PANEVĖŽIO KOLEGIJA ISSN 2029-1280. Taikomieji tyrimai studijose ir praktikoje – Applied Research in Studies and Practice, 2016, 12.

THE SOCIAL AND ETHICAL AFTERMATH OF BIG DATA

Anna Waligórska-Kotfas

Państwowa Wyższa Szkoła Zawodowa w Koninie

Abstract. The revolution in information and communication technologies induced the exponential growth of produced and acquired, processed, handled and consumed data. Epistemic foundations of the knowledge hierarchy outlined in the first part of the article give the footing to the reflections on the Big Data phenomenon, its attributes, architecture and practical implementations. The last part of the paper is devoted to the social and ethical menace behind Big Data which may lead to reinforcement of digital divide and discrimination. In conclusion a need of public education and open discussion regarding the notion of Big Data is postulated.

Key words: knowledge pyramid; data processing; information ethics; Big Data.

INTRODUCTION

Technological transformations of the 20th century have brought extensive socio-economical changes referred to as the *third wave* (Toffler, 1997), *knowledge society* (Drucker, 2001), *network society* (Castells, 2013). After the industrial society based on manufacturing, and the post-industrial one, in which a key role has been played by the service sector, time has come for the information society where information has become a special intangible asset, frequently more valuable than material possessions. The present-day society founded on acquisition, retrieval, distribution, and manipulation of data and information and consequently on creation of both explicit and tacit knowledge, is universally recognized as giving the footing to the knowledge-based economy (KBE) which in turn becomes more interdependent on a current phenomenon by the name of *Big Data*.

The reflections on the topic of this paper are focused on Big Data as the major instrument of today's data processing in order to obtain information and knowledge. The article begins with a brief description of the knowledge hierarchy. Subsequently, the Big Data phenomenon is discussed focusing on its promise and peril from the social, economic, ethical, and ontological perspective. The paper is elaborated on English and Polish literature studies available both in print and online as well as statistical data and research.

THE HIERARCHY OF KNOWLEDGE

Epistemologically, *knowledge* as a key resource of today's organizations is created by processing, interpretation and interconnection of *information*; which in turn is based on *data*, raw facts, signs and symbols that by categorization, condensation, contextualization and correction become purposeful and relevant. The pinnacle of the hierarchy is *wisdom* recognized as accumulated knowledge allowing to To describe method in annotation; to formulate research problem and to make references to some information sources in the introduction; to insert an official data sources used in the references list; to correct some illustrations design (figure 4). act critically in new situations and interconnected with ethical judgement (Rowley, 2007). The review of selected definition of the terms *data, information,* and *knowledge* is presented in table 1.

Historically, T.S. Eliot is regarded in the literature discourse as the 20th century precursor of the concept of the knowledge hierarchical structure since he contemplated the following question: "Where is the wisdom we have lost in knowledge? Where is the knowledge we have lost in information?" (Eliot, 1991, p. 147). Subsequently, the knowledge hierarchy founded on the *data* → *information* → *knowledge* → *wisdom* scheme was formulated by M. Zeleny (1987), and R. Ackoff (1989) completed it with the fifth degree, *understanding*, which was positioned between *knowledge* and *wisdom*, while stating that each subsequent level is founded on the former one and emerges out of it. The criticism of the Ackoff's concept of the five-level knowledge pyramid can be found in the work of G. Bellinger, D. Castro and A. Mills (2004), who argued that the understanding cannot be seen as a separate tier since its inextricable linkage to each stage of cognition. Therefore, the following cognitive chain with the corresponding processes was proposed: *data* →

(understanding relationships) → *information* → (understanding patterns) → *knowledge* (understanding principles) → *wisdom*.

Table 1

The review of selected definitions of *data*, *information*, *knowledge* (Source: own elaboration based on (Ackoff, 1989; Zeleny, 1987; Davenport and Prusak, 1998; OECD, 2000; Kisielnicki and Sroka, 2005; Hey, 2004; Materska, 2007; Robertson, 2013; Stefanowicz, 2013)

Autor	*Data*	*Information*	*Knowledge*
Zeleny (1987)	Know-Nothing	Know-How	Know-What
Ackoff (1989)	Symbols representing the properties of objects or events	Processed data containing descriptions and answers to who, what, where, how many questions	Instructions answering how-to questions
Davenport, Prusak (1998)	A set of discrete, objective facts about events	A message meant to change the way the receiver perceives something. It has meaning and is organized to some purpose	A mix of framed experience, values, contextual information, and expert insight that provides a framework for evaluating and incorporating new experiences and information
OECD (2000)	The actual state of the world	Indicators that are accessible to the agents representing the state of the world)	Both an input (competence) and output (innovation) in the production process
Kisielnicki, Sroka (2001)	Can be processed using computer equipment	Resource which allows to incur To describe method in annotation; to formulate research problem and to make references to some information sources in the introduction; to insert an official data sources used in the references list; to correct some illustrations design (figure 4).ease the level of knowledge about the surrounding world	The source of all action performed by intelligent people. Interdependent facts and the strategy for solving problems
Hey (2004)	A resource. Manipulable Objects. A solid, physical, thing with an objective existence	Can be processed and accessed, generated and created, transmitted, stored, sent, distributed, produced and consumed, searched for, used, compressed and duplicated	Personal, subjective and inherently local. Internalized by the knower, and as such is 'shaped' by their existing perceptions and experiences
Materska (2007)	Facts, signs, or observations recorded "on" or "in" medium	Data in relation to the specified context organized according to certain categories	The translation of complex information structures into new performance contexts. Intuitive, so that it is complicated to define and analyze
Robertson (2013)	Depending on the research methods data can be inductively used to form conclusions or deductively taken from the conclusions	The intellectual content of commodified and reified documents	Justified true belief
Stefanowicz (2013)	Elements currently stored by the respective characters	The relationship existing between single elements of the message	Simultaneous consideration of information, context and experience

Additionally, T. Davenport and L. Prusak (1998) enriched *the knowledge pyramid* adding five methods of converting *data* into *information*, such as contextualization, categorization, calculation, correction and condensation. In the ongoing dispute, the knowledge hierarchy was supplemented by further indicators related to *meaning* and *value* or *programmability*, as well as the levels, namely *enlightenment* or *truth*. Some researchers postulated that the individual steps of the hierarchy should be regarded as the continuum rather

than the separate parts, while J. Rowley (2007) turned the knowledge pyramid by 180 degrees creating *the knowledge funnel*, where data is recognized as an input whilst knowledge is an output, so that the overall cognitive construction is teetering on knowledge and is about to collapse without it.

Despite the diversities in the study literature there is a universal agreement on the following features: firstly, the key structure is DIKW scheme, and its elements are arranged in the same order; secondly, the succeeding tier is explained by the previous one with the implementation of the appropriate transformation pattern; finally, the predominant issue is to comprehend and clarify the processes occurring between the consecutive layers. The knowledge hierarchy, or DIKW (The acronym comes from Data → Information → Knowledge → Wisdom) pyramid, is presented in picture 1.

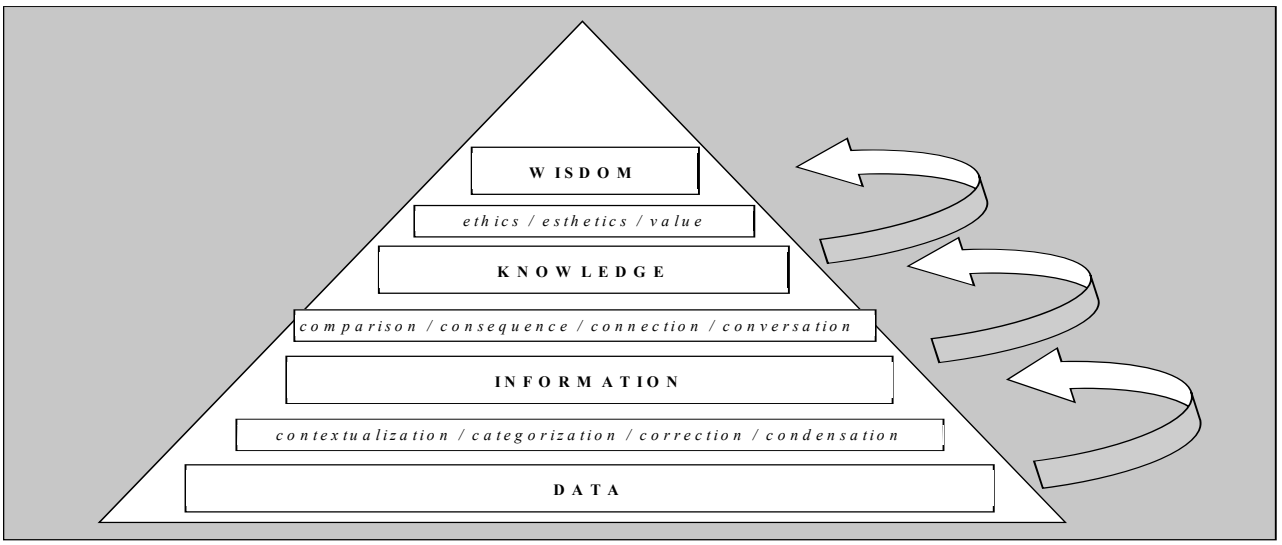

Figure 1. **DIKW Pyramid and the Processes between its Levels** (Source: own elaboration based on (Ackoff, 1989; Davenport and Prusak, 1998; Zeleny, 1987)

THE BIG PHENOMENON OF BIG DATA

The revolution in information and communication technologies (ICTs) has implied the revaluation of the classical paradigm of the socio-economical realms towards information society (IS) and knowledge-based economy (KBE), where the informational mode, in which data and information processing together with knowledge generation and diffusion by means of multimodal, flexible, and networked communication, is both the main form of social interaction and the fundamental source of productivity and competitive advantage (Castells, 2013).

The 21st century paradox of IS and KBE lies the fact that the development of ICTs allows almost instant access to data and information, their creation and transmission, production and consumption, but parallels it generates the abundance of data and triggers an abnormality recognized as information explosion leading to information overload (Ruff, 2002; Eppler and Mengis, 2004). The problem is clearly illustrated in a series of studies under the common name Digital Universe prepared by IDC enterprise on behalf of EMC Corporation. The ongoing research in which the amount of data generated each year is being measured gives the footing to formulate forecasts. The seventh report, published in 2014 and entitled The Digital Universe of Opportunities: Rich Data and the Increasing Value of the Internet of Things shows the increasing importance of wireless technologies, mobile devices, intelligent products and businesses defined by the software in the data output (IDC, 2014).

The development of digital and mobile devices together with improved, accelerated and predominantly wireless Internet access, the expansion of interactive communication through social media, the universality of city monitoring, and the Internet of Things (Internet of Things (IoT) refers to a global infrastructure for the IS, enabling advanced services by interconnecting (physical and virtual) things based on existing and evolving interoperable information and communication technologies (ITU, 2012) catalyze the double upsurge in the size of the digital universe every two years. Therefore, between 2013 and 2020 the amount of produced and consumed data is to increase tenfold and will reach the level of 44 ZB (Zettabyte equals 10^{21} bytes (1,000,000,000 TB), which translates to more than 5 TB (5,000 GB) for each person on Earth. According to IDC, the number of devices or objects that can be computerized and connected to the

Internet is approaching 200 billion, and 7% of them (14 billion) are already interacting online. The data generated by these devices currently represents 2% of the world data and the IDC's predictions state that by 2020 the number of network-connected devices will have multiplied to reach the level of 32 billion and will have produced 10% of the world data.

The exponential growth of data together with the increase in the computing power, the research on artificial intelligence, and data mining has led to the creation of tools to analyze large, diverse, variable, and often unstructured data sets, referred to as Big Data, in order to transform them into purposeful information and consequently to interconnect and interpret into knowledge. The most concisely the term of Big Data is defined as datasets that cannot be processed or analyzed using traditional processes or tools (Zikopoulos et al., 2012). To put it differently, the data growth has multiplied sharply to such extend that conventional database management tools or data processing tools have become inefficient to handle those complex sets of data. The ambiguity and subjectivity of the definition is implied by the constant technology advancement over time which in turn results in assumption that the size of datasets that could be qualified as Big Data will parallelly increase (McKinsey Global Institute, 2011).

Big Data, as a process of analyzing data, is characterized by increasing number of parameters. Currently the literature discourse mentions four dimensions (V4) distinguished in Big Data description, namely volume, variety, velocity, veracity. Volume interrelates the amount of data currently counted in petabytes or exabytes (Petabyte equals 10^{15} bytes (1,000 TB), Exabyte equals 10^{18} bytes (1,000,000 TB). Variety refers to the semantic heterogeneity represented by numerous sources and types of data both structured and unstructured, not only from the traditional sources such as spreadsheets and databases, but also coming in the formats of emails, photos, videos, PDFs, audio, websites, tweets, blogs, comments, tags, SMS messages, WiFi location tracking, etc. Velocity corresponds real-time data processing, which can in turn provide the strategic competitive advantage. Veracity concentrates on data relevance in order to avoid bias, noise and abnormality in data (Hitzler and Janowicz 2013). The Big Data dimensions are illustrated in figure 2.

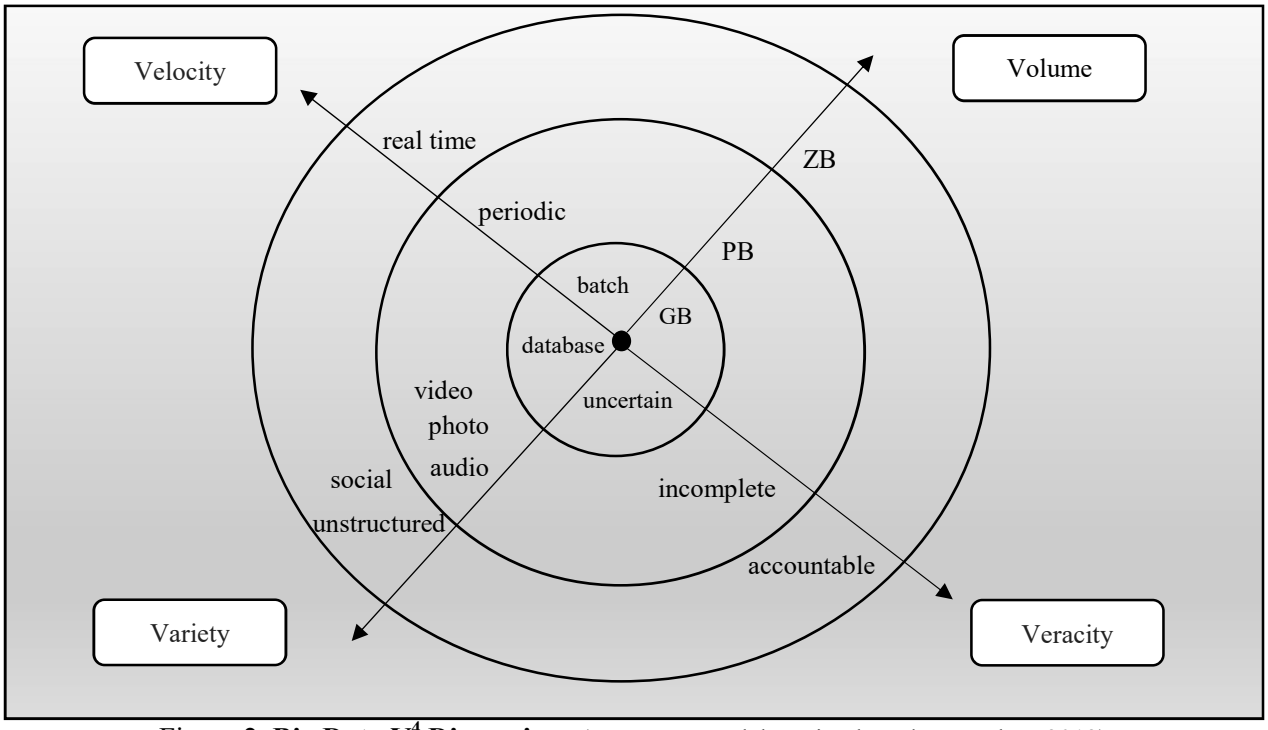

Figure 2. **Big Data V[4] Dimensions** (Source: own elaboration based on Soubra, 2012)

From the broader, not only technological, but interdisciplinary perspective Big Data is the coexistence and mutual impact of technology, analysis and mythology. The technology aspect is realized in maximizing computation power together with algorithmic preciseness to acquire and process, interrelate and correlate, parallel and juxtapose data. The analysis component focuses on identifying patterns in order that economic, social, technical, and legal claims could be made. The mythology facet lies in the widespread belief that large data sets offer a higher form of intelligence and knowledge which can generate insights that were previously impossible, with the aura of truth, objectivity, and accuracy (Boyd and Crawford, 2012).

The Big Data pipeline architecture, presented in figure 3, consists of five stages such as: acquisition, extraction, aggregation, analysis, and interpretation. Data collection and recording is performed regardless the sector or industry and can include: the IP address of the computer; online activity tracked by cookies and stored by the browser; keywords and passwords entered in the search engines; information voluntarily shared in social media (circle of friends, recommended or popular content or location, uploaded photos and videos), payments and financial transactions, health conditions, telephone calls, shopping receipts, and geolocation coordinates. Consecutively, data needs to undergone the extraction process that pulls out the required bits from the underlying sources and expresses it in a structured form suitable for analysis. Consequently, the integration and representation is required to enable data structure and semantics to be computer understandable, and then automatically processed. Furthermore, analysis and modeling are implemented to obtain general statistics and to examine linkages collected from frequent patterns in order to disclose hidden correlations and convert them into knowledge. The phenomenon of Big Data interpretation embraces the fact that the large heterogeneous and interconnected information datasets are generated so that information redundancy can be explored to compensate for missing data, to crosscheck conflicting cases, to validate data trustworthiness, to disclose fundamental clusters, and to uncover unknown relationships and models.

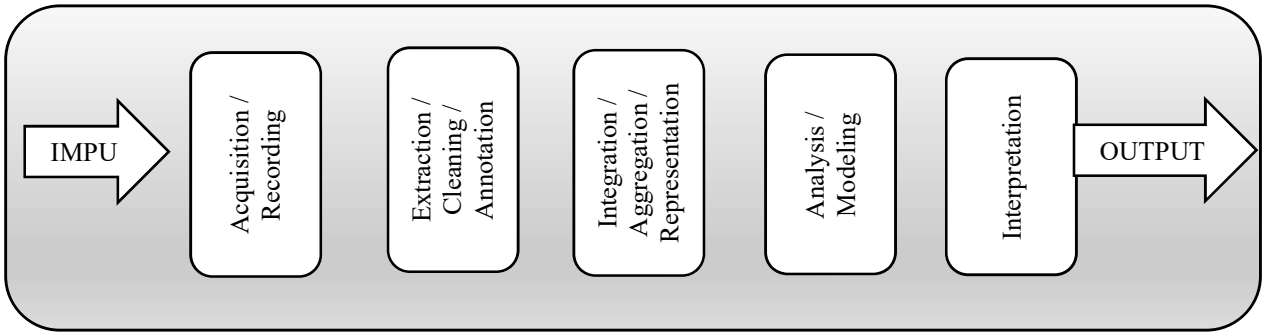

Figure 3. **The Big Data Pipeline Architecture** (Source: own elaboration based on Labrinidis and Hosagrahar, 2012)

THE PROMISE AND PERIL OF BIG DATA

data are used to identify stolen vehicles in real time, which in turn translates into reduction of illegal behavior since many crimes are committed in stolen automobiles (Jewell et al., 2014). A search engine tracks the incidence of flu-related search terms and can identify possible flu outbreaks one to two weeks earlier than official health reports (Bollier, 2010).

Notwithstanding the enormous potential benefits of using Big Data, the other side of the coin needs to be recognized, particularly in the light of information ethics (IE). Among the IE research areas; there are three fields of ethical dilemmas corresponding to the production of information, its classification, and finally the access and dissemination of information (ICIE, 2015). Information is considered in the IE conceptualization as:

– a resource (input) – moral issues arising from *the triple A*: *availability*, *accessibility*, *accuracy* of informational resources, independently of their format and physical representation; additionally, from the information consumer's point of view the questions concerning reliability, trustworthiness, and relevance of information sources need to be raised here;

– a product (output) – ethical problems regarding pragmatic rules of communication such as: accountability, liability, libel, plagiarism, advertising, propaganda, misinformation;

– a target (environment) – the dilemmas over information security, its vandalism, piracy, and hacking, intellectual property rights, freedom of expression, and censorship, filtering, and contents control; the social dimension includes the digital divide and the ICTs illiteracy (Floridi, 2006).

In view of IE theory, the ethical consideration facing Big Data regarded as both the input and the environment should include the fact that although the data for analysis is collected anonymously and from legitimate sources with law-restricted access, one can doubt whether Internet users consciously accept the rules and privacy policy of the services they use. Furthermore, the ongoing controversy of so-called

public data in the social media sites is the battleground between privacy campaigners and transparency supporters. Additionally, acquired data may not be adequately protected and therefore can leak, be traded or stolen. On top of that, since anonymity in the Internet is apparent, even aggregated data can be sensitive and may unleash surveillance, or even discrimination, against citizens by both the state and corporations.

The ethical dilemmas interrelated the Big Data adoption are illustrated by the morally questionable examples. The first of them relates to a supermarket chain which, in order to analyze its customers' shopping behaviors, implemented software allowing pregnancy prediction score to be established by connecting the purchase to Customer ID number, their credit card, and email address and consequently the product offer was adjusted. Therefore, even though the customer does not inform anyone, let alone the retail outlet, about their condition, the retail network has information on the subject (Hill, 2012). The second case is of a drugstore chain which applied an algorithm to predict whether an individual suffers from one of 17 diseases, including diabetes, tobacco related cancer, cardiovascular disease, and depression based on analyzing his or her buying habits. Accordingly, a person's life expectancy is estimated and insurance rates may increase (Marr, 2015).

Regarding Big Data as the output brings the question of data analysis and interpretation. The process of deciding which data attributes and variables are essential and which could be ignored together with the algorithms applied in computerized data sifting tend to be error-prone so that patterns are identified where none are existing. The noticeable instance of that is the case of Jeff Bezos, who demonstrated the Amazon recommendation engine in front of an audience. To validate its efficiency and practicality he accessed his own set of recommendations to surprisingly notice the first recommendation as *Slave Girls from Beyond Infinity,* a choice triggered by Bezos's purchase of a DVD of *Barbarella,* a classic Jane Fonda movie, the previous week (Bollier, 2010).

The social approach to the Big Data menace is correlated with creation of the digital divide. Despite the fact that common perception of Big Data is that it offers easy access to massive amounts of data, in real circumstances the access is not straightforward and frequently limited by funds. Additionally, the advantaged are those with extensive and sophisticated computational skills. Furthermore, Big Data ensues a transformation in a social stratification system from the hierarchy founded on socioeconomic variables to the one established on the modality of the digital competence since within its realm are "those who create data (both consciously and by leaving digital footprints), those who have the means to collect it, and those who have the expertise to analyze it" (Manovich, 2011, p. 10). Albeit the ultimate class is the smallest, it is the most privileged and endowed with power to determine the rules about the use of Big Data and participants of the process (Boyd and Crawford, 2012). Therefore, the divide into proletariat, digitariat, and cogitariat is reinforced.

From the ontological perspective the peril of Big Data is induced by the question whether a scientific construction of theories and models, particularly in social sciences, might be discarded and replaced by search for correlation, whatever intricate and complex. Another words, can science advance without coherent models, unified theories, or mechanistic explanation and rely on automatically generated correlations? Can enormous datasets and advanced linkage techniques prevail over hypothesis in scientific inquiry? Can quantity mean quality to such extend that the aura of Big Data infallibility is able overpower epistemological objectivity? (Bollier, 2010; Boyd and Crawford, 2012).

The invoked above issues corresponding the Big Data hazards are perceived by EU citizens: 81% feel that they do not have complete control over their personal data online; 70% are anxious about their information being used for a different purpose from the one it was collected for; 69% would like to give their explicit approval before the collection and processing of their personal data; 55% are concerned about the recording of their activities via payment cards and via mobile phones; only 24% have trust in online businesses such as search engines, social networking sites and e-mail services; only 18% fully read privacy statements of the network services they use (EC, 2015). The subject matter is recognized by the European Parliament, the Council and the Commission and resulted in reaching agreement on December 15^{th,} 2015 on the new data protection rules, establishing a modern and harmonized data protection framework across the EU. Hence, the Regulation 2016/679 and the Directive 2016/680 are entered into force and will have to be applied and transposed into national level legislative regulations by May 2018. Therefore, the protection of natural persons with regard to the processing of personal data will be strengthened by a right to access and correction of personal data, an explicit right to be forgotten, a right to object to data processing, and the right to be informed when data security is breached. Additionally, the collection and processing of such data will be determined and controlled (EC, 2016).

CONCLUSION

The essential nature of converting raw data into information, since it is fundamental to modern existence, has been considered by many disciplines, including communications theory, library and information science, information systems, cognitive science, cybernetics, organization science, and managerial studies to name a few. This has generated multiple perspectives on the notion of data, information, and knowledge. Simultaneously, the ongoing transformations in the socio-economic environment arising from the development of ICTs are correlated with the exponential growth of generated, processed, stored, and shared data. Currently Big Data, counted in terabytes and petabytes, acquired from diverse sources in varied formats, classified, aggregated, profiled, analyzed, and interpreted in real time, is frequently seen as the most valuable asset giving a unique insight into remarkable patterns and correlations which can accordingly be repurposed by various entities in the wide range of sectors in order to create incalculable possibilities to accelerate innovation.

In the author's opinion Big Data is an essential and inevitable element of the IS and KBE in 21st century, and its processing can definitely support organizational and business progress of enterprises and institutions, both public and private. Notwithstanding, information founded on Big Data can inflict some ethical, social and ontological challenges regarding coexisting elements such as users' identity (both on- and offline), their privacy and digital abilities on one hand interrelated with data access, its ownership and confidentiality (particularly in view of the social media communication), and its reputation on the other. The actors of socio-economic stage should be well aware of potential threats behind the Big Data implementation since it may inflict privacy interference, manipulation, misinterpretation, and consequently culminate in social engineering or deepening social stratification by surveillance, let alone discrimination. It is therefore postulated to tackle the issue by fostering public education and implementing legal regulations which could execute some elementary principles, namely the transparent data collection, the introduction of data lifespan so that data would be eradicated after a given period of time, and the empowerment of users to have greater control over the access to and processing of their personal data in order to protect privacy, identity, and confidentiality. Simultaneously, organizations and enterprises should construct appropriate relationships with the stakeholders in order to explain the essence of Big Data and to clarify the notion of its collection, computation, processing, handling, implementing and exploitation. As K. Davis and D. Patterson (2012, p. 8) correctly acknowledge: "Big Data is, like all technology, ethically neutral. […] Individuals and corporations, however, do have value systems, and it is only by asking and seeking answers to ethical questions that we can ensure Big Data is used in a way that aligns with those values".

REFERENCES

1. Ackoff R.L. (1989). From Data to Wisdom // *Journal of Applies Systems Analysis*. Vol. 16, pp. 3-9.
2. Bellinger G.; Castro D.; Mills A. (2004). *Data, Information, Knowledge, and Wisdom*. Retrieved from: http://www.systems-thinking.org/dikw/dikw.htm (6.07.2016).
3. Bollier D. (2010). *The promise and Peril of Big Data*. Washington: The Aspen Institute.
4. Boyd D.; Crawford K. (2012). Critical Questions for Big Data: Provocations for a Cultural, Technological, and Scholarly Phenomenon. *Information, Communication & Society.* Vol. 15, (5), pp. 662-679.
5. Castells M. (2013). *Społeczeństwo sieci.* Warszawa: Wydawnictwo Naukowe PWN.
6. Davenport T.H.; Prusak L. (1998). *Working Knowledge: How Organizations Manage What They Know.* Harvard: Harvard Business Shool Press.
7. Davis K.;VPatterson D. (2012). *Ethics of Big Data.* Beijing: O'Reilly.
8. Drucker P. (2001). The next society. *The Economist.* No. 52.
9. EC (2015). *Special Eurobarometer 431: Data Protection.*
10. EC. (2016). *The EU Data Protection Reform and Big Data: Factsheet.* Retrieved from: http://ec.europa.eu/ justice/data-protection/files/data-protection-big-data_factsheet_web_en.pdf (10.07.2016).
11. Eliot T.S. (1991). *Collected Poems 1909-1962.* Orlando: Harcourt.
12. Eppler M.J.; Mengis J. (2004). The Concept of Information Overload - A Review of Literature from Organization Science, Accounting, Marketing, MIS, and Related Disciplines // *The Information Society: An International Journal.* No. 20(5), pp. 1–20.
13. Floridi L. (2006). Information Ethics, its Nature and Scope // *SIGCAS Computers and Society*. No. 36 (3), pp. 21-36.

14.Hey J. (2004). *The Data, Information, Knowledge, Wisdom Chain: The Metaphorical Link*. Retrieved from: http://www.dataschemata.com/uploads/7/4/8/7/7487334/dikwchain.pdf (7.07.2016).
15.Hill K. (2012). How Target Figured Out a Teen Girl Was Pregnant Before Her Father Did. *Forbes* 2012-02-16. Retrieved from: http://www.forbes.com/sites/kashmirhill/2012/02/16/how-target-figured-out-a-teen-girl-was-preg nant-before-her-father-did/ (4.07.2016).
16.Hitzler P.; Janowicz K. (2013). Linked Data, Big Data, and the 4th Paradigm // *Semantic Web.* No. 4(3), pp. 233-235.
17.ICIE (2015). *International Centre for Information Ethics*. Retrieved from: http://icie.zkm.de/research (4.07.2016).
18.IDC (2014). *The Digital Universe of Opportunities: Rich Data and the Increasing Value of the Internet of Things*. Retrieved from: https://www.emc.com/collateral/analyst-reports/idc-digital-universe-2014.pdf (3.07.2016).
19.ITU (2012). *Recommendation ITU-T Y.2060: Overview of the Internet of Things*. Retrieved from: http://www.itu.int/ITU-T/recommendations/rec.aspx?rec=y.2060 (3.07.2016).
20.Jewell D.; et al. (2014). *Performance and capacity implications for big data*. IBM Redbooks. Retrieved from: http://www.redbooks.ibm.com/redpapers/pdfs/redp5070.pdf (6.07.2016).
21.Kisielnicki J.; Sroka H. (2005). *Systemy informacyjne biznesu: informatyka dla zarządzania.* Warszawa: Placet.
22.Labrinidis A.; Hosagrahar V.J. (2012). Challenges and opportunities with big data // *Proceedings of the VLDB Endowment*. Vol. 5(12), pp. 2032-2033.
23.Manovich L. (2012). Trending: the promises and the challenges of big social data. In: M.K. Gold (ed.), *Debates in the Digital Humanities* (pp. 460-475). Minneapolis: University of Minnesota Press.
24.Marr B. (2015). *The 5 Scariest Ways Big Data is Used Today*. Retrieved from: http://data-informed.com/the-5-sca riest-ways-big-data-is-used-today/ (10.07.2016).
25.Materska K. (2007). *Informacja w organizacjach społeczeństwa wiedzy.* Warszawa: Wydawnictwo SBP.
26.McKinsey Global Institute. (2011). *Big Data: the Next Frontier for Innovation, Competition, and Productivity*. Retrieved from: http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innova tion (3.07.2016).
27.Nadolny M. (2012). *Data Economy – gospodarka oparta na danych*. Warszawa: MGG Conferences.
28.OECD (2000). *Knowledge Management in the Learning Society*. Paris.
29.Robertson T.D. (2013). The Data/Information/Knowledge/Wisdom Hierarchy Goes to Seminary // *Advances in the Study of Information and Religion.* Vol. 3, article 7.
30.Rowley J. (2007). The Wisdom Hierarchy: Representations of the DIKW Hierarchy // *Journal of Information Science.* No. 33, pp. 163-180.
31.Ruff J. (2002). Information Overload: Causes, Symptoms and Solutions // *Harvard Graduate School of Education*, pp. 1-13.
32.Soubra D. (2012). *The 3V that define Big Data.*. Retrieved from: http://www.datasciencecentral.com/forum/ topics /the-3vs-that-define-big-data (5.07.2016).
33.Stefanowicz B. (2013). *Informacja, wiedza, mądrość*. Warszawa: Główny Urząd Statystyczny.
34.Toffler A. (1997). *Trzecia fala.* Warszawa: Państwowy Instytut Wydawniczy.
35.Zeleny M. (1987). Management support systems: towards integrated knowledge management. *Human Systems Management.* No. 7, pp. 59-70.
36.Zikopoulos P.C., et al. (2012). *Understanding Big Data: Analytics for Enterprise Class Hadoop and Streaming Data*. New York: McGraw Hill.

Summary

Technological transformations of the 20th century have brought extensive socio-economical changes. The present-day society and economy are founded on acquisition, retrieval, distribution, and manipulation of data and information and consequently on creation of both explicit and tacit knowledge. The reflections of this paper are focused on Big Data as the major instrument of today's data processing in order to obtain information and knowledge. The objective of the article is to discuss the Big Data phenomenon and present its promise and peril from the social, economic, ethical, and ontological perspective. The article is based on literature studies.

Epistemologically, *knowledge* as a key resource of today's enterprises and organizations is created by processing, interpretation and interconnection of *information*; which in turn is based on *data*, raw facts, signs and symbols that by categorization, condensation, contextualization and correction become purposeful and relevant. The

pinnacle of the hierarchy is *wisdom* recognized as accumulated knowledge allowing to act critically in new situations and interconnected with ethical judgement.

The 21st century exponential growth of data together with the increase in the computing power, the research on artificial intelligence, and data mining has led to the creation of tools to analyze large, diverse, variable, and often unstructured data sets, referred to as Big Data, in order to transform them into purposeful information and consequently to interconnect and interpret into knowledge. The most concisely the term of Big Data is defined as datasets that cannot be processed or analyzed using traditional processes or tools. Big Data is characterized by the following parameters: volume (currently counted in petabytes or exabytes), variety (various data sources such as: spreadsheets, databases, emails, photos, videos, PDFs, audio, websites, tweets, blogs, comments, tags, SMS messages, WiFi location tracking), velocity (real-time data processing), veracity (data relevance in order to avoid bias and noise). The Big Data pipeline architecture consists of five stages such as: acquisition, extraction, aggregation, analysis, and interpretation.

The reasoning founded on Big Data can offer extensive insight into many complex issues and can improve the quality of administration on both government and local level, as well as scientific research, let alone business decision-making. Notwithstanding the enormous potential benefits of using Big Data, its menace needs to be recognized. Although the data for analysis is collected anonymously and from legitimate sources with law-restricted access, one can doubt whether Internet users consciously accept the rules and privacy policy of the services they use. Furthermore, the ongoing controversy of so-called public data in the social media sites is the battleground between privacy campaigners and transparency supporters. Additionally, acquired data may not be adequately protected and therefore can leak, be traded or stolen. Moreover since anonymity in the Internet is apparent, even aggregated data can be sensitive and may unleash surveillance, or even discrimination, against citizens by both the state and corporations. On top of that, the process of deciding which data attributes and variables are essential and which could be ignored together with the algorithms applied in computerized data sifting tend to be error-prone so that patterns are identified where none are existing.

The social approach to the Big Data threats is correlated with reinforcement of the digital divide due to the fact that the advantaged are those with extensive and sophisticated computational skills. From the ontological perspective the peril of Big Data is induced by the question whether a scientific construction of theories and models, particularly in social sciences, might be replaced by search for correlation, whatever intricate and complex. Can science advance without coherent models, unified theories, or mechanistic explanation and rely on automatically generated correlations? Can quantity mean quality to such extend that the aura of Big Data infallibility is able overpower epistemological objectivity?

Concluding, in the author's opinion Big Data is an essential and inevitable element of the realms of the information society and the knowledge-based economy in the 21st century. Big Data processing can definitely support organizational and business progress of enterprises and institutions, both public and private. Notwithstanding, information and knowledge founded on Big Data can inflict some ethical, social and ontological challenges regarding coexisting elements such as users' identity (both on- and offline), their privacy and digital abilities. The Big Data analysis may inflict privacy interference, manipulation, misinterpretation, and consequently culminate in social engineering or deepening social stratification by surveillance, let alone discrimination. It is therefore postulated to foster public education and implement legal regulations which could execute some elementary principles, namely the transparent data collection, the introduction of data lifespan so that data would be eradicated after a given period of time, and the empowerment of users to have greater control over the access to and processing of their personal data in order to protect privacy, identity, and confidentiality.

