International Forum
Vol. 21, No. 1
June 2018
pp. 192-209

FEATURE

Big Data, Education, and Social Responsibility

Yaroslav Ovdiyenko

Abstract. *The Internet has become one of the basic tools in everyday life of society. It is intensively used for education, as well as for entertainment, communication, and many other social activities. However, considering the evidence of digital acceleration, schools have become more disconnected from the World Wide Web in different ways. This paper identifies education opportunities and challenges in the contemporary informational realm addressing certain issues related to big data. On the one hand, it can revolutionize academia by giving it a new life. On the other hand, it can be dangerous, deceiving, and even killing. Thus, touching the contemporary reality linked to the daily life of students, parents, teachers, administrators, and other segments of society, this work aims at staying on guard of social responsibility pointing toward the new research horizons concerning the ways of innovative improvement of teaching products and services, informed learning strategies, and big-data-driven decisions related to both theory and practice in the settings of educational institutions.*

Keywords: big data, education, e-learning, data science, data mining, social responsibility

Introduction

In the 19th century, loping off a leg without anesthesia was a common practice, antiseptics and germ theory of disease had not been discovered yet, and the doctors "did not feel any necessity to change aprons or clean knives or even wash their hands between surgeries,". It was when the best method of treatment was "bleeding off some of the excess blood," and the "average life expectancy at birth was 32 in 1800, 41 by 1850, 50 by 1900, and 67 by 1950" (Knight, 1998, pp. 30-34). At the same time, the modern scientists admit that "the characterization of complex diseases at the molecular level combined with medical and treatment history, diagnostic or imaging tests offer unprecedented opportunities for personalized medicine" (Rossell, 2015, p. 1). Definitely, the more advanced microbiological knowledge is, the more assurance of success in medical treatment.

Relating to this story, the contemporary education can be as harmful as medicine in the 19th century. Without special microbiological knowledge, the ways of treating patients were based on the shallow, superficial "scientific" knowledge. Today, as it is put by Wang (2017), "In the big data era, the wealth of data available to social scientists has been considered as the microscope to microbiologists" (p. 6). Therefore, the offered similitude, which compares both the medical field with its progress and education with its big data perspective, helps to see the parallels in the decision-making approaches based on the available scientific opportunities. The treatment of the nowadays schools might be as poor in some way as it was in the 19th-century hospital where people were merely brought to die. Thus, to reflect on the harmful and dangerous past will be only possible in the future. Yet, for teaching and learning science, it is time to catch up with the global trend and align with the health care industry progress. Indeed, doing hi-tech brain surgeries might save a human life, but education makes difference in its quality.

Being carried by the wind of the progressive and dynamic scientific reality mentioned above, this paper attempts to spot possible opportunities and challenges related to big data and education. Through this, it aims at awakening scientific interest in the minds of researchers in the learning field. It concerns especially those who are in charge of the educational services sector. Thus, the focus of this theoretical study is to get a better perception of the available strategic choices in the teaching and learning science that may affect teachers, students, curriculum materials, instructional strategies, and society. The whole article is divided into three top-level units such as *Big Data in Research*, *Big Data in Education*, and *The Issues of Big Data*.

Big Data: Main Ideas

The purpose of this section is to introduce the initial ideas of big data. It starts with the contemporary trends, definitions, and various big data sources. Then, it raises a question about the importance of the discussed concepts through the market trend, decision-making pyramid, and scientific paradigm. Finally, it presents a technical side of big data discovering the expected expertise, knowledge discovery process, data-mining circle, and methodological kit.

What Is Big Data?

Big data trend. August 4, 2010, was the day when Eric Schmidt, the CEO of Google announced that every two days a human race generates the same amount of data as it did from the beginning of human era up until the year 2003 (Siegler, 2010). It is a tremendous, large, massive information that opens immense opportunities for retrieving essential knowledge. It is a well-known fact that in the 21st century the Internet websites produce millions of records, comments, recommendations, and photos. E-learning platforms such as MOOC (massive open online courses) and mobile applications live a trail of the digital ore (Blackmon & Major, 2016). As it is stated by Wang (2017), when "we digitalize education, our online activities create a vast digital trove of data" (p. 381). Therefore, it is evident that big data has saturated and penetrated every aspect of human life.

Big data definition. There are different opinions related to the definitions of big data. Historically, the term "originated in the lunch-table conversations at Silicon Graphics in the mid-1990s, in which John Mashey figured prominently" (Lohr, 2013, p. 5). However, what is more important is to understand the very up-to-date meaning of the concept. According to some authors, there are no standard definitions (Diebold, 2012; Porche, Wilson, Johnson, Tierney, & Saltzman, 2014). Yet, it is important to look at some explanations of the concept.

SAS Institute provides the first, direct explanation of big data. They claim that "Big data is a term that describes the large volume of data–both structured and unstructured" ("Big data," n.d.). In addition, Manyika et al. (2011) attempt to explain the concept, comparing it to the "large data sets" (para. 2). Later on, Oussous (2017) adds another flavor, saying that "Big Data refers to large growing data sets that include heterogeneous formats: structured, unstructured, and *semi-structured data* [italics added]" (p.3). He continues that "Big Data has a complex nature that requires powerful technologies and advanced algorithms" (p. 3). Notwithstanding the diversity of opinions, there is another figurative description of the term that was provided before by Laney (2001) who defined big data by establishing the three conceptual pillars such as volume, variety, and velocity.

Indeed, thousands and millions of records that are coming from different sources count the massive volume of big data. The large *variety* of big data is nothing else as the existing media that is classified as structured (represented by numbers and categories), unstructured (does not have a predefined data model), and semi-structured data. In fact, the unstructured data constitutes 95% of the existing data in the world (Gandomi & Haider, 2015). Regarding *velocity*, it explains the idea that the content is created rapidly, constantly, and everywhere. As it was noted, "more data has been created in the past two years than in the entire previous history of the human race" (Marr, 2015). This fact leads to another discussion that is more focused on the sources of big data.

Big data sources. As of 2017, based on the research provided by the *Smart Insights* marketing agency, every 60 seconds the Internet users produce 500 YouTube videos, send 149513 emails, generate 3.3 million Facebook posts, conduct 3.8 million Google searches, upload 65972 Instagram photos, create 448800 twits, publish 1440 WordPress posts, and deliver 29 million WhatsApp messages (Allen, 2017). In such a way, videos, music, photos, scanned and typed text, posts, comments, blogs, interviews, and surveys are the materials of digital data coming from the different devices such as mobiles, laptops, and desktops, for example. This is a reason why there is an urgent necessity to look at the big data trend and relate it to education.

Thinking about big data in the context of this paper, it may be produced by schools, communities, and organizations. In fact, it may be generated by students, teachers, administrators, parents, and friends in the settings of educational institutions. All of them are characterized as the creators of the "digital footprints" in the daily life. For example, almost every accountability system in education has its own database. Twitter and Facebook are full of comments and posts from both happy and angry parents in relation to their children and schools. In fact, they may express their opinions on teachers, education systems, or technologies, for example. Logs or user-generated content may represent another source of data after working with mobile applications, for instance. To this account, one can add OCR systems with its capacity to produce immense amounts of digitalized text and images (Daniel, 2014). So, even in education, there are a lot of springs of big data that can flow into further research and analysis.

Categorizing the sources of big data, one may conclude that there are four essential groups: archives (scanned docs, emails, medical records), media (images, audio, videos), social networking (twits, likes, chats), sensor data (medical devices, wearable devices), and log data (application, server, click stream) (Poddar, 2016). All of them contain certain components such as scanned documents, email, and certain records for archives; images, audio, and videos for media; twits, likes, and chats for social networking; electronic devices, wearable devices for sensor data; and application, server, and clickstream for log data. In fact, there might be other aggregation and categorization of the same nature.

Why Is It Important?

Proven results. The value of big data has been associated with a hype created by the digital analytics. Yet, passing this period researchers from Dell suggest to "focus on the proven results that big data can deliver" ("Big data and analytics," n.d., para. 3) such as product quality improvement, understanding the needs of a client, decision-making quality and speed, recognizing new opportunities. It would contribute to improving human existence in every sphere of life and education, in particular, which is in constant need of improvement toward innovative products and services, transformed decision-making theories, and informed strategies (Daniel, 2014). Thus, paying attention to the advantages of functioning in the big data world would "translate its potential into actual social and economic value" (Günther, 2017, p. 191).

Market perspective. Looking at the trend from the business perspective based on the reports done by Statista, one can notice that by 2018 the big data market is projected to grow up to $40.8 billion and by 2026, the size is predicted to be $92.2

billion., So it will probably double in less than 8 years ("Global big data market size 2011-2026," 2017). It is another tip to be socially responsible for those who are in charge of the administrative and curriculum and instruction development as well as for the scholars in the field of education. A good example of such social responsibility is explained by Dell ("Big data and analytics," n.d.): "Every minute of every day, business leaders and IT face the challenge of translating data into insights for smart *decision-making* [italics added]" (para. 3).

Wisdom perspective. The important point mentioned in the business world is striving toward wisdom. Undeniable is the fact that the Bible teaches the very same thing from its inspired perspective. "Blessed is the one who finds wisdom, and the one who gets understanding, for the gain from her is better than gain from silver and her profit better than gold," says King Solomon (Prov 3:3-18). From the *information processing theory* perspective, based on DIKW (data-to-information-to-knowledge-to-wisdom) hierarchy, in order to reach the level of scientific wisdom one should start with exploring data while moving toward extracting information, then knowledge, and, finally, wisdom (Glasbeek, 2014). That is why the ideal goal of any scientific research is to find wisdom around, both in the natural and social world. Considering the big data trend, the DIKW paradigm becomes even more relevant as mining the raw digital oil will eventually result in practical data-driven insights (Provost & Fawcett, 2013; Institute, 2013). Probably, it is time to consider big data as a historical moment which is predicted in the Scriptures by the angel of God as he instructed Daniel to "seal up the book until the *end of time* [italics added]" when "knowledge will increase" (Dan 12:4). Thus, along with the rapid increase of data in the information age, the increase of knowledge and scientific wisdom is a possibility.

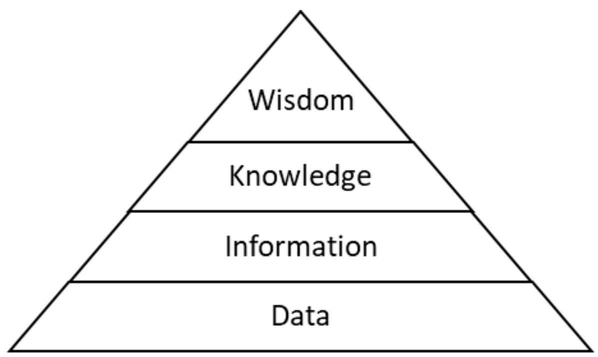

Figure 1. DIKW pyramid.

Scientific paradigms. The information has definitely grown in the big data civilization and the scientific paradigm is changing, at least according to Kitchin (2014), who discusses the development of the scientific patterns pertaining to different periods. Thus, he starts from the experimental science that began during the pre-Renaissance period in a form of empiricism. Back then, it was intended to describe the natural phenomena. Then, in the pre-computer era, the focus was on the theoretical science employing such processes as modeling and generalization. The third, pre-big data paradigm presented a computational science approach attempting to simulate complex digital models by means of computational resources. The final period in this hierarchy is characterized as data-intensive being famous for advanced statistical exploration and data mining. The author names it as exploratory science that is actually symbiotic in nature.

Research approaches. The first symbiotic nature of big data research is based on its attribute to combine both inductive and deductive approaches fusing them together based on a meta-analysis (Günther, Mehrizi, Huysman, & Feldberg, 2017). The scholars explain that the inductive approach is about emerging categories, which finally produce a theoretical explanation of the researched patterns. This approach may remind the qualitative research design where an academic does not limit a study with some predetermined variables. In the process, all the qualitative "variables" or themes emerge naturally without any control. Another way in which data science works is determined by the deductive approach. It starts with the existing theory which defines the way research will be conducted. In this case, decision makers test the prepared in advanced hypotheses analyzing big data. The main aim here is to justify the already settled options. The second symbiotic nature manifests itself in the ability of big data techniques to work with different types of data that can make a difference and be valuable in the research field.

Technology of Big Data

Data science. Following the zeitgeist of the 21st century, the contemporary researchers invented so-called data science to deal with big data. According to Dhar (2013), this term "is becoming increasingly common" (p. 2). Saltz and Stanton (2017) define it as "an emerging area of work concerned with the collection, preparation, analysis, visualization, management, and preservation of large collections of information" (p. ii). In fact, data science is a compound discipline that is a blend of computer science, math and statistics, and business knowledge ("Data Science Institute," 2017). Moreover, the paired intersections of these three areas create another three segments such as machine learning (computer science plus math and statistics), software development (computer science plus business knowledge), and traditional quantitative research (math and statistics plus business knowledge). In such a way, data science employs the three above-mentioned areas in order to be able to comply the big data requirements.

KDD process. The Knowledge Discovery in Database (KDD) process (Pazmiño-Maji, García-Peñalvo, & Conde-González, 2017) presents another model that describes how data science works. The initial purpose of it is to extract knowledge out of data when the databases are large. Algorithmically, it moves from selection to processing, transformation, data mining, and evaluation. On its way from data to knowledge, it leaves such byproducts as target data, processed data, transformed data, and patterns. Though the process is technically complicated, it moves from the raw material to the useful knowledge and insights.

Data mining process. One of the steps in the mentioned above process is data mining. The technology of data mining, described by the *Cross Industry Standard*

Process for Data Mining (CRISP-DM), has been fundamentally and widely used until the present day (Abbott, 2014). It is a data mining model that splits the whole process of finding informational patterns and knowledge into six stages, such as business understanding, data understanding, data preparation, modeling, evaluation, and, finally, deployment. Based on its iterative idea, it moves in cycles until the final point of reaching the desired level of satisfaction.

Gandomi and Haider (2015) offered another similar idea. As they explained, the procedure that outlines the mining of insights out of big data consists of the following steps: acquisition and recording; extraction, cleaning, and annotation; integration, aggregation, and representation; modeling and analysis; and, finally, interpretation. The first two items relate to the data management process, while the last three are in charge of data analysis.

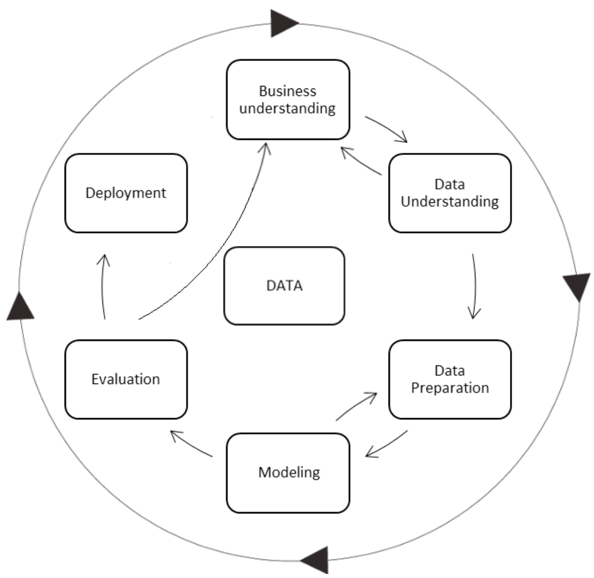

Figure 2. Cross Industry Standard Process for Data Mining (CRISP-DM).

One more data mining process model is offered by Daniel (2015) who suggests making it simple paying attention to the four stages only: data collection, data analysis, data visualization, and application. Of course, all of these three concepts are similar in a way; yet, they try to approach the technological process related to data mining from different perspectives and angles. At the same time, there is a room for improvement; however, the analysis and evaluation of the data science models goes beyond the main purpose of this journal article.

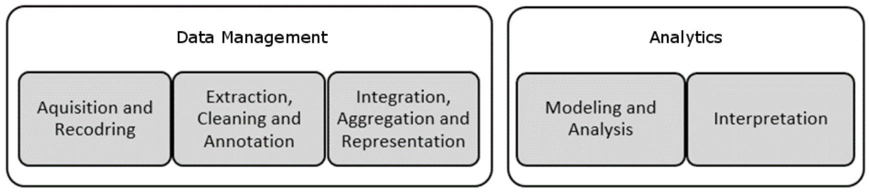

Figure 3. Big data process.

Data mining methods. Another area of the technological aspect of data mining is methodological. According to Liu & Huang (2017), there are seven items in the methodological list: classification, value estimation, clustering, frequent pattern mining, text mining, structural analysis, and behavior profiling. Gandomi and Haider (2015) give another set of techniques who view the list slightly different, including text analytics, information extraction, text summarization, question answering, sentiment analysis, audio analytics, video analytics, social media analytics and predictive analytics. Some scholars may be even more specific in certain areas. For example, Wang (2017) marks out three positions in the text analysis selecting only topic models, network analysis, and sentimental (or opinion) analysis. All of such methods may serve a researcher as the scientific algorithms that can help to elicit knowledge from data.

Big Data in Education

Peter Fix (2015), in his book *Gamechangers,* is raising up an important question about establishing a disruptive and innovative business. Education is not an exception in this regard. Some scholars believe that big data and data science can transform, change, and improve the processes of teaching and learning (Baer & Campbell, 2012). As it is nicely put by Wang (2017), "For all the richness of big data and emerging methodological tools at our disposal, if used properly, big data could be a goldmine of education policy research" (p. 10). This opinion points toward the opportunities of the current digital game-changing trend.

The first of such changes may be approached from the perspective of resolving a long-term question regarding how to measure the affective domain (Bloom, 1968). Indeed, education can go beyond measuring the 3R success. It can investigate socio-emotional learning as well. In this regard, Liu and Huang (2017) suggest that data analytics can be used to measure non-academic factors that are extremely important in the theory of student motivation.

The second change in big data game is that it can offer a new customizing predictive tool. In other words, it would measure weaknesses and opportunities and affect learning outcomes for any individual student. The personalized learning experience was a dream of the mastery learning theorists (Airasian, Bloom, & Carroll, 1971; Khan, 2013). What was not possible years ago, has become possible today not in the offline face-to-face settings, but in the e-learning environments. Thus, employing data science in education can lead to the analytical reports of different variety and quality. For example, Daniel (2015) talks about the possibility of using descriptive analytics, predictive analytics, prescriptive analytics, performance outcomes, and process outcomes in the settings of educational institutions going blended. Yet, the mentioned analytical tools can be as well used considering other online e-learning platforms such as MOOC, educational games services, and online quizzes, for example (Popescu et al., 2017). All of them are potential producers of educational big data and, in this connection, are recommended for further analysis. Such results eventually inform the adaptive learning systems which in turn customize and personalize online learning experience (Graf, 2011; Sclater, 2017). The age of hard copies and large size classes with a lack of personal approach is coming to its shrinkage and, probably, will be obsolete someday.

One more useful application of big data is education policy (Wang, 2016). Thus, analyses of a digital trace of the online public services could illuminate social opinions and bring awareness regarding the contemporary educational issues. Even the data that is coming from the advanced learning management systems can be a good source for it as it may reveal the current situation of the academic opportunities and weaknesses. For example, public opinions regarding standardized tests, effective learning experience, positive or negative feedback can be collected and analyzed using the data science technologies.

Improvement and innovation in the educational sector is the result of big-data-driven decisions. As Günther, Mehrizi, Huysman, and Feldberg (2017) describe, "organizations can leverage big data while generally continuing to function in the same manner, only more *effectively and efficiently* [italics added]" (p. 197). Indeed, schools, academies, and universities can gain a lot from operating in the field of data science. Yet, there are certain issues to be discussed as they primarily are regarded as the reasons for possible obstacles to success.

The Issues of Big Data

There are certain matters in big data that have to be addressed in this paper. These are the concerns of better education and data-driven decisions in the contemporary school and academia. Such issues will be discussed starting from some big data misconceptions to the matters of educational capability, portability and interconnectivity, and finally ethical responsibility. All of them challenge the nowadays society and bring awareness of better learning opportunities.

Misconceptions

There are three misconceptions that are discussed by Wang (2017). The first one is the idea that big data is good because it is big itself. In other words, the bigger the data is the better. This concept reduces the need for the methodological rigidity. The author argues that it is not bad to have a large volume of a raw digital material for further knowledge and insights. Yes, but the necessary methodological care should be applied as well. The way in which a sample represents population is important and should be considered always no matter how big the data is in this case. It is not possible yet to get the entire population for a study. For example, if Twitter has millions of open records in its databases, it does not mean that the whole world population is on the Twitter. Thus, the traditional sampling approach should be also considered.

The second misconception brought by Wang (2017) is about a theoretical framing that may become obsolete. Some believe that big data is enough to figure out all the theory behind merely by crunching the digital matter with profound algorithms. The author argues that even the patterns discovered using mathematics may not be sufficient to prove or come with a theory not using special theoretical knowledge behind the study. There might be certain chance-associations that will eventually lead to the data mining misconceptions. Thus, if scholars simplify human behavior reducing it to numbers, they may lose some other important facets.

The third most profound misconception explained by Wang (2017) is that the idea that a computer and algorithms can substitute humans in terms of big data research. In reality, artificial intelligence does a great job for humankind; yet, it is not sufficient to make human involvement and intervention obsolete. Humans do the job that the machines cannot do. First, the initial quality of big data should be assessed by human intelligence; otherwise, it might be error-prone and lead to the wrong results. Second, the analysis process and the model parameters should be set by scientists. Third, algorithms do not work if the context is slightly changed and the conditions are unknown; only humans can operate there, as only humans can capture such reality thanks to common sense. Thus, the presence of big data does not make things happen automatically: there is a need for the rigid methodology, previous theoretical foundations, and human ability to be flexible in a new context. At the same time, there is a need for algorithmic intelligence and smart machines.

That is why, balancing two facets of the scientific world, big data can benefit society if used correctly.

Educational Capability

Big data is a relatively new phenomenon which became evident and progressive during the age of computer science and digital technologies. Its analysis uses mathematical algorithms along with machine computation technology and statistical methods (Liu & Huang, 2017). All of such directions are covered by the unifying umbrella of data science. In this regard, the educational dimension of the new area is not developed enough. Relating to the teaching and learning views it is noted that "By taking a close look at education in big data and data science, we should try to fill the gap among data, technologies, and people" (Song & Zhu, 2016). That is why, some scholars are talking about the best possible curriculum for the future data scientists, who can integrate even more areas such as social science, education technology, computer science, data visualization, and statistics (Daniel, 2014; Dos Santos, 2014; Wallach, 2016). Thus, the capability of the education scholars to do data science reaching new horizons in big data analysis depends on the expertise and professional capacity of the researchers. Therefore, the recommendation for all education policymakers is to move toward interdisciplinary training programs preparing future educators in the data science field.

Portability and Interconnectivity

Two more big data issues are portability and interconnectivity (Günther et al., 2017). The first one refers to the possibility of providing a special compatible interface for different data. It may relate both to (a) data itself or the types of data distribution as well as to (b) the application programming interface between different types of software. Mainly, it resolves the problem of a remote data access from one digital software context to another. Regarding the interconnectivity, its idea is to combine different data sources for the benefits of the final result. Thus, synthesizing digital content from numerous big data sources and combining them in a new way can lead to the new insights. Such a fusion may go beyond the expectations from the analysis of pre-existing datasets. In this case, the balance between human and artificial intelligence can be seen as the balance between creativity and computation ability as the algorithms alone are not yet able to step outside the programmed predetermined reality.

Ethical Considerations

It is not surprising for an educational researcher to discuss ethical issues and related topics. That is why, there are some ethical concerns in regards to big data that find their place in this paper too (Martin, 2015; Pence, 2015). Basically, they can be divided into two subclasses depending on whether big data is open or controlled. For the open big data, the question that is raised by the ethical committees is whether the open data is really open and can be used for research.

For the controlled big data, the issue is about how the data can be used and targeted against subjects.

Starting from the open data, it is clear that the assumptions regarding the ethical considerations that have been developed almost 35 years ago and applied to social science until now, historically do not reflect the need of the contemporary society and in their roots are not the same for the field of mathematics, statistics, and computer science (Metcalf & Crawford, 2016). Indeed, “The ethical implications of data-driven knowledge are hard to observe at an aggregate level where data are impersonal and anonymous” (Schroeder & Cowls, 2014). The research conducted in the latter areas has no direct contact with human data as it is considered as a substrate for testing algorithms and systems and not the direct object of interest. Moreover, the division between public and private big data is not very well defined today (Vayena & Tasioulas, 2016; Vayena, Gasser, Wood, O’Brien, & Altman, 2016). For example, taking into consideration that the Twitter records are open for everyone on the Internet, does it mean they are free to use as a research raw material? After analyzing the digital ore, there might be some personal and sensitive findings related to certain individuals that are not obvious from the human standpoint but clear from the perspective of artificial intelligence. Here, the institutional review boards are overwhelmed with such an issue as they see some risks to human subjects. Thus, some scholars see a need of creating an ethical framework adaptable to the field of data science as the ideas that were described in the Common Rule (Coleman, Menikoff, Goldner, & Dubler, 2003) in 1981 do not cover the ethical issues which pop up during the era of big data, Internet, and artificial intelligence.

The second idea of big data ethical considerations stands on the point that even the controlled data may be dangerous for individuals. This kind of thoughts started with the development of learning analytics. Despite the great benefits of understanding of the learners, their capacities and capabilities by means of artificial intelligence, “collection of data and their use face a number of ethical challenges, including location and interpretation of data; informed consent, privacy, and deidentification of data; and classification and management of data” (Slade & Prinsloo, 2013). Thus, predicting the success of one student based on the statistics obtained from the learning accountability platform means predicting the failure of another student. The issue of such learning analytics results would probably be a personal issue of an individual student after his or her “academic diagnosis” has been released before the decision-making committees. In other words, in cases where a human nature would allow a person to become a better student, artificial intelligence would advise administrators to make an “algorithmically grounded” decision to ban such an individual. In this way, the whole cohort of learners may be separated, discriminated, and predestined toward a programmed and predetermined by the artificial intelligence future. Therefore, it is important to have a balanced view analyzing big data that the decisions made would be more human-minded.

It is clear that the issue of ethical considerations does exist. Moreover, it is observed from different angles and made to appeal for the truth. Yet, for some, in the contrary, the data protection mechanism and policies do not work anymore as the digital age changes the foundations of reality (Ambrose, 2015). Maybe, it is possible to prevent a thesis or dissertation from conducting the big data research in an unethical way; yet, it is not possible to stop a common Internet user or even some organization from doing data science for their own benefits. Probably, some educational institutions will not follow all the necessary recommendations of the revised ethical framework or make their inhuman decisions based on the available student data.

In this regards, talking about social responsibility, the new scientific paradigm inhering in the 21st century should be accepted that it will move on other issues pertaining to this reality. In connection with the teaching and learning science, it should definitely inhale a breath of digital atmosphere where big data is dissolved. As it is well said by one scholar, "Looking to the past for a richer understanding of the issues, I do not find that this is all a palaver over nothing – that big data is nothing new or that it is a lot of hype about nothing" (Ambrose, 2015, p. 277). Yes, reflecting on the aforementioned ideas, the author is right – it is something. She continues: "Instead, I find that there is potential for a significant epistemic shift on the horizon – one that presents the law with a great deal of uncertainty and responsibility" (p. 277). Probably, she is talking about the social responsibility entrusted to the gatekeepers of the educational sector. Therefore, the idea of the message is to move on, extending the timeline as from the market, decision making, biblical, and scientific paradigm perspective, big data is at the door.

Conclusion

After having been presenting the concept of big data from different perspectives and angles, it is time to ask a logical question: "What should be done with this?" The comparison of education with the medical science can be applied in further visualization while thinking about big data trends, definitions, sources, results, market and wisdom perspectives, scientific paradigms, research approaches, and technological opportunities. The idea of employing big data in education would probably draw other insightful ideas that can be elaborated and applied in the real schools and institutions. Finally, the issues and ethical considerations would make people think and resolve the existing problems mentioned above, not only theoretical but those which are directly related to the contemporary society, its needs, and education system in particular.

Concluding this theoretical paper, whose aim was at extending a sense of social awareness related to both big data and education, it is appropriate to tame heed of some recommendations from the researchers regarding the topic. First, "it is imperative for organizations to continuously realign work practices, organizational models, and stakeholder interests to realize value from big data" (Günther et al., 2017, p. 205). Second, big data research purpose is "to inform education policy-making, implementation, and evaluation" (Wang, 2017, p. 5). Finally, "there are many points to be considered, discussed, improved, developed, analyzed" with regards to big data (Sagiroglu & Sinanc, 2013, p. 47). Therefore, the recommendation of this article is to diligently conduct further big data research, especially, in the field of education.

References

Abbott, D. (2014) *Applied Predictive Analytics: Principles and Techniques for the Professional Data Analyst*. Hoboken, NJ: Willey.

Airasian, P. W., Bloom, B. S., & Carroll, J. B. (1971). *Mastery learning: Theory and practice*. (J. H. Block, Ed.). New York: Holt Rinehart and Winston.

Allen, R. (2017, February 6). What happens online in 60 seconds? Retrieved November 15, 2017, from https://www.smartinsights.com/internet-marketing-statistics/happens-online-60-seconds/

Ambrose, M. L. (2015). Lessons from the avalanche of numbers: Big data in historical perspective. *A Journal of Law and Policy for the Information Society*, *11*, 201.

Baer, L., & Campbell, J. (2012). From metrics to analytics, reporting to action: Analytics' role in changing the learning environment. In D. G. Oblinger (Ed.), *Game changers: Education and information technologies*. Washington, D.C.: EDUCAUSE.

Big data. (n.d.). Retrieved November 9, 2017, from https://www.sas.com/en_ph/insights/big-data/what-is-big-data.html

Big data and analytics. (n.d.). Retrieved November 10, 2017, from http://www.dell.com/learn/no/en/nochn1/shared-content~solutions~en/documents~getting-past-the-bigdata-hype-hero2.pdf

Blackmon, S. J., & Major, C. H. (2016). *MOOCs and higher education: Implications for institutional research*. Hoboken, N.J: John Wiley and Sons.

Bloom, B. S. (1968). *Taxonomy of educational objectives: The classification of educational goals*. London, England: Longmans.

Coleman, C. H., Menikoff, J. A., Goldner, J. A., & Dubler, N. N. (2003). *The ethics and regulation of research with human subjects*. Newark, NJ: LexisNexis.

Daniel, B. (2015). Big data and analytics in higher education: opportunities and challenges. *British Journal of Educational Technology*, *46*(5), 904–920. https://doi.org/10.1111/bjet.12230

Data science institute. (2017). Retrieved November 15, 2017, from http://www.fox.temple.edu/institutes-and-centers/data-science/

Dhar, V. (2013). Data science and prediction. *Commun. ACM*, *56*(12), 64–73. https://doi.org/10.1145/2500499

Diebold, F. X. (2012). A personal perspective on the origin(s) and development of 'big data: The phenomenon, the term, and the discipline, second version. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2202843

Dos Santos, R. (2014). Big data as a mediator in science teaching: A proposal. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2441534

Fisk, P. (2015). *Gamechangers: Creating innovative strategies for business and brands*. West Sussex: England: Wiley.

Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management*, *35*(2), 137–144. https://doi.org/10.1016/j.ijinfomgt.2014.10.007

Glasbeek, L. (2014). *What is knowledge?: A practical review of six philosophical perspectives*. Amsterdam, The Netherlands: Workplace Prosperity.

Global big data market size 2011-2026. (2017). Retrieved November 10, 2017, from https://www.statista.com/statistics/254266/global-big-data-market-forecast/

Graf, S. (2011). *Intelligent and adaptive learning systems: Technology enhanced support for learners and teachers*. (F. Lin, Kinshuk, & R. McGreal, Eds.). Hershey, PA: IGI Global.

Günther, W. A., Mehrizi, M. H. R., Huysman, M., & Feldberg, F. (2017). Debating big data: A literature review on realizing value from big data. *The Journal of Strategic Information Systems*, *26*(3), 191–209. https://doi.org/10.1016/j.jsis.2017.07.003

Institute, P. M. (2013). *A guide to the project management body of knowledge*. Newtown Square, PA: Project Management Institute.

Khan, S. (2013). *The one world schoolhouse: Education reimagined* (Reprint edition). New York: Twelve.

Kitchin, R. (2014). Big data, new epistemologies and paradigm shifts. *Big Data & Society*, *1*(1), 1–12. https://doi.org/10.1177/2053951714528481

Knight, G. A. F. (1998). *Ellen White's World: A fascinating look at the times in which she lived*. Hagerstown, Md: Review and Herald Publishing.

Laney, D. (2001). 3-d data management: controlling data volume, velocity, and variety. *META Group Research Note*, *6*.

Liu, M.-C., & Huang, Y.-M. (2017). The use of data science for education: The case of social-emotional learning. *Smart Learning Environments*, *4*, 1. https://doi.org/10.1186/s40561-016-0040-4

Lohr, S. (2013, February 1). The origins of "big data": An etymological detective story. Retrieved November 16, 2017, from https://bits.blogs.nytimes.com/2013/02/01/the-origins-of-big-data-an-etymological-detective-story/

Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Byers, A. H. (2011, May). Big data: The next frontier for innovation, competition, and productivity.

Marr, B. (2015, September 30). Big data: 20 mind-boggling facts everyone must read. Retrieved November 9, 2017, from https://www.forbes.com/sites/bernardmarr/2015/09/30/big-data-20-mind-boggling-facts-everyone-must-read/

Martin, K. E. (2015). *Ethical issues in the big data industry* (SSRN Scholarly Paper No. ID 2598956). Rochester, NY: Social Science Research Network. Retrieved from https://papers.ssrn.com/abstract=2598956

Metcalf, J., & Crawford, K. (2016). Where are human subjects in big data research? The emerging ethics divide. *Big Data and Society*, *3*(1), 2053951716650211. https://doi.org/10.1177/2053951716650211

Oussous, A., Benjelloun, F.-Z., Ait Lahcen, A., & Belfkih, S. (2017). Big data technologies: A survey. *Journal of King Saud University - Computer and Information Sciences*. https://doi.org/10.1016/j.jksuci.2017.06.001

Pazmiño-Maji, R. A., García-Peñalvo, F., & Conde-González, M. (2017). Statistical implicative analysis approximation to KDD and data mining: A systematic and mapping review in knowledge discovery database framework. In A. Schmidt, F. Laux, D. Hristovski, & S. Ohnishi (Eds.), *Proceedings of the Ninth International Conference on Advances in Databases, Knowledge, and Data Applications* (pp. 70–77). Wilmington, DE: IARIA XPS. Retrieved from https://www.researchgate.net/publication/317084890_Statistical_Implicative_Analysis_Approximation_to_KDD_and_Data_Mining_A_Systematic_and_Mapping_Review_in_Knowledge_Discovery_Database_Framework

Pence, H. E. (2015). Will big data mean the end of privacy? *Journal of Educational Technology Systems*, *44*(2), 253–267. https://doi.org/10.1177/0047239515617146

Poddar, D. (2016, May 30). A beginner's guide to big data! Retrieved November 15, 2017, from https://upxacademy.com/beginners-guide-to-big-data/

Popescu, E., Kinshuk, Khribi, M. K., Huang, R., Jemni, M., Chen, N.-S., & Sampson, D. G. (Eds.). (2017). *Innovations in smart learning*. New York, NY: Springer.

Porche, I. R. I., Wilson, B., Johnson, E.-E., Tierney, S., & Saltzman, E. (2014). *Data flood: Helping the navy address the rising tide of sensor information*. Santa Monica, CA: Rand Corporation.

Provost, F., & Fawcett, T. (2013). Data science and its relationship to big data and data-driven decision making. *Big Data*, *1*(1), 51–59. https://doi.org/10.1089/big.2013.1508

Rossell, D. (2015). Big data and statistics: A statistician's perspective. *Metode Science Studies Journal*, *5*, 143–149. https://doi.org/10.7203/metode.83.3590

Sagiroglu, S., & Sinanc, D. (2013). Big data: A review (pp. 42–47). https://doi.org/10.1109/CTS.2013.6567202

Saltz, J. S., & Stanton, J. M. (2017). *An introduction to data science*. Los Angeles, CA: SAGE Publications.

Schroeder, R., & Cowls, J. (2014). Big data, ethics, and the social implications of knowledge production. Retrieved from https://www.oii.ox.ac.uk/research/projects/big-data-to-advance-social-science-knowledge/

Sclater, N. (2017). *Learning analytics explained*. New York: Routledge.

Siegler, M. G. (2010, August 4). Eric Schmidt: Every 2 days we create as much information as we did up to 2003. Retrieved November 9, 2017, from http://social.techcrunch.com/2010/08/04/schmidt-data/

Slade, S., & Prinsloo, P. (2013). Learning analytics ethical issues and dilemmas. *American Behavioral Scientist*, *57*, 1510–1529. https://doi.org/10.1177/0002764213479366

Song, I.-Y., & Zhu, Y. (2016). Big data and data science: What should we teach? *Expert Systems*, *33*(4). https://doi.org/10.1111/exsy.12130

Vayena, E., Gasser, U., Wood, A., O'Brien, D. R., & Altman, M. (2016). Elements of a new ethical framework for big data research. *Washington and Lee Law Review*, *72*(3).

Vayena, E., & Tasioulas, J. (2016). The dynamics of big data and human rights: The case of scientific research. *Phil. Trans. R. Soc. A*, *374*(2083), 20160129. https://doi.org/10.1098/rsta.2016.0129

Wallach, H. (2016). Computational social science: Discovery and prediction. In R. M. Alvarez (Ed.), *Data science for politics, policy, and government* (pp. 1–11). Singapore: Cambridge University. Retrieved from http://dirichlet.net/pdf/wallach15computational.pdf

Wang, Y. (2016). Big opportunities and big concerns of big data in education. *TechTrends*, *60*(4), 381–384. https://doi.org/10.1007/s11528-016-0072-1

Wang, Y. (2017). Education policy research in the big data era: Methodological frontiers, misconceptions, and challenges. *Education Policy Analysis Archives*, *25*(0), 94. https://doi.org/10.14507/epaa.25.3037

Yaroslav Ovdiyenko, MS in Applied Mathematics
Ph.D. Student
Adventist International Institute of Advanced Studies
Silang, Philippines