The Use of Big Data in Education

Athanasios S. Drigas[1] and Panagiotis Leliopoulos[2]

**[1] Institute of Informatics & Telecommunications, Telecoms Lab - Net Media Lab, N.C.S.R. "Demokritos"
Agia Paraskevi, Athens, 15310, Greece**

**[2] Department of Digital Systems, University of Piraeus
Piraeus, Attiki, 18534, Greece**

Abstract

This paper is a study on the use of Big Data in Education. Analyzed how the Big Data and Open Data technology can actually involve to education. Furthermore how big mounts of unused data can benefit and improve education. Providing some new tools and methods bypassing the traditional difficulties and open a new way of education.

Keywords: *Big Data, Open Data, Education, Data Analytics, Cloud Computing, Benefits and affects in Education, Big Data analysis, Data Mining.*

1. Introduction

In this study we will see the use of Big Data and Open Data in Education. Also how big amount of data can be used and extracted to something useful, helping the industry to raise their profits. Hence a definition of Big Data and Open Data make clear how that these two technologies are classified.

After that it is very important to mention the goals and purposes of Big Data in education, giving a clear picture of the value and effects of Big Data in education. We study how is the value potential of Big Data of recent years and what will be the development in the near future.

Finally we are analyzing the learning benefits from Big Data and Open Data giving a brief description of how these technologies can contribute to a renowned education system. Helping teachers and students to make more targeted choices in the sector of education.

2. Big Mount of Data that is not Used

The Computing technology in the last 20 years has fast grown replacing all the traditional methods. One of the missions of Computing after all is to shares content. More business and government agencies are discovering the strategic uses of large databases. New software tools and techniques are assumed to analyze the data for beneficial inferences; a radically new type of "knowledge infrastructure" is materializing [1].

As Bollier and Firestone says, we can also mention, Computer databases are very used the last decades. Data is not just a back-office, accounts-settling tool any more. And can be used as a real-time decision-making tool. Researchers using progressive interaction methods can now tease out possibly useful kinds of information that would otherwise remain hidden in petabytes of data (a petabyte is a number starting with 1 and having 15 zeros after it) [1].

Furthermore, the web establishes new needs for information development. The number of information on the web is highly growth, as well as the number of new users using more, web research. People are likely to surf the web, starting with high quality human maintained indexes such as Yahoo! or with search engines, like Google [2].

According to a report, the amount of digital content on the Internet is now close to five hundred billion gigabytes (for the Year 2010). This amount is expected to double within a year. Ten years ago, a single terabyte of data seemed like a huge amount of information. Now, we commonly hear of data stored in petabytes. Some even talk of exabytes or the yottabyte, which is a trillion terabytes or, as one website describes it, "everything that there is." [1].

From the Year 2012, about 2.5 exabytes of data are produce each day, and that number is doubling every 40 months or so. More data cross the internet every second than were stored in the entire internet just 20 years ago. This gives businesses a possibility to work with many petabyes of data in a single data set—and not just from the internet. For example, it is approximated that Walmart

collects more than 2.5 petabytes of data every hour from its customer transactions. A petabyte is one quadrillion bytes, or the equivalent of about 20 million filing cabinets' worth of text. An exabyte is 1,000 times that volume, or one billion gigabytes [3].
The true profit of such data lies in the users' capability to select helpful reports, spot attractive events and trends, support decisions and policy based on statistical analysis and reasoning, and exploit the data to achieve business, practical, or scientific goals. When the scale of data manipulation, exploration, and inference grows beyond human skills, people look to computer technology to make the things easier [4].

Across a big range of fields, data are being collected and expand with fast speed. There is a critical need for a new creation of computational theories and tools to helps humans in taking useful information (knowledge) from the extremely spreading volumes of digital data. Where can found a lot of applies such as education [5].

We are overflowing in an overflow of data today. In an expansive variety of application areas, data is being gathered at extraordinary scale. Decisions that earlier were established on estimation, or on thoroughly created models of reality, can now be made established on the data itself. Such Big Data analysis now efforts nearly every side of our present culture, containing mobile services, retail, manufacturing, financial services, life sciences, physical sciences and of course education [6].

3. Definition of Big Data

So, what is Big Data? We can give some definitions about them. We can say one aspect of them is the expansion of mobile networks, cloud computing and a rise of unintelligibly of big amounts of data, are described as "Big Data." [1].

Furthermore is the description of state-of-the-art techniques and technologies to catch, collection, allocate, accomplish and explore petabyte- or larger-sized datasets with high-speed and varied patterns that predictable data management methods are unable of control [7].

What make most Big Data big are frequent explanations over time and/or space. The Web log archives lots of visits a day to a bit of pages. The cellphone database stores time and position frequently, for each mobile user. The seller has thousands of stores, tens of thousands of products, and millions of consumers but logs billions and billions of separate connections in a year [8].

According to the previous definitions, Big Data can provide advanced parallel techniques, such as, data analysts (both human and machine), massive swaths of data to forecast situations like activities and procedures in ways unimagined only years earlier [1].
Another aspect as Siemens and Long said, "Big Data are datasets whose size is beyond the ability of typical database software tools to capture, store, manage and analyze." In this case new technology techniques are used like Hadoop framework, where is used for the Map-Reduce processes and Databases techniques like Data-Mining for extracting knowledge from Databases [9].

Hence, Big Data has confirmed the size to develop estimates, save money, boost productivity and improve decision-making in fields such as traffic control, weather forecasting, disaster prevention, finance, fraud control, business transaction, national security, health care and education [7]. Finally as we can see from the above, all have as a result many basic changes with the data. Creating a new era in which all processes and interactions including scientific research are affected [9].

3.1 Big Data and Open Data

As McKinsey Global Institute mentions Big Data refers to data sets that are voluminous, diverse, and timely. On the other hand Open Data is actually Big Data but smaller and the information is open to everyone [10].

As we can see Open Data comes mostly from Government data sets or other institutions and enterprises, and from individuals. Actually the public sector feeds the Open Data and in the public sector includes schools, Universities and other Educational Institutes [10].

In Figure 1 we can see the distribution from all the global data; we can see the Big Data are a major part of them. Also in the Big Data included the Open Data and then the Open Government Data. Finally, the open data concept is associated with "MyData," which involves sharing information collected about an individual or organization with that individual, such as hospitals [10].

So as we can make an overview, Open Data are an important part of Big Data Category and have a major impact of them and also to our study, as we can see below.

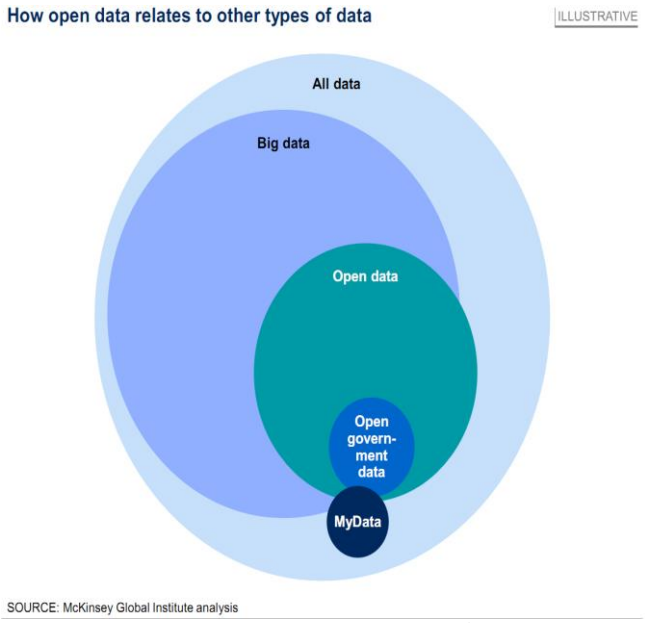

Fig. 1 How open data relates to other types of data [10]

4. Goals and Purposes of Big Data in Education

4.1 The Value of Big Data in Education

Big Data has the future to change not just research, but also education. A late accurate significant similarity of many approaches taken by 35 charter schools in NYC has discovered that one of the top five policies connected with significant academic effects was the use of data to guide instruction [6]. Other collaboration Technologies which the Big Data are based on them is the Cloud Computing. These technologies can improve educational services, giving young and adult students alike access to low-cost content, online instructors, and communities of fellow learners [11].

Furthermore, as West says "Big Data can support the classic educational system helping teachers to analyze what students know and what techniques are most effective for each pupil." In this way, also teachers are able to learn new techniques and methods about their education work [12].

Hence Technologies such as Data mining and Data analytics can provide a fast feedback to students and teachers about their academic performance. These methods can provide a deep analysis of some education patterns and extract valuable knowledge from them [12]. In this way, collective and big scale data can predict who student needs more help from the education system, avoiding the danger of failure or drop out [12],[13]. This has as a result to find pedagogic approaches that seem most effective with particular students and special needs [12].

On the other hand, as Siemens and Gasevic say "Big Data can easily find apply at online education." As we can see, the online education has a very big development at recent years and has a very increasing impact of the education sector. Furthermore digital learning is actually a collection of data and analytics which can contribute to teaching and learning. In this way many students participate in online or mobile learning, where are crated new data [13]. These new data, also with the help of social networks, are helping the students with the different background to correlate between them and help them to understand core course concepts [13].

4.2 The Effects of Big Data in Education

Below we can see how the Big Data effects to Education. As we can see from Figure 2, educational services for the years 2000 to 2008 in Unites States, have a small impact from the Big Data compare to other industries.

A reason of this negative productivity growth is probably the educational system meets strong systemic obstacles to increasing productivity and be affected positive from Big Data. Another reason is education still meets a lack of data-driven mind-set and available data [14].

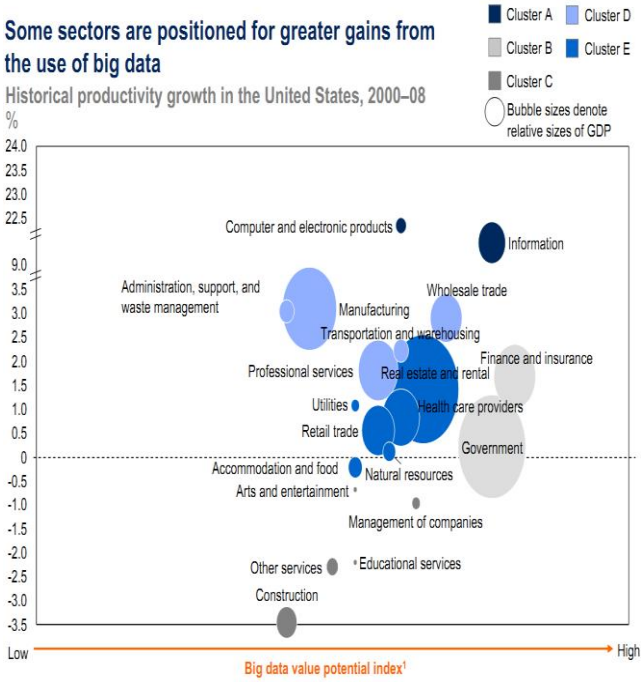

Fig. 2 Some sectors are positioned for greater gains from the use of big data. Historical productivity growth in the United States, 2000–2008 [14]

Although this not so strong effect of Big Data in education, if we will see in the near future a growth of data skills personnel in education like companies, is very possible education meets an tremendous effect from Big Data [15].

Additionally big companies are investing and training their personnel to have familiarity to Big Data as 1.5 million managers and analysts are deeply understanding of how big data can be applied [15]. So we can conclude also the education will have the same way.

As we can see from the Figure 3 the employment with deep analytically talent and skills in 2009 in education was 7.9 thousand. A very small number comparing to other industries and of course the increasingly requirements that come in the near future for Data Analyzing in education. We can definitely see Data Scientists and Data Analysts will be the most needed specializations in the near future for all the industry and most for education.

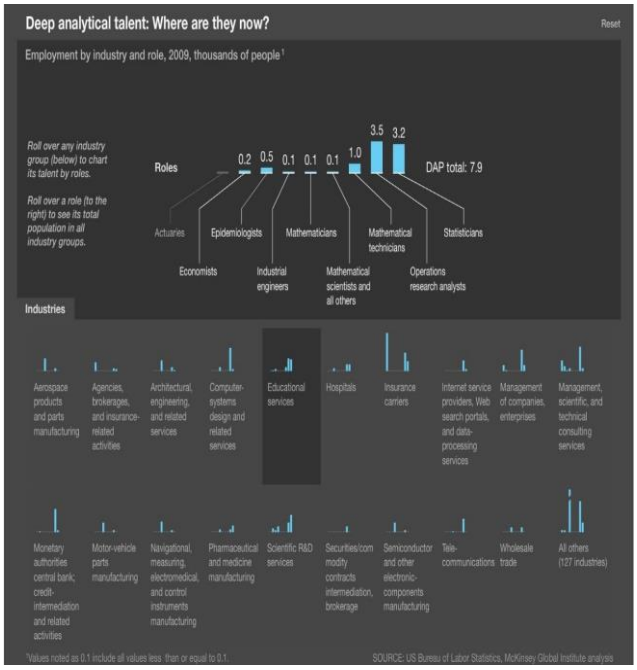

Fig. 3 Deep analytical talent: Where are they now? [16]

Also as we can see at the Figure 4 which is the use of Data Analytics with Big Data the education compared to other industries has only a 3%. But also other industries such as advertising, computer manufacturing and utilities have the same use and the rest of them are at 4% - 7%. So we can see the use of Data Analytics with Big Data in education is not bad compared to others.

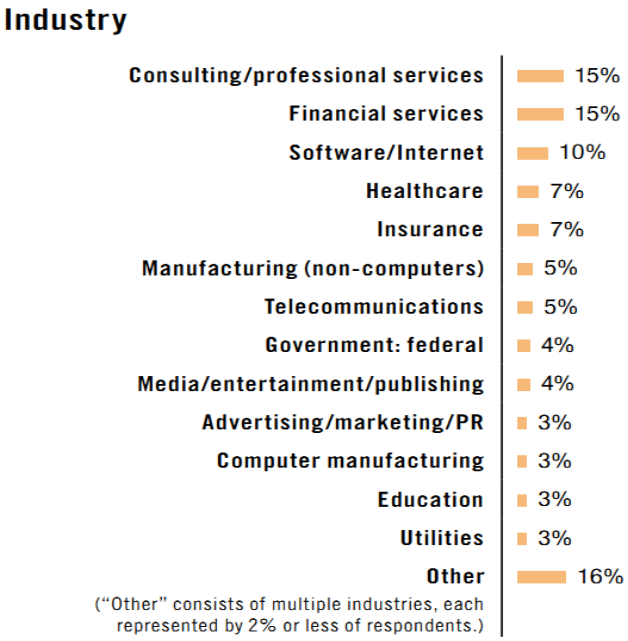

Fig. 4 Use of Data Analytics with Big Data in Industry [17]

Furthermore as we can mention from Figure 5, Companies in the sector of education has a fair data growth. For the Year 2009 the stored data in the United States are 269 Petabytes as the data stored per firm are 319 Terabytes. This amount of data will expanded in the coming years so the need for more educated personnel in Big Data will be absolute needed and the number of employees per firm will be much more than 843 employees for the Year 2009.

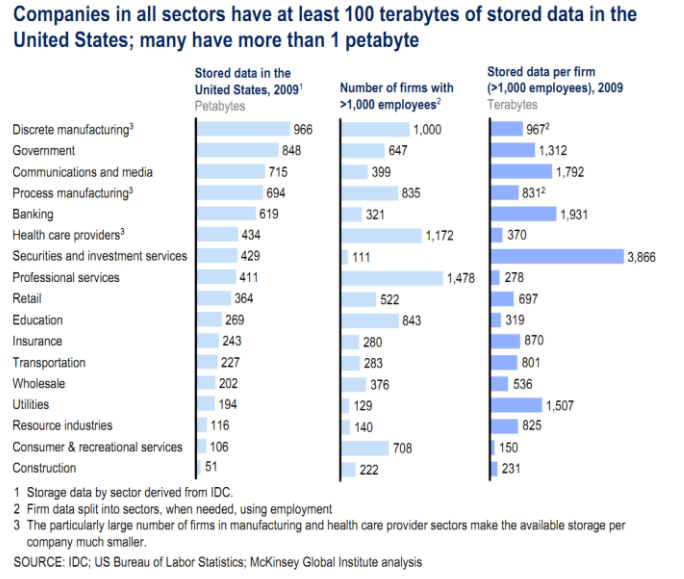

Fig. 5 Companies in all sectors have at least 100 terabytes of stored data in the United States; many have more than 1 petabyte [14]

5. Benefits of Big Data and Open Data in Education

5.1 Learning Benefits from Big Data and Open Data

Except from making education more personal and executive, also new types of data help researchers' ability to learn about learning. All these data comes from online courses or other technology-based learning platforms. In this case the analytics from them can improve students' ability learning and guide them to more efficient results than the traditional education [18].

In this case Big Data can provide more opportunities for new learning experience for children and young adults. Hence students can share information with educational institutions in this way they can expand their knowledge and skills. Furthermore, Educational institutes and Universities are able to help and prepare their future students [18].

Below we can see in summary some benefits of Big Data.

Table 1: Benefits of Big Data and Open Data in Education [10]

Benefits of Big Data and Open Data in Education	
1. Improved instruction	Can improve students' performance and learning abilities making the lessons more personal. The courses can be adjusted from the teachers with the help of analytics.
2. Matching students to programs	Open Data are able to help parents and students to find the best school or educational program.
3. Matching students to employment	Companies and candidate employees can discover alternative and more effective tools to use open data to qualify their skills with the needed skills. Also students can find and make applications for jobs which can match with their abilities, more efficient than before.
4. Transparent education financing	This leaves to students to participate in education activities, which previously they don't have the ability. Furthermore are able to choose anything about higher education and to discover the most proper education programs for them.
5. Efficient system administration	School education systems are able to develop a skillful school supply which can help administrators to allow more affective education resources. In that way this secures a high performance and afford to a versatile and smart plan for future education interests.

6. Conclusions

As we can see from the previous chapters, Big Data can really improve the education. Can afford to shape a modern and dynamic education system, which every individual student can have the maximum benefit from that. Furthermore teachers have valuable tools, were they do not have before, which can make their decisions more specific and are able to choose a big variety of new learning methods.

Hence the Big Data are actually involved to change the way of industries including the education. In the new era of Data the traditional difficulties will be no longer exists, keeping the good methods. The education system will be enriched with new learning ways, making more efficient and targeted.

But the way of this new era, have just began and there are many difficulties such as the lack of experienced personnel on the science of Big Data and Data analytics. Furthermore the teachers and academics must actually train and involved on them and finally the students must accept and use these new tools.

References

[1] D. Bollier and C. M. Firestone, The promise and peril of big data. Aspen Institute, Communications and Society Program, 2010.
[2] S. Brin and L. Page, "The anatomy of a large-scale hypertextual Web search engine," Comput. Netw. ISDN Syst., vol. 30, no. 1, pp. 107–117, 1998.
[3] A. McAfee and E. Brynjolfsson, "Big data: the management revolution," Harv. Bus. Rev., vol. 90, no. 10, pp. 60–66, 2012.
[4] U. Fayyad, G. Piatetsky-Shapiro, and P. Smyth, "The KDD process for extracting useful knowledge from volumes of data," Commun. ACM, vol. 39, no. 11, pp. 27–34, 1996.
[5] U. Fayyad, G. Piatetsky-Shapiro, and P. Smyth, "From data mining to knowledge discovery in databases," AI Mag., vol. 17, no. 3, p. 37, 1996.
[6] A. Bhatia and G. Vaswani, "BIG Data–A Review."
[7] J. Yan, "Big Data, Bigger Opportunities," 2013.
[8] A. Jacobs, "The pathologies of big data," Commun. ACM, vol. 52, no. 8, pp. 36–44, 2009.
[9] G. Siemens and P. Long, "Penetrating the fog: Analytics in learning and education," Educ. Rev., vol. 46, no. 5, pp. 30–32, 2011.
[10] "Open data: Unlocking innovation and performance with liquid information | McKinsey & Company." [Online].

Available: http://www.mckinsey.com/insights/business_technology/open_data_unlocking_innovation_and_performance_with_liquid_information. [Accessed: 19-Jun-2014].

[11] J. Bughin, M. Chui, and J. Manyika, "Clouds, big data, and smart assets: Ten tech-enabled business trends to watch," McKinsey Q., vol. 56, 2010.

[12] D. M. West, "Big Data for Education: Data Mining, Data Analytics, and Web Dashboards," Gov. Stud. Brook. US Reuters, 2012.

[13] G. Siemens and D. Gasevic, "Guest Editorial-Learning and Knowledge Analytics.," Educ. Technol. Soc., vol. 15, no. 3, pp. 1–2, 2012.

[14] "Big data: The next frontier for innovation, competition, and productivity | McKinsey & Company." [Online]. Available: http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation. [Accessed: 19-Jun-2014].

[15] B. Brown, M. Chui, and J. Manyika, "Are you ready for the era of 'big data'?," McKinsey Q., vol. 4, pp. 24–35, 2011.

[16] J. Manyika, M. Chui, B. Brown, J. Bughin, R. Dobbs, C. Roxburgh, and A. H. Byers, "Big data: The next frontier for innovation, competition, and productivity," McKinsey Glob. Inst., pp. 1–137, 2011.

[17] P. Russom, "Big data analytics," TDWI Best Pract. Rep. Fourth Quart., 2011.

[18] J. PODESTA, P. PRITZKER, E. MONIZ, J. HOLDREN, and J. ZIENTS, "BIG DATA: SEIZING OPPORTUNITIES, PRESERVING VALUES," Executive Office of the President, The White House Washington, Study, May 2014.

Athanasios Drigas is a Senior Researcher at N.C.S.R. Demokritos. He is the Coordinator of Telecoms and founder of Net Media Lab since 1996. From 1985 to 1999 he was Operational manager of the Greek Academic network. He has been the Coordinator of Several International Projects, in the fields of ICTs, and e-services (e-learning, e-psychology, e-government, e-inclusion, e-culture etc). He has published more than 260 articles, 7 books, 25 educational CD-Roms and several patents. He has been a member of several International committees for the design and coordination of Network and ICT activities and of international conferences and journals.

Panagiotis Leliopoulos is holding a Bachelor Degree at Electronic Engineering and a Master Degree on ICT from the University of Piraeus, Department of Digital Systems with direction on the Network-Centric Systems and Cloud Computing. He has experience on the Hadoop framework and Map-Reduce processes and developed a Big Data Business Intelligence application for the needs of his Master Thesis. He has published articles on e-Commerce, ICT and Education at international conferences and journals.