Data definitions, personal data protection.

*Автори статті фокусують увагу на деяких теоретичних та практичних питаннях Big Data. У статті представлений короткий огляд теоретичних підходів у вивченні концепту Big Data в сучасній англійській соціогуманітарній літературі, літературі з комп'ютерних наук та інформатики; а також наведені різні визначення цього поняття. Автори відзначають інтерес до цього концепту в українському соціологічному дискурсі та акцентують на певній міфологізації причин, що спонукали такий інтерес. Статтю ілюстровано прикладами практики використання Big Data бізнесом в Україні. Автори приділяють увагу найважливішим тенденціям Big Data на 2017-2018 рр., які були відзначені експертами низки спеціалізованих агенцій – Dataversity, the Economist, Gartner, Tableau. Зокрема, автори вказують на наступні тенденції: Big Data стає швидкими та доступними, Big Data зростають; використання бізнес-аналітики, заснованій на Big Data, стає більш важливим для малого та середнього бізнесу та стартапів; різноманітність, а не обсяг або швидкість, є драйвером інвестицій Big Data; конвергенція Інтернету речей, хмарних технологій і Big Data створює нові можливості для аналітики самообслуговування; зміна викликів безпеки. Автори зосереджують увагу на деяких дослідницьких перспективах Big Data для соціогуманітарних наук, зокрема в контексті дослідження безпекових питань.*

**Ключові слова:** Big Data, дані, інформація, тенденції Big Data, визначення Big Data, захист персональних даних.

There are many arguments that we are witnessing the beginning of the Big Data era. The phenomenon called Big Data is becoming increasingly important for the nowadays world, it's actually one of the biggest IT trends of the last few years, and one of the most attractive and promising areas of business, which penetrates (with high-powered effect) into our everyday life<sup>1</sup>. Recently we can see a growing interest in Big Data in Ukrainian

© Nikolayevskyy V., Omelchenko V., 2018

DOI: 10.26565/2227-6521-2018-40-09

<sup>1</sup> Perhaps, there will not be any global business player, even out of IT, which neglects Big Data and, so, would not build strategies for its using. For instance, in the field of education draws attention to the example of California State University's work. This University has urged its faculty to use free, or low-cost, materials in teaching classes. To simplify the process, (replacing previous course material with such materials) Intellus Learning (an educational platform that enables colleges and universities to organize and find digital content and resources. It focuses on supporting teaching and learning in higher education with intelligent analytics that help faculty select and recommend the suitable content for each student – authors) provided a solution, by indexing over 45 million online resources and teaching (by way of machine learning) the program/algorithm to make recommendations. By that the instructor can upload the free (or low-cost) materials into

sociology also, although from the beginning this process is not free from certain mythologization<sup>2</sup>. Sure, because of its technical nature there are large amount of questions regarding Big Data querying and processing which is swiftly increasing<sup>3</sup>, but mostly analysis of such issues is beyond the scope of this article<sup>4</sup>.

IT is dramatically changing the world. According to the IT experts what we have nowadays is not the era of petabyte, but the exabytes one. That changes concern both individuals and society. The very nature of economic relations is undergoing fundamental changes. Everybody has become both a consumer and a producer. Other change is that, earlier, before, the production of books and music etc. went through a strict guarantying quality procedure. Nowadays, all of us are witnesses, that everyone can put Youtube videos or text in social network, regardless of their quality. People no longer accept to pay perceived high prices. So, companies, organizations have to rethink their business model and what their added value is. The way we behave is changing also. It was common to see people reading a book or watching a movie with continuously searching web, checking their mail or social media. The ever increasing usage of a variety of digital devices and (especially) remote sensors that generate continuous streams of digital data, resulting in what has been called as Big Data. And this is no longer surprising. Furthermore, web gives us the feeling that we have access to all the information in the world, resulting in delaying decisions because obtaining more data may lead to perceived better decisions<sup>5</sup>.

Sure, researches in many fields of sciences require the analysis of data. As was mentioned above, one can encounter data at every turn and it seems that data is everything. At the same time, it's deceptive that extracting meaningful information from the data is easy task. Often we meet different terms such as Big Data, large data, small data, high-dimensional data, data visualization, research data, digital data, linked data, web data, open data, etc. without a proper definition of such words. The rapid growth in the size and scope of data sets in a variety of sciences has created a need for innovative strategies for analyzing such data. A number of instruments (statistical, computational tools) are needed to deal with such type of data.

What is Big Data? What does it mean? There are many other questions concerning Big Data. Try to answer some of them.

It is rather difficult to overview literature to define Big Data, because it's extensive. But even brief analysis outlines some essential features. Firstly, we have to note, that the stable tradition exists, when Big Data is analyzing in connection with data.

C. Borgman provocatively offers instead of asking «What are data?» to question: «When are data?», because data are ubiquitous, ephemeral and «because recognizing that some phenomenon could be treated as data

the course materials management system, and make them available to students [1]. Of course, there are examples and practices of using of Big Data in Ukraine too, but they are rather modest and can be counted «on the fingers». We can say, for instance, about the projects of Ukrainian transport company «UZ» (launching in 2016 for intellectual ticket sales using Big Data and mobile applications etc.), the project «Big Data Lab», provided by mobile operator Vodafone Ukraine with the goal of creating a holistic and capable ecosystem of entrepreneurial activity built on the using of Big Data tools and of forming the Big Data market in Ukraine [2, 3, 4].

<sup>2</sup> In fact, from our point of view, interesting in the phenomenon of Big Data of Ukrainian sociologists was actualized during the last two years (more precisely, since the fall of 2016) [5, 6], and has been much stimulated by the Big Data wide usage in the presidential victorious campaign of D.Trump (2016). Usually, it's a question of the algorithm for constructing the social networks (Facebook) profiles in accordance to their preferences (classification of each person according to a «Big 5 Personality» traits lines), breakthrough in nature and widely used by analysts of Cambridge Analytica in that campaign (simultaneously, we consciously pay no attention to Cambridge Analytica and Facebook related scandal caused by accusations of illegal use of Big Data in that campaign). In reality, this type of profiling is not new, it has been in use for years and even in other USA presidential campaigns: in B. Obama's and H. Clinton's election campaigns too [7]. So, in any case, without doubting the facts of the widespread use of the Big Data in the D. Trump's recent presidential campaign, it should be noted that there is hardly any reason to assert that this criterion refers to the uniqueness of that campaign. Most likely, attracting attention to this circumstance (at the beginning, until mid-March of 2018) it looked like a successful PR campaign by Cambridge Analytica. For some details – see below.

<sup>3</sup> For this reason the article is inevitably burdened by a number of technical terms.

<sup>4</sup> Nevertheless, because Big Data is directly related to data, data science it is important to emphasize that data science is a scientific field that extracts useful knowledge and predictions from large datasets, using techniques from data mining, machine learning, predictive analytics and business intelligence. It comprises two fundamental activities: data preprocessing and data modeling. As to the first of them (data preprocessing) its purpose, as to our opinion, is to clean the data and put it in a proper format, because data comes from sources that are different in nature, such as paper notes, text files, relational databases, access logs, web documents, reports and data warehouses. This data has missing, redundant, irrelevant, or inconsistent values and is untidy, i.e., is not formatted in a way that is adequate to further processing by data science algorithms. As to the latter (data modeling and exploration) we guess the following: while prepared, data is explored in order to find patterns, relations between variables, association rules, tendencies, etc. that may be not obvious to the researcher. The pattern to search depends on the specific problem at hand. This is an iterative process that is largely improved with the use of visualization techniques. Data mining focuses on pattern extraction and knowledge extraction; machine learning focuses on algorithms that learn from past evidence and that are used as predictive techniques, classification and decision support; predictive analytics focuses on the use of statistical methods, such as regression analyses, to predict future values from past experience; and business intelligence focuses on presenting complex information to decision makers [8, p. 447].

<sup>5</sup> We agree with Peter Apers (University of Twente, The Netherlands) that above-mentioned results in a number of issues require attention, for example, the amount of data is increasing at a tremendous speed (by one account, it is more than doubling every two years [9]), making us actually close to blind; the average quality of the data is going down, many data is contradicting each other, which reflects on the level of trust to the data. Is it possible to extract reliable knowledge from the Internet, asks P. Apers also [10, p. 1].

is itself a scholarly act» [11, p.5]. Other conceptualizations of data include data as a reified, external resource, as information that is generated as data, and as a process of ascribing meaning. For example, one of the five distinct definitions of information in the discipline of information science that J. Furner outlines is «information-as-data» which is defined as «Any object, event, or property (or aggregate of such) that takes material form and to which it is possible to ascribe meaning». S. Zins proposes to define these concepts by creating a knowledge map of the field that involved analyzing 130 definitions of data, information, and knowledge formulated by scholars in information science. He found that the conceptual model of data, information, and knowledge most often used by scholars of that field is one in which data and information are conceptualized as external phenomenon and where knowledge is viewed as internal [12, p. 4-5].

Similar to the data definitions' approaches, H. Ekbja et al. give a critical overview of Big Data through the discussion of dilemmas – «a situation that presents itself as a set of indeterminate outcomes that do not easily lend themselves to a compromise or resolution» [13] – that serve to frame their discussions of Big Data and epistemology, methodology, aesthetics, ethics, and technology.

Other definitions of Big Data vary from simple definitions that Big Data are those datasets that are so large and complex, that it's difficult for traditional tools and software applications to process them [14] to those (for instance, of V. Mosco) that give greater consideration to the political economy of Big Data [cit. 12, p. 5]. V. Mayer-Schonberger and K. Cukier suggest that Big Data create new forms of value and reshape ideas about innovation and relationships as it «overturns centuries of established practices and challenges our most basic understanding of how to make decisions and comprehend reality» [15, p. 6-7].

The experts of Opentracker note that the definition of Big Data is a moving target, but propose list of 33 definitions of Big Data. Here are some of them:

1. R. Magoulas' definition: «Big Data is when the size of the data becomes part of the problem»;
2. M. Gualtieri: «A more pragmatic definition of Big Data must acknowledge that: Exponential data growth makes it continuously difficult to manage – store, process, and access. Data contains nonobvious information that firms can discover to improve business outcomes. Measures of data are relative; one firm's big data is another firm's peanut. A pragmatic definition of Big Data must be actionable for both IT and business professionals. Big Data is the frontier of a firm's ability to store, process, and access (SPA) all the data it needs to operate effectively, make decisions, reduce risks, and serve customers»;

3. J. Ebbert: «The world has always had «Big Data». What makes «Big Data» the catch phrase of 2012 is not simply about the size of the data. «Big data» also refers to the size of available data for analysis, as well as the access methods and manipulation technologies to make sense of the data»;

4. D. Boyd and K. Crawford: «...Big Data is a cultural, technological, and scholarly phenomenon that rests on the interplay of: 1) technology: maximizing computation power and algorithmic accuracy to gather, analyze, link, and compare large data sets; 2) analysis: drawing on large data sets to identify patterns in order to make economic, social, technical, and legal claims; 3) mythology: the widespread belief that large data sets offer a higher form of intelligence and knowledge that can generate insights that were previously impossible, with the aura of truth, objectivity, and accuracy» [16].

T. Elliott proposes to «compress» definitions of Big Data into 4 groups:

- 1) The current use of the term Big Data took off because it was used to refer to new technologies like Hadoop and MapReduce;

- 2) Most definitions refer to the data characteristics outlined by analyst D. Laney in 2001: the 3Vs: extremes of volume, velocity, or variety (other Vs sometimes evoked include Validity or Veracity for data quality, and the importance of business Value). Wikipedia's definition of Big Data is among them;

- 3) Others prefer emphasizing the difficulties of manipulating the data in different phases such as Sense, Collect, Store, and Analyze;

- 4) Finally, for many business people, Big Data is now becoming a generic term for «analytics», except it's done by «data scientists» rather than «business analysts». Business people appear to be more comfortable with the term Big Data than some of its predecessors (DSS or Decision Support Solutions) [17].

It may seem interesting how Big Data is understood by practitioners, executives. According to a survey by SAP, nearly 76% of them see Big Data as an opportunity. However, respondents' definition of Big Data varied to a considerable degree. Nearly a quarter of the 154 C-suite executives felt that Big Data was the technologies designed to handle the massive amounts of data swamping organizations. Another 28% defined Big Data as that flood of data itself. Still another group (19%) equated Big Data with storing data for regulatory compliance. Around 18% viewed Big Data as the increase in data sources, including social networks and mobile devices [16].

What is the way of Big Data evolution, its main trends? As to our opinion, some answers were suggested by experts of Dataversity, Gartner, Tableau who highlighted top Big Data trends for 2017-2018. Some of them are the following:

1. Big Data becomes fast and approachable: options expand to speed up Hadoop. Clarification. The first question people often ask is: How fast is the interactive SQL? The need for speed has fueled the adoption of faster

- databases like Exasol and MemSQL, Hadoop-based stores like Kudu, and technologies that enable faster queries.
2. Big Data is no longer just Hadoop: purpose-built tools for Hadoop become obsolete. Clarification. In previous years, we saw several technologies rise with the big-data wave to fulfill the need for analytics on Hadoop. But enterprises with complex, heterogeneous environments no longer want to adopt an isolate Business Intelligence (BI) access point just for one data source (Hadoop). Answers to their questions are buried in a host of sources ranging from systems of record to cloud warehouses, to structured and unstructured data from both Hadoop and non-Hadoop sources. Customers will demand analytics on all data. Platforms that are data-and source-agnostic will thrive while those that are purpose-built for Hadoop and fail to deploy across use cases will fall by the wayside. The exit of Platfora (Big Data analytics company) serves as an early indicator of this trend.
  3. Organizational decision-making is currently undergoing a shift which will continue into 2018. In 2017, the goal of processing Big Data promoted ever-increasing efficiency and steadily decreasing costs. In turn, this has made the use of BI, based on Big Data, more important to small and medium-sized businesses, and even start-ups. This trend will continue with the cost of processing Big Data continuing to drop.
  4. Organizations leverage data lakes from the get-go to drive value. Clarification. A data lake is like a man-made reservoir. First you dam the end (build a cluster), then you let it fill up with water (data). Once you establish the lake, you start using the water (data) for various purposes like generating electricity, drinking, and recreating (predictive analytics, ML, cyber security, etc.). Up until now, such a «hydrating the lake process» has been an end in itself. That will change as the business justification for Hadoop tightens. Organizations will demand repeatable and agile use of the lake for quicker answers. They'll carefully consider business outcomes before investing in personnel, data, and infrastructure. This will foster a stronger partnership between the business and IT. And self-service platforms will gain deeper recognition as the tool for harnessing big-data assets.
  5. Architectures mature to reject one-size-fits all frameworks. Clarification. Hadoop is no longer just a batch-processing platform for data science use cases. It has become a multi-purpose engine for ad hoc analysis. Organizations will respond to these hybrid needs by pursuing use case-specific architecture design. They will research a host of factors including user personas, questions, volumes, frequency of access, speed of data, and level of aggregation before committing to a data strategy. These modern-reference architectures will be needs-driven. They will combine the best self-service data prep tools, Hadoop Core, and end-user analytics platforms in ways that can be reconfigured as those needs evolve. The flexibility of these architectures will ultimately drive technology choices.
  6. Variety, not volume or velocity, drives Big Data investments. As well known, Gartner defined Big Data as the three Vs: high-volume, high-velocity, and high-variety information assets. While all three Vs are growing variety is becoming the single biggest driver of Big Data investments. This trend will continue to grow as business organizations seek to integrate more sources and focus on the «long tail» of Big Data. Data formats are multiplying and connectors are becoming crucial. Analytics platforms are evaluating based on their ability to provide live direct connectivity to these disparate sources.
  7. Spark and machine learning light up Big Data. Clarification. Apache Spark once a component of the Hadoop ecosystem, is now becoming the Big Data platform of choice for enterprises. These big-compute-on-big-data capabilities have elevated platforms featuring computation-intensive machine learning, AI, and graph algorithms. Opening up machine learning to the masses will lead to the creation of more models and applications generating petabytes of data. As machines learn and systems get smart, all eyes will be on self-service software providers to see how they make this data approachable to the end user.
  8. The convergence of Internet of Things (IoT), cloud and Big Data creates new opportunities for self-service analytics. Clarification. It seems that everything soon will have a sensor that sends information back to the mothership. IoT is generating massive volumes of structured and unstructured data, and an increasing share of this data is being deployed on cloud services. The data is often heterogeneous and lives across multiple relational and non-relational systems. While innovations in storage and managed services have sped up the capture process, accessing and understanding the data itself still pose a significant last-mile challenge. As a result, demand is growing for analytical tools that seamlessly connect to and combine a wide variety of cloud-hosted data sources. Such tools enable businesses to explore and visualize any type of data stored anywhere, helping them discover hidden opportunity in their IoT investment.
  9. Self-service data prep becomes mainstream as end users begin to share Big Data. Clarification. Making Hadoop data accessible to business users is one of the biggest challenges of our time. The rise of self-service analytics platforms has improved this process. But business users want to further reduce the time and complexity of preparing data for analysis, which is especially important when dealing with a variety of data types and formats. Agile self-service data-prep tools not only allow Hadoop data to be prepped at the source, but also make the data available as snapshots for faster and easier exploration. We've seen a host of innovation in this space from companies focused on enduser data prep for Big Data (such as Alteryx, Trifacta, etc.). Tools which are lowering the barriers to entry for late Hadoop adopters and laggards will continue to gain traction.
  10. Big Data grows up: Hadoop adds to enterprise standards. Clarification. One of a growing trend is that

Hadoop is becoming a core part of the enterprise IT landscape. And there will be more investments in the security and governance components surrounding enterprise systems. Apache Atlas, created as part of the data governance initiative, empowers organizations to apply consistent data classification across the data ecosystem. Apache Ranger provides centralized security administration for Hadoop. Customers are starting to expect these types of capabilities from their enterprise-grade RDBMS platforms. These capabilities are moving to the forefront of emerging big-data technologies, thereby eliminating yet another barrier to enterprise adoption.

11. Rise of metadata catalogs helps find analysis-worthy Big Data. Clarification. For a long time companies threw away data because they had too much to process. With Hadoop, they can process lots of data, but the data is not generally organized in a way that can be found. Metadata catalogs can help users discover and understand relevant data worth analyzing using self-service tools. Cataloging helps both data consumers and data stewards reduce the time it takes to trust, find and accurately query the data. We shall see more awareness and demand for self-service discovery, which will grow as a natural extension of self-service analytics.

12. Changing security challenges. New Internet security challenges will become a problem in 2018. Security becomes a significant issue as more people are given access to sensitive information. It is predicted hackers will seek to access the IoT for destructive purposes [1, 18, 19].

As one can conclude from the analysis of literature, mostly researches (and this is quite understandable) explore issues related to technical components of the problem. But at the same time, many problems are common for different disciplines. For example, we can say about biometrics applications' booming nowadays, which make it a new Big Data challenge in its streaming, processing, classification and storage.

Big Data gives lots of opportunities for researchers who want to make analysis based on several data sources available on the Internet. It can be used for marketing and sales optimization, market forecasting, making informed management decisions, etc. One of the most common examples is to make analysis based on the web content. Usually such data sources are unstructured, but in many cases webpages used for a specific analysis can be semi-structured. For example, webpages which are related to the analysis of real-estate market, job offers, etc. And, if there is a need to make analysis of real-estate market, several different webpages present offers in a similar way, which means that some data can be extracted using specific structure enhanced with tags. This allows preparing a good quality dataset for further analysis.

S. Sicular (research director at Gartner) suggests that business people' main goals for now are to learn how to identify and formulate Big Data problems, and to grow their own skills and experience with Big Data technologies, while these technologies are evolving and maturing. Good solutions are possible although not easy [20].

J. Maślankowski's (University of Gdańsk, Poland) logic of reasoning is approximately the following: Big Data gives lots of opportunities for researchers who want to make analysis based on several data sources available on the Internet. If more than one data source are used, the same information can be collected two or more times. Although the popularity of Big Data analysis is increasing, the importance of Big Data's quality is arising, there is still no framework that can be used to eliminate duplicates from large unstructured datasets. Therefore, J. Maslankowski says, it is necessary to focus on a framework that can be used to eliminate duplicates for Big Data analysis [21, p.104-105].

A number of questions for social sciences and humanities, as well as information studies practitioners, researchers in the context of Big data security arise.

In general, as is well known, there are four different aspects of Big Data security: infrastructure security, data privacy, data management, and integrity and reactive security, which are the focus of the article of J. Moreno, M. A. Serrano and E. Fernández-Medina [22].

Data can be contradictory in that it can be used for surveillance, violating on privacy, be used for secondary purposes (often without consent), and can be totalizing in that we continually create data discharge, it can be hacked, searched, aggregated, and preserved for years [12, p. 1]. An illustrative, vivid and convincing (far from the first and, most likely, not the last) example is the loud scandal associated with the D. Trump's election campaign in 2016, which directly touched the Facebook<sup>6</sup>. Vice versa, data can be used for the public good, to promote social change, and to empower people. It seems that rather important and promising to provide the analysis based on the methodology of critical theory.

---

<sup>6</sup> As is well known in March (2018) Facebook was in the center of a scandal after The New York Times, The Guardian conducted an investigation into the illegal use of personal data of social network users. Among other things, it was also about that Cambridge Analytica (consulting company, UK), who worked for the D.Trump campaign, used personal information from over 50 million Facebook users to target political advertising without their permission, which allowed to influence the course of the election. [23] Is anyone really surprised that personal data in social networks is not sufficiently protected? So, the question is, what explains the transformation of this, as it seems, an ordinary, non-unique incident in the scandal? From our point of view, perhaps such a loud effect is due to a certain coincidence in time of several circumstances (factors): the legal (Cambridge Analytica is a British company, it was involved in the Brexit campaign also), political (firstly, it refers to the election campaign of D.Trump, who has tense relations with media, in particular The New York Times, and secondly, the «trail» stretches to the SPbSU, Russia), as well as the size and status of the Facebook company itself.

As for Big Data analysis and security issues, personal data protection it seems that the situation may get worse. In support of this «The Economist» notes in one of its recent editorial «The myth of cyber-security»: «Computers will never be secure. To manage the risks, look to economics rather than technology» [24, p. 7]. The pessimism is grounded on the following: «The problem is about to get worse. Computers increasingly deal not just with abstract data like credit-card details and databases, but also with the real world of physical objects and vulnerable human bodies. A modern car is a computer on wheels; an aeroplane is a computer with wings. The arrival of the IoT (Internet of Things) will see computers baked into everything from road signs and MRI scanners to prosthetics and insulin pumps. There is little evidence that these gadgets will be any more trustworthy than their desktop counterparts. Hackers have already proved that they can take remote control of connected cars and pacemakers». And «...it is certainly true that many firms still fail to take security seriously enough» [24, p. 7]. In our opinion, paying attention to the importance and efficiency of the technical and legal mechanisms for ensuring the security of the Big Data, more attention should be given to the complex tools, increasing of person's media culture, information security as well as person's media security culture are among them.

**REFERENCES:** 1. Foote K.D. Big Data Trends for 2018. (November 28, 2017) / Dataversity. URL: <http://www.dataversity.net/big-data-trends-2018/> (accessed – 15.01.2018). 2. Kulesh S. Vodafone Ukraine zapuskayet proekt Big Data Lab, v ramkakh kotoroho otkroyet massiv svoikh realnykh dannykh IT-razrabotchikam. ITCua. URL: <https://itc.ua/news/vodafone-ukraina-zapuskayet-proekt-big-data-lab-v-ramkakh-kotorogo-otkroet-massiv-svoih-realnyih-dannyih-it-razrabotchikam/> (accessed – 7.11.2017) (in Russian). 3. «UZ» nachinayet prodazh kvytiv z vykorystannyam Big Data. «Slovo i dilo». URL: <http://www.slovoidilo.ua/2016/12/16/novyna/ekonomika/ukrzaliznycya-zapuskaye-prohramu-prodazhiv-kvytkiv-z-vykorystannyam-big-data> (accessed – 16.12.2016) (in Ukrainian)/ 4. UZ prisoedinyaetsya k portalu E-data i nachnet publikovat tam dannye o dvizhenii svoikh deneh. CFTS. URL: [http://cfts.org.ua/news/2017/12/20/uz\\_prisoedinyaetsya\\_k\\_portalu\\_e\\_data\\_i\\_nachnet\\_publicovat\\_tam\\_dannye\\_o\\_dvizhenii\\_svoikh\\_deneg\\_44635](http://cfts.org.ua/news/2017/12/20/uz_prisoedinyaetsya_k_portalu_e_data_i_nachnet_publicovat_tam_dannye_o_dvizhenii_svoikh_deneg_44635) (accessed – 20.12.2017) (in Russian). 5. Metodologiya i metody sotsiologichnykh doslidzhen' v Ukraini: istoriya ta suchasni problemy. Do 70-richchya Volodymyra Paniotto: zbirnyk statej za materialamy konferentsii, 22 sichnya, m. Kyiv / za nauk.red. V. S. Bakirova ta Ye. I. Holovakhy. Kyiv: Instytut sotsiologii NAN Ukrainy, V. N. Karazin KhNU. 2017. 208 p. 6. New Inequalities – New Conflicts: Ways to Overcome. Abstracts of the III Congress of Sociological Association of Ukraine, Kharkiv, 12-13 October 2017. –Kharkiv, 2017. – 516 P. 7. Davis J. Psycho-demographics: Drilling Down on Trump's Campaign Analytics. (13.04.2017). All Analytics. URL: [https://www.allanalytics.com/author.asp?elq\\_mid=77960&elq\\_cid=23083305&section\\_id=3858&doc\\_id=282448&mc=NL\\_NJ\\_EDT\\_20170419&cid=NL\\_NJ\\_EDT\\_20170419&elqTrackId=c6fa7b08a34042d59c3bf67c02339334&elq=fce4fd9098b54640ac673f3a1cae6bbb&elqaid=77960&elqat=1&elqCampaignId=26307](https://www.allanalytics.com/author.asp?elq_mid=77960&elq_cid=23083305&section_id=3858&doc_id=282448&mc=NL_NJ_EDT_20170419&cid=NL_NJ_EDT_20170419&elqTrackId=c6fa7b08a34042d59c3bf67c02339334&elq=fce4fd9098b54640ac673f3a1cae6bbb&elqaid=77960&elqat=1&elqCampaignId=26307) (accessed – 12.02.2018). 8. Urbano J., Pedro Nogueira, Ana Paula Rocha, and Henrique L. Cardoso. Analysis of Data Science Tools for Sensor-Based Assessment of Quality of Life in Health Care. *Recent Advances in Information Systems and Technologies: Vol.I-II.* / Alvaro Rocha, Ana Maria Correia, Hojjat Adeli, Luis Paulo Reis, Sandra Costanzo (Editors). Vol. I. Springer, 2017. P. 446-455. 9. Kakutani M. Watched by the Web: Surveillance Is Reborn. 'Big Data,' by Viktor Mayer-Schönberger and Kenneth Cukier. *The New York Times*. URL: <http://www.nytimes.com/2013/06/11/books/big-data-by-viktor-mayer-schonberger-and-kenneth-cukier.html> (accessed – 10.06.2013). 10. Apers P. Data Overload: What Can We Do? *Database Systems for Advanced Applications*. 18th International Conference, DASFAA 2013, Wuhan, China, April 22-25, 2013. Proceedings, Part I. / Weiyi Meng Ling Feng, Stephane Bressan WernerWiniwarter, Wei Song (Eds.). Springer, 2013. P. 1-2. 11. Borgman C. Big Data, Little Data, No Data: Scholarship in a Networked World. Cambridge, MA: MIT Press, 2015. 12. Oliphant T. A Case for Critical Data Studies in Library and Information Studies. *Journal of Critical Library and Information Studies*. No. 1. 2017. URL: <http://libraryjuicepress.com/journals/index.php/jclis/article/view/22/13> 13. Ekbia H. et al. Big Data, Bigger Dilemmas: A Critical Review. *Journal of the Association for Information Science and Technology*. 2015. Vol. 66. No. 8. P. 1523-1545. 14. Sherman C. What's the Big Deal About Big Data? *Online Searcher*. 2014. Vol. 38. No. 2. P. 10–16. 15. Mayer-Schonberger V., Cukier K. Big data: A Revolution That Will Transform How We Live, Work, and Think. New York: Houghton Mifflin, 2013. 242 P. 16. Definitions of Big Data. Opentracker. Event tracking and analytics (updated: April 26, 2013). URL: <http://www.opentracker.net/article/definitions-big-data> (accessed – 23.03.2017). 17. Elliott T. What is Big Data? / Business Analytics & Digital Business. URL: <http://timoelliott.com/blog/2012/12/what-is-big-data.html> (accessed – 12.02.2017). 18. Top ten Big Data trends for 2017. Tableau. URL: [https://www.tableau.com/sites/default/files/media/Whitepapers/whitepaper\\_top\\_10\\_big\\_data\\_trends\\_2017.pdf?ref=lp&signin=095b34d5b95542a260f40aa8c02d4d0c](https://www.tableau.com/sites/default/files/media/Whitepapers/whitepaper_top_10_big_data_trends_2017.pdf?ref=lp&signin=095b34d5b95542a260f40aa8c02d4d0c) (accessed – 25.03.2017). 19. The World in 2018 / The Economist (UK). The Economist Newspaper Ltd., 2017. – 154 P. 20. Sicular S. Gartner's Big Data Definition Consists of Three Parts, Not to Be Confused with Three «V»s / Forbes.com. URL: <https://www.forbes.com/sites/gartnergroup/2013/03/27/gartners-big-data-definition-consists-of-three-parts-not-to-be-confused-with-three-vs/#1a8332c342f6> (accessed – 27.03.2013). 21. Maślankowski J. Towards De-duplication Framework in Big Data Analysis. A Case Study. Information Systems: Development, Research, Applications, Education: 9th SIGSAND/PLAIS EuroSymposium 2016 Gdansk, Poland, September 29, 2016 Proceedings / Stanislaw Wrycza (Ed.) : Springer, 2016. P. 104-113. 22. Moreno J., Manuel A. Serrano and Eduardo Fernández-Medina. Main Issues in Big Data Security. Future Internet 2016, 8(3). URL: <http://www.mdpi.com/1999-5903/8/3>. 23. Cadwalladr C., Graham-Harrison E. Revealed: 50 million Facebook profiles harvested for Cambridge Analytica in major data breach. The Guardian. URL: <https://www.theguardian.com/news/2018/mar/17/cambridge-analytica-facebook-influence-us-election> (17.03.2018). 24. The myth of cyber-security. The Economist (UK). 2017. Vol.423. No. 9035. April 8<sup>th</sup>. P. 7.

---

**Big Data: концепт, тенденції, безпека**

Автори статті фокусують увагу на деяких теоретичних і практичних питаннях Big Data. В статті представлено короткий огляд теоретичних підходів в вивченні концепта Big Data в сучасній англійській соціогуманитарній літературі, літературі по комп'ютерним наукам і інформатиці; а також наведені різні визначення цього поняття. Авторами відзначається інтерес до цього концепту в українському соціологічному дискусії і акцентують на певній міфологізації причин, побудивши такої інтерес. Стаття проілюстрована прикладами практики використання Big Data бізнесом в Україні. Авторами віддається увагу найважливішим тенденціям Big Data на 2017-2018 рр., які були названі експертами ряду спеціалізованих агентств – Dataversity, the Economist, Gartner, Tableau. В частині, в статті пропонується звернути увагу на наступні тенденції: Big Data стають швидкими і доступними, Big Data ростуть; використання бізнес-аналітики, заснованої на Big Data, стає більш важливим для малого і середнього бізнесу, а також стартапів; різноманітність, а не обсяг або швидкість, є драйвером інвестицій Big Data; конвергенція Інтернету речей, хмарних технологій і Big Data створює нові можливості для аналітики самообслуговування; зміна викликів безпеки. Авторами зосереджується увага на деяких дослідницьких перспективах Big Data для соціогуманитарних наук, зокрема в контексті дослідження питань безпеки.

**Ключові слова:** Big Data, дані, інформація, тенденції Big Data, визначення Big Data, захист персональних даних.

Отримано 22.03.2018