

About Big Data Measurement Methodologies and Indicators

Makrufa Sh. Hajirahimova

Institute of Information Technology of Azerbaijan National Academy of Sciences,
B. Vahabzade str., Baku, AZ1141, Azerbaijan
Email: makrufa@science.az

Aybeniz S. Aliyeva

Institute of Information Technology of Azerbaijan National Academy of Sciences,
B. Vahabzade str., Baku, AZ1141, Azerbaijan
Email: aybeniz63@rambler.ru

Received: 28 July 2017; Accepted: 12 September 2017; Published: 08 October 2017

Abstract—The digitization of nearly all media and the increasing migration of social and economic activities to the Internet, the development of social networking technologies, the Internet of Things and cloud computing caused rapid increase in the volume of data and the formation of Big Data paradigm. Big Data involves technologies and tools for collecting, processing, analyzing and extracting useful knowledge from structured and unstructured data of large volumes generated at high speed by different sources. Increasing the volume, speed, diversity and value of Big Data began to play an important role in the creation of social relationships, competitive advantage and innovative fields. The development of the information society, the formation of digital economy, and the application of Big Data technologies in different spheres of human activity required the quantitative and qualitative assessment of Big Data. In this article some approaches relate to the definition of Big Data have been reviewed. Methodological approaches and indicators for measuring Big Data have been researched. At the end, the indicators have been proposed for the measurement of factors that affected the growth and development of Big Data.

Index Terms—Big Data, Big Data indicators, Big Data measurement, information infrastructure, innovation factors, technological factors, human factor.

I. INTRODUCTION

The beginning of the third millennium is characterized by the emergence of the information society, many significant technological changes, such as cloud computing, the Internet of Things and social networking. The development of these technologies has made the amount of data increase continuously and accumulate at an unprecedented speed, and announce the coming of Big Data. This digital universe will be growing 40% per year into the next decade. An IDC report [1] predicts that, from 2013 to 2020, the global data volume will grow

exponentially and will reach from 4.4 zettabytes to 44 zettabytes. The advent of the Internet of Things (IoT) working with intelligent systems in the billions, and involving millions of applications is a fourth growth spurt for the digital universe. But also all the “things” (smart devices) connected to the Internet, unleashing a new wave of opportunities for businesses and people around the world. Growing of volume, speed, diversity and value has resulted to significant changes in transition direction to the model that managed with data in socio-economic activity paradigm. So, Big Data began to play an important role in the creation of social impact, competitive advantage and innovation field [2]. In the near it is inevitable future economic and political competition between the countries will be based on using of Big Data's potential opportunity [3].

Therefore, there is a need to develop indicators that allow monitoring of the current situation of Big Data for find out how Big Data can give rise to the socioeconomic development of society.

II. WHAT IS BIG DATA?

Big Data is a revolutionary phenomenon, which is one of the most frequently, occurred topics in scientific and practical discussions in fact, several definitions for Big Data are in the literatures. For example, in Oxford Dictionaries [4] define Big Data as: “Extremely large data sets that may be analyzed computationally to reveal patterns, trends, and associations, especially relating to human behavior and interactions: much IT investment is going towards managing and maintaining Big Data”. Essentially, Big Data is not only a large volume of data, but there are also other features which differentiate it from the of massive data and very large data. There are several explanations of Big Data, but three types of definitions play a significant role in forming how Big Data is viewed: attributive definition (in there includes IDC, IBM, Gartner, Microsoft researcher's definition), comparative definition (in there includes McKinsey

definition) and architectural definition (in there includes The National Institute of Standards and Technology (NIST) definition) [5, 6].

Gartner defines Big Data in [7]: “high-volume, high-velocity and high-variety information assets that demands cost-effective innovative forms of information processing for enhanced insight and decision-making”. This definition of Big Data is referred to as a 3V model.

Microsoft researchers D. Boyd and K. Crawford define Big Data as 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 [8].

IDC defines Big Data in [9]: “Big Data technologies describe a new generation of technologies and architectures, designed to economically extract value from very large volumes of a wide variety of data, by enabling high-velocity capture, discovery, and analysis”. This definition delineates the four salient features of Big Data, i.e., volume, variety, velocity and value (4Vs).

McKinsey's report in 2011 [10] described Big Data as huge data sets whose volume is over the potency of typical database software tools to capture, store, manage, and analyze.

The National Institute of Standards and Technology (NIST) suggests that, “Big Data is where the data volume, acquisition velocity or data representation limits the ability to perform effective analysis using traditional relational approaches or requires the use of significant horizontal scaling for efficient processing”. NIST introduces five attributes such as volume, velocity, variety, horizontal scalability and relational limitation to classify Big Data [11].

UN Global Pulse [12] defines as: Big Data is a popular phrase used to describe a massive volume of both structured and unstructured data that is so large that it's difficult to process with traditional database and software techniques. The characteristics, which broadly distinguish Big Data, are sometimes called the “3Vs”: more volume, more variety and higher rates of velocity.

“Integrating Big Data into the Monitoring and Evaluation of Development Programmers” was published in UN Global Pulse in 2016. In fact, from the perspective of international development, Big Data is an integrated approach to research and development that involves three interlinked components: 1) Data generation: Generation of new sources of data; 2) Data analytics: Organization, integration, analysis and dissemination of Big Data; 3) Data ecosystem - Involving producers, analysts, users and regulators of Big Data [13].

The theoretical fundament of Big Data is missing. Because relevant topics and related theories are unknown. To overcome this obstacle Marco Pospiech and Carsten

Felden introduce a Big Data theory model to describe the current relationships and concepts. It (the model) unveils the underlying characteristics of Big Data [14].

The existence of altered attitudes in Big Data's definition has led to the formation of different approaches methodologies in the field of measurement.

III. RELATED WORKS

Big Data incorporates along the volume of collected information as well as storage, computing technologies, combines and devices.

A. Methodological Approaches in Measuring Big Data Volume

Numerous researches have been conducted in the field of volume of rapidly increasing information and the rate of growth in the world. Although these studies have differing trends and results, they are of great importance.

Hal Varian and Peter Lyman at the University of California Berkeley is one of the first researchers in the field of measurement of volume of data produced, stored, and transmitted. As part of their “How much information?” project that ran from 2000 to 2003, they assessed that 5 exabytes of new data were stocked universally in 2002 and that 92% of the new information was stored on magnetic media, mostly in hard disks. It is an initial exhaustive exploration to quantify, in computer storage terms, the total amount of new and original information created in the world by the year and stored in four physical media: paper, film, optical (CDs and DVDs), and magnetic[15].

Since 2007 the study firm IDC to produce an annual series of reports on the “Digital Universe” to measure the amount of digital information created and recurred each year. The scientists appraised that in 2007 all the on hard drives, tapes, CDs, DVDs, and memory in the market equaled 264 exabytes [16]. The basic methodological approach of the IDC in measuring of the Digital Universe was described as follows:

- Progress a prediction for the installed base of devices or applications that could capture or create digital information.
- Guesstimate how many units of information (files, images, songs, minutes of video, phone calls and so on) were created per year.
- Modified these units to megabytes
- Evaluate the number of times a unit of information is replicated.

IDC research is based on more than 40 devices. Performs digital data calculation across the world and nearly 90 countries. IDC estimates that in 2016, the amount of digital information being created, captured, and replicated passed over 9.3ZB [17].

Measurement of information in global storage, transmissions and computing devices was carried out by a team of researchers from the University of Southern California of the USA guidance by Martin Hilbert. The

total amount of information was calculated using the following formula and applied to each of the telecommunications, storage and computing technologies [18]:

$$IC = \sum_i Q_i P_i$$

There IC - informational capacity, Q_i - shows the i -th device, P_i - shows the performance of the i -th device, i - all considered technologies.

As a result of the research, in 2007, all data were estimated to be 295 exabytes volumes. During the evaluation of information, the technological factors such as the amount of infrastructure and devices and productivity of each device are taken as a basis [18].

The trend of formation of Big Data can be characterized by the speed of data generation. In particular, the speed of the data generation increasingly depends on the technological achievements.

The Cisco System links data rates with video, Internet and cameras. During the evaluation of Internet traffic, the factors such as the number of users, the number of Internet video users, minutes of use, the bit rates. The methodology is to multiply the bit rates, minutes of use (MOU), and users together to get average PB per month [19].

The employees of San Diego University of California Roger E. Bohn, James E. Short approaches to information as consumption point of view and in 2008 they carried out the measurement of information that consumed by individuals (consumable information in the workplace was not included) in the USA. According to R. Bohn and J. Short assessing, in 2008, the USA population consumed 3.6 zettabytes of information per year. Average consumption time per user during the year is assessed [20].

More than 200 states and industrial enterprises' sources, population registration, Labor Statistics Bureau and other state sources of USA, as well as various academic, industry and state research sources were used when the amount of data and hours were calculated. Time, word, and bytes are used as measurement units for consumed information. Factors such as the number of users during the measurement of information, the average daily hour per person using each technology, and the amount of information transmitted (consumed) by technology (bytes or words) in one second are taken as basis [20].

McKinsey Global Institute (MGI) investigator group research the amount of data that enterprises and individuals are generating, storing, and consuming throughout the global economy [10]. MGI assesses that enterprises globally stored more than 7 exabytes of new data on disk drives in 2010, while consumers stored more than 6 exabytes of new data on devices such as PCs and notebooks. MGI's estimation of the size of Big Data includes:

- New data stored in enterprise external disk storage in a year.
- New data stored by consumers in a year.

To approximate the amount of data generated and stored in total, model is based on four key inputs and presumptions: 1) annual storage capacities shipped by sector; 2) average replacement cycle of storage; 3) utilization rate; 4) duplication rate.

B. Big Data Measurement Indicators

Big Data indicators can play a role in assessing the progress and impact of Big Data-related growth and performance in the information society (and innovation economy). But generally accepted indicators systems for measurement of Big Data are not still available.

At below let we look at some of Big Data indicators.

C. Measurement of the Big Data Ecosystem

As mentioned above, data ecosystem is one of main components of Big Data. In 2012 in Massachusetts state within the framework of "Massachusetts Big Data Initiative" was created Big Data ecosystem with high competitive advantage and its measurement was carried out [21]. Measurement of the ecosystem was approached from business, technology, talent, and capital aspect. The Innovation Institute at Massachusetts Technology Collaborative has identified eight key indicators that summarize Massachusetts' competitive position in Big Data and the expansion of the Big Data ecosystem [22]. These indicators are:

- 1) A number of Big Data and data-driven related companies;
- 2) Volume investments in data-driven and Big Data companies;
- 3) A number of data-driven research centers;
- 4) A number of Big Data related meetup groups;
- 5) A number of Big Data related patents;
- 6) Data science related programs;
- 7) Big Data projects received federal investment;
- 8) A number of data-related STEM (Science, Technology, Engineering and Mathematics) fields graduate.

Big Data ecosystem indicators in Massachusetts in 2015 is outlined in Table 1. below.

D. Data Market Indicators

Big Data technologies play a special role as the enabler of most of the innovative services and applications. The diffusion of mobile and social technologies, IoT and cloud computing in turn generates huge amounts of consumer and business data. IDC tracks digital innovation developments worldwide and in Europe.

In 2014 initial indicators have been proposed for measurement of Big Data in Europe market by IDC company. These indicators are as follows [23, 24]:

- A number of data workers (workers who collect, store, manage and/or analyses, interpret and visualize data);
- A number of data related companies (the organizations providing data (data-suppliers) and making a strong reliance on data (data-users));
- Revenue of data related companies;
- Data market size (the market where digital data is exchanged as products or services derived from raw data);
- Data workers' skill gaps (the potential gap emerging between the demand and supply of data workers);
- Citizen's data economy.

Table 1. Indicators of Big Data Ecosystem in Massachusetts

Indicators	Related findings
Big Data and data-driven related companies	537 companies driving the Massachusetts Big Data ecosystem
Investments in data-driven and Big Data companies	\$2.4B invested in Mass Big Data companies
Data-driven research centers	44 research centers at eight universities and five hospitals
Big Data related meetup groups	52 active meetup groups
Big Data related patents	3191 patents were awarded in 23 categories
Data science related programs	10 data science programs (two Bachelors, three Masters, one combined BS/MS, three certificates, and one doctorate)
Big Data projects received federal investment;	200 new Big Data projects have received federal investment (over \$115 million)
Data-related STEM fields graduate	6170 students in data-related STEM fields graduated from Massachusetts colleges and universities in 2013

In particular, the total number of data workers - their shares on the total employment and their average numbers per company, data companies are the organizations providing data/data-suppliers and making a strong reliance on data/data-users, the value of the overall data market - the market where digital data is exchanged as products or services derived from raw data and the value of the overall data economy including the economic impacts generated by data-suppliers, data users and the whole. The data workers' skill gap - the potential gap emerging between the demand and supply of data workers.

This table 2. was presented a set of indicators for the years 2015 and 2016, that forecasted to 2020 on the basis of measuring the European population of data workers, the value of the data market, a number of data user enterprises, a number of data companies, their revenues and data workers' skill gaps.

Table 2. European Data Market Indicators

Indicators	2015	2016	2020
Data workers (Million unit)	6.00	6.16	10.43
Data companies	249100	255000	359050
Data companies revenues (billion EUR)	€56	€61	€130
Data market (billion EUR)	€54	€59	€106
Data workers skill gaps	393000	420000	769000

As seen from the table the total number of data workers, data companies, the value of the overall data market and the data worker's skill gap have increased in number in the EU from 2015 to 2016 and are projected to continue their growth throughout 2020.

The data economy is a global phenomenon, not only are investments in data-related technologies, but data itself is a production factor than can be easily transferred other region or country of the world. The key indicators of the data market in the countries (the ICT-oriented regions) as the USA, Japan and Brazil in 2016 were given at the following table (Table 3.). According to assessment of 2016, a number of data workers, a number of data companies, indicators of revenues of data companies in the USA is higher than any other country, including Europe.

Table 3. Worldwide Data Market Indicators in 2016

Indicators	EU	USA	Brazil	Japan
Data workers (Million unit)	6.16	12.732	1.160	3.740
Data companies	255000	28 9556	35979	101612
Data companies value (billion EUR)	€61	€129	€ 6	€25
Data market (billion EUR)	€59	€129	€ 6	€25

The digital skills gap is one of the most urgent policy challenges facing the developed, and also developing countries. Unlike conventional analytics, mining Big Data requires an extremely diverse set of skills like data visualization, statistics, machine learning, deep business insights and computer programming. Estimates shows that the USA alone faced a shortage 190,000 people in 2016.

The market value of Big Data is measured in accordance with revenues from the sale of hardware, software and ICT services [24].

The increase of Big Data and the application in various fields has led to the expansion of scientific researches. In scientific sources, the number of published documents

related to Big Data can be viewed as an indicator of the Big Data measurement. Big Data research has begun in 2001 with one published document in Scopus [25]. In recent years, published Big Data documents have increased, in 2014 only in this base the number of published Big Data documents have reached to 3472. In order to number of these documents the USA, China, Germany, Great Britain and India were the first place [26, 27].

E. Data Quality Indicators

High-quality data are the precondition for analyzing and using Big Data and for guaranteeing the value of the data. In [28] some detailed data quality indicators are given. Authors chose data quality dimensions such as availability, usability, reliability, relevance, and presentation accepted and widely used as Big Data quality standards. At the same time, each dimension was divided into many typical elements associated with it, and each element has its own corresponding quality indicators.

Many of the authors show that the quantification of information does not necessarily say anything about the quality or value of this information. In fact Information quantity is not equal to information quality or information value. However, per definition of “value of information” or “quality of information,” any quantifiable measure of value and quality will first of all suppose a quantifiable measure of information [29].

F. Big Data Indicators for Social Measurement

The complication of the forms of organizing social actions in the network requires special online methods of research of virtual space. In Big Data, many parameters of the functioning of the Internet environment are taken into account (frequency of visits, the click map, file uploads, replies to the forum and so on). The overall structure of these indicators is presented in [30] and proposed Online Big Data indicators for social measurement:

- A number and list of sites / pages viewed.
- Frequency of access to the Internet as a whole or visiting a particular site.
- Time of day of the greatest activity.
- The growth rate of the number of web servers.
- A number of domains registered in the Network.
- The amount of information resources provided on the Internet.
- The size of the Internet-audience -a number of Internet users in the world, by country.
- A number of households with high-speed Internet access.
- The proportion of wireless users in the Internet audience.
- Frequency and volumes of Internet use (hours per week).
- And so on.

G. Big Data for Monitoring the Information Society

The growth of ICT has resulted in a rapid increase of new data sources, in particular from the ICT industry. ITU is looking into innovative ways to utilize Big Data as a new data source, and to overcome important data gaps. To this purpose, in 2016 ITU recently launched a project on “Big Data for Measuring the Information Society [31]. Within the project proposed Big Data indicators for measuring the Information Society:

- Percentage of the land area covered by mobile-cellular network, by technology;
- Percentage of the population covered by a mobile-cellular network, by technology;
- Usage of mobile-cellular networks for non-IP related activities, by technology;
- Usage of mobile-cellular networks for internet access, by technology;
- Number of subscriptions with access to technology;
- Active mobile voice and broadband subscriptions, by contract type;
- Average number of active mobile subscriptions per day, by contract type;
- Active mobile devices;
- IMEI conversion rate;
- Fixed domestic broadband traffic, by speed, contract type;
- Mobile domestic broadband traffic, by speed, contract type, technology;
- Mobile international broadband traffic, by contract type;
- Inbound roaming subscriptions per foreign tourist;
- Fixed broadband subscriptions, by technology;
- Fixed broadband subscriptions, by speed;
- Any proposed indicators from the country stakeholders.

Under this project, ITU will apply a number of pilot studies in designated countries and produce novel and policy-relevant ICT statistics with Big Data.

IV. FACTORS CONTRIBUTING TO THE GROWTH AND DEVELOPMENT OF BIG DATA

Big Data covers technologies and tools for collecting, processing, analyzing and extracting useful knowledge from structured and unstructured data of large volumes generated at high speed by different sources. Like the physical universe, digital is diverse. It created by everyone using a digital camera, by the more than 2 billion people and millions of enterprises living their lives and doing their works online, and by the millions of sensors and communicating devices sending and receiving data over the Internet. The OECD relates data growth to the following factors [32]:

- 1) The increasing ubiquity of broadband access and the proliferation of smart devices and smart ICT applications;
- 2) The large minimizing in Internet access costs over the last 20 years;
- 3) The large reducing in storage costs. The storage costs have decreased to the point at which data can generally be kept for long periods if not indefinitely;
- 4) The data processing tools have become increasingly powerful, sophisticated, ubiquitous and inexpensive, making data easily searchable, linkable.

In general, the following information infrastructure indicators can be used to assess the factors affecting the growth and development of Big Data:

- A number of Internet users;
- Worldwide per capita information;
- A number of devices connected to the Internet;
- A number of mobile devices connected to the Internet;
- A number of mobile phone users;
- A number of Internet-connected enterprises;
- Productivity of large data-related devices (storage, computing, generating);
- Input speed of broadband internet network (mbit / sec);
- A number of data centers;
- And so on.

In 2012 the number of Internet users in the world was 2.5 billion, per capita volume of information was 369 GB (2.8 zettabytes by worldwide). According to IDC forecast the number of Internet users will reach 4.1 billion in 2020 and the per capita volume of information will be 5200 GB [1].

Businesses (consumers) who have access to the Internet in the field of using of Big Data for new purposes (to learn about customers, speed business cycles, flatten organizational structures and so on) acquire more opportunities. Large application of Big Data in the fields as commerce, banking, governance, science, health and etc. has also led to an increase in the volume of data. Healthcare is one of the fastest-growing segments of the Digital universe. If the volume of data generated in the healthcare was 153 Exabyte in 2013, forecast that it will be 2314 Exabyte in 2020 (annual growth 48%). This is more than Digital Universe's annual rate of growth (40% per year for the overall Digital Universe) [1].

Expanding to include not only the increasing number of people and enterprises doing everything online, but also all the “things” – smart devices – connected to the Internet, unleashing a new wave of opportunities for businesses and people around the world.

While not all “things” are connected to the Internet, 20 billion of them were in 2013, and will be 32 billion by 2020. The network connecting devices in the Internet of Things is characterized by automatic provisioning, management, and technology. It includes Intelligent

systems and devices, Connectivity enablement, Platforms for device, network, and application enablement, Analytics and social business, Vertical industry solutions. IoT is growing over three times as fast as traditional ICT, and by 2020 will nearly equal all other ICT spending. Mobile “Connected Things” generate of 18% of the Digital Universe in 2014. In 2020, will grow to 27% generated by Mobile “Connected Things” [1].

The Internet of Things will generate a staggering 400 zettabytes (ZB) of data per year by 2018, up from 113.4ZB per year in 2013, according to a report from Cisco. This will still cause the total amount of traffic sent to data centers to grow to 8.6ZB, up from 3.1ZB in 2013. [33]

According to Cisco’s forecast the number of devices connected to IP networks will be more than three times the global population by 2021. There will be 3.5 networked devices per capita by 2021, up from 2.3 networked devices per capita in 2016. There will be 27.1 billion networked devices in 2021, up from 17.1 billion in 2016. Broadband speeds will nearly double by 2021. By 2021, global fixed broadband speeds will reach 53 Mbps, up from 27.5 Mbps in 2016 [34].

Accruing the volume of data in the world causes the arising of new data center. As the demand increases for cloud services these centers include dozens, sometimes hundreds of thousands of servers, saving systems by petabytes. So, if 4 % of all data in 2010, 16 % of all data in 2012 was held in the clouds, in 2020 this figure is expected will organize 37 %. According to information of IDC consulting company, the number of data center in the world in 2015 amounted to 8.55 million. In [35] are given Inside Ten of the World’s Largest Data Centers.

In order to leverage the potential of Big Data, a key challenge is to ensure the availability of highly and rightly skilled people.

Double increasing of digital data in every two years enhanced the demand of cadres, qualification called “Data Scientist”. Worldwide in 2014: A number of IT specialists were 28 million; volume of information per IT specialist was 230 GB. In 2020 it is expected that a number of IT specialists will be 36 million, volume of information per IT specialist will be 1231 GB [1].

The technologies that used for the collection and processing of Big Data can be divide to 3 places: software, devices (equipment) and services.

The digital universe is created and defined by software that analyzes this expanding universe of digital data, creates new opportunities to extract value from the digital universe that have created. The most widely technologies for software contain SQL, NoSQL, MapReduce, Hadoop, SAP HANA [36, 37]. According to information of IDC the volume of Big Data world market amounted to 21.3 billion dollars in 2015. According to information of IDC company related to application of different fields of the industry, in 2019 the volume of Big Data technologies and services will be increased to \$48.6 billion in the world [1].

According to the researches, the following indicators can be proposed for measuring of Big Data.

1) *Information Infrastructure:*

- A number of Internet users;
- Worldwide per capita information;
- A number of devices connected to the Internet;
- A number of mobile devices connected to the Internet;
- A number of mobile phone users;
- A number of Internet-connected enterprises;
- Productivity of large data-related devices (storage, computing, generating);
- Input speed of broadband internet network (mbit / sec);
- A number of data centers;
- And so on.

2) *Innovative Factor:*

- A number of research centers associated with Big Data;
- A number of large data-related studies;
- A number of scientific publications related to Big Data;
- A number of patents associated with Big Data;
- A number of invested Big Data projects;
- Volume of investment in Big Data projects;
- And so on.

3) *Human Capital Factor:*

- A number of enterprises (universities, colleges) teaching Big Data;
- A number of data science programs;
- A number of graduates in science, technology, engineering and mathematics related to data;
- Courses related to data;
- And so on.

4) *Economic Factor:*

- Number of businesses offering Big Data products and services;
- Volume of investment in Big Data related businesses;
- Volume of revenues of businesses associated with Big Data;
- Big Data market volume (software, hardware, services);
- Number of jobs associated with Big Data;
- And so on.

The presented list of indicators is not complete, but only opens up the horizons of possible classification.

For this purpose, it may be expedient to carry out the work on toward a harmonization of the different methodologies. It is suitable to set up an institutional mechanism to regularly accumulate imperative and influential indicators, and harmonized methodologies are certainly being required in order to do so.

V. DISCUSSION

The development of modern IT technologies with the rapid growth of Big Data and its application in different spheres of human activity has also led to the assessment of its current state (by the world and countries) along with its volume.

According to researches, numerous investigations have been carried out in the field of measurement of large volumes of data. It proves that the quantification of information and communication presents a theoretical, methodological, as well as statistical challenge. A few researchers have studied the total amount of data generated, stored, and consumed around the world. On the other hand, the scope of their predictions and accordingly their results vary. Different studies come up with largely varying numbers. For instance, Hilbert and Lopez report that the amount of globally communicated amount of information amount up to 1.15 zettabytes in 2007, while Bohn and Short report that only one year later, in 2008, American's alone consume more than 3.6 zettabytes. The reason for these differences is methodological nature. Some of the methodological distinctions between approaches are rooted simply in difference in the research focus.

Also, the measurement methodologies of large volume of data are not generally disclosed. For example, the IDC company has been conducting researches in the field of measuring of Big Data by the world and in many countries for over a decade. However, its measurement methodology is not fully disclosed.

All surveys, each with varied methodologies and definitions, adopt on one fundamental matter—the amount of data in the world has been expanding rapidly.

As usual, the chosen measurement indicator is defined by the selected theoretical framework and the particular exploration question on the researcher's mind.

Some indicators for measurement of Big Data (measurement of the Big Data ecosystem, Data market indicators, Big Data indicators for social measurement, etc.) have been included. As seen from the investigations generally accepted indicators systems for measurement of Big data are not still available.

International organizations have begun to use Big Data as a new data source, and to overcome important data gaps for improve the quality of the official statistics. As noted above, Big Data Indicators have been proposed to measure the Information Society by the ITU. Also, by other countries around the world have been provided with the opportunity to offer new Big Data indicators. These indicators can allow to measure various aspects of Big Data application.

VI. CONCLUSION

Information is the key resource and development factor of information society in the modern world. Rapid

development of this field in many other developed and developing countries indicates Big Data's potential capabilities. However, there is still no system of indicators that allows to find out how Big Data influence the socio-economic development of the society and to assess the current state of the data. First of all, this can be related to the lack of unambiguous adopted definition of Big Data, and Big Data theory model to describe the current relationships and concepts. That's why there is a need for serious research in this area.

REFERENCES

- [1] The Digital Universe of Opportunities: Rich Data and the Increasing Value of the Internet of Things, 2014, <https://www.emc.com/leadership/digital-universe/>
- [2] Big Data, Big Impact: New Possibilities for International Development, 2012, <http://www.weforum.org/reports>
- [3] X. Jina, W Benjamin, X, Chenga, Y.Wanga, "Significance and Challenges of Big Data Research", *I.J. Big Data Research*, vol. 2(2), pp. 59–64, 2015.
- [4] Oxford Dictionaries: www.oxforddictionaries.com/definition/big-data, 2015.
- [5] H. Hu, Y. Wen et al., "Toward Scalable Systems for Big Data Analytics", *IEEE Access Journal*, vol. 2, 2014, pp. 652-689.
- [6] P. Deep Kaur, A. Kaur and S. Kaur, "Performance Analysis in Bigdata", *International Journal of Information Technology and Computer Science (IJITCS)*, 2015, vol.7, no.11, pp. 55-61. DOI: 10.5815/ijitcs.2015.11.07
- [7] Gartner. IT Glossary Big Data, 2014, <http://www.gartner.com/it-glossary/big-data/>
- [8] D. Boyd, K. Crawford, "Critical Questions for Big Data", *I.J. Information Communication & Society*, vol.15, no.5, pp. 662-679, 2012.
- [9] J. Gantz and D. Reinsel, "Extracting value from chaos", in *Proc. IDC iView*, 2011, pp. 1-12
- [10] J. Manyika, M. Chui, B. Brown et al., "Big Data: The Next Frontier for Innovation, Competition, and Productivity", San Francisco, CA, USA: McKinsey Global Institute, 2011, pp. 1-137.
- [11] M. Cooper, P. Mell, *Tackling Big Data*, 2012, <http://csrc.nist.gov/groups/SMA/forum/documents/>
- [12] UN Global Pulse, "Big Data for Development: Challenges & Opportunities", 2012, <http://unglobalpulse.org>
- [13] UN Global Pulse, "Integrating Big Data into the Monitoring and Evaluation of Development Programmes", 2016, <http://unglobalpulse.org/>
- [14] M. Pospiech, C. Felden, "Towards A Big Data Theory Model", *Proceedings 2015, IEEE International conference on Big Data*, 2015, pp. 2082-2090.
- [15] P. Lyman, and Hal R. Varian, "How Much Information", 2003. <http://groups.ischool.berkeley.edu/>
- [16] J. Gantz et. al., *The Diverse and Exploding Digital Universe: An Updated Forecast of Worldwide Information Growth Through 2011*, IDC White Paper, March 2008.
- [17] R. Westervelt *IDC White Paper: Information-Centric Security: Why Data Protection Is the Cornerstone of Modern Enterprise Security Programs*, March 2017, symantec.com/content/dam/
- [18] M. Hilbert, H. Lopez, "The World's Technological Capacity to Store, Communicate, and Compute", *Information Science*, 2011, vol.332(6025), pp. 60 –65.
- [19] Cisco Visual Networking Index: Forecast and Methodology, 2016–2021, <http://www.cisco.com>
- [20] R. Bohn, J. Short, "How much information? 2009 report on American consumers", *Global Information Industry Center of University of California*, 2009, <http://hmi.ucsd.edu/howmuchinfo.php>
- [21] 2014 The Massachusetts Big Data Report: A Foundation for Global Leadership, <http://www.masstech.org/sites/mtc/files/documents/Full-Report-2014-Mass-Big-Data-Report.pdf>
- [22] *Massachusetts Big Data Indicators 2015*, <http://massbigdata.org/assets/Uploads/Final-Big-Data-Report-2015.pdf>
- [23] G. Catteneo, "The European Data Market", *NESSI summit in Brussels on 27 May 2014*, <http://www.nessi-europe.eu/>
- [24] Final results of the European Data Market study measuring the size and trends of the EU data economy, 2017, <https://ec.europa.eu/digital-single-market/en/news/>
- [25] G. Halevi, "The Evolution of Big Data as a Research and Scientific Topic Overview of the Literature", *Research Trends*, Issue 30, 2012, <https://www.researchtrends.com/wp-content/>
- [26] M. Hajirahimova, A. Aliyeva, "Some indicators of Big Data", *IOSR Journal of Engineering (IOSRJEN)*, 2016, vol. 06, Issue 10, pp. 01-06
- [27] R. M. Aliguliyev, M. Sh. Hajirahimova, A. S. Aliyeva, "Current scientific and theoretical problems of Big Data", *Problems of information society*, 2016, №2, 34–45.
- [28] L. Cai, Y. Zhu, "The Challenges of Data Quality and Data Quality Assessment in the Big Data Era", *Data Science Journal*. 2015, vol.14, p.2.
- [29] M. Hilbert, "How to Measure "How Much Information"? Theoretical, Methodological, and Statistical Challenges for the Social Sciences", *International Journal of Communication*, 2012, vol.6., pp.1042–1055
- [30] N.V Korytnikova, "Online Big Data as a source of analytic information in online research", *Sotsiologicheskie Issledovaniya*, 2015, Issue 8, pp. 14-24.
- [31] Big Data for Measuring the Information Society, 2016, <http://www.itu.int/net4/ITU-D/CDS/projects/display.asp?ProjectNo=2GLO16081>
- [32] OECD, "Exploring data-driven innovation as a new source of growth: Mapping the policy issues raised by "big data", *Supporting Investment in Knowledge Capital, Growth and Innovation*, OECD Publishing, 2013, DOI:<http://dx.doi.org/10.1787/9789264193307-12-en>
- [33] *Internet of Things to generate 400 zettabytes of data by 2018*, v3.co.uk
- [34] *The Zettabytes Era: Trends and Analysis*, 2017, www.cisco.com
- [35] *Inside Ten of the World's Largest Data Centers*, 2010, <http://wikibon.org/blog/inside-ten-of-the-worlds-largest-data-centers/>
- [36] B. Jena, M. Kumar Gourisaria, S. Swarup Rautaray, M.Pandey, "A Survey Work on Optimization Techniques Utilizing Map Reduce Framework in Hadoop Cluster", *International Journal of Intelligent Systems and Applications (IJISA)*, vol.9, no.4, pp.61-68, 2017. DOI: 10.5815/ijisa.2017.04.07
- [37] M. Abdrabo, M. Elmogy, G. Eltaweel, Sh. Barakat, "Enhancing Big Data Value Using Knowledge Discovery Techniques", *I.J. Information Technology and Computer Science*, 2016, vol.8, pp. 1-12, <http://www.mecspress.org/>, DOI: 10.5815/ijitcs.2016.08.01

Authors' Profiles

Makrufa Sharif Hajirahimova is an associate professor, project chief engineer at the Institute of Information Technology of Azerbaijan National Academy of Sciences. She teaches at the Training Innovation Center of the institute. She defended the thesis on the "Development of models and algorithms for intelligent management of text

document in e-government" at the Institute of Information Technology of Azerbaijan National Academy of Sciences and received PhD in Technical sciences in 2013. Currently she conducts research on the "Big Data", and is actively involved in the development of the "e-Science" software project under the "E-Azerbaijan" program.



Aybeniz Salman Aliyeva is a researcher, a senior researcher at the Institute of Information Technology of Azerbaijan National Academy of Sciences. Her research interests are Big Data technologies. She has published many papers in international journals and conferences.

How to cite this paper: Makrufa Sh. Hajirahimova, Aybeniz S. Aliyeva, "About Big Data Measurement Methodologies and Indicators", International Journal of Modern Education and Computer Science(IJMECS), Vol.9, No.10, pp. 1-9, 2017.DOI: 10.5815/ijmecs.2017.10.01