International Journal of Latest Trends in Engineering and Technology
Vol.(10)Issue(3), pp.035-039
DOI: http://dx.doi.org/10.21172/1.103.06
e-ISSN:2278-621X

A LITERATURE REVIEW ON ETHICAL ANALYSIS OF BIG DATA SHARING

Salman Khan[1], Manish Madhav Tripathi[2]

Abstract- Big data can be contrasted with small data, another evolving term that's often used to describe data whose volume and format can be easily used for self-service analytics. A commonly quoted axiom is that "big data is for machines; small data is for people."Many big data-driven companies today are moving to protect certain types of data against intrusion, leaks, or unauthorized eyes. This study analyzes the managing and sharing styles, more specifically in context to Big Data. It questions how Big Data is transforming and evolving so far with the change in technologies. The study will be conducted through the use of an interaction analysis. Its goal is to increase the amount of knowledge regarding current state of Big Data and how it relates to development in sharing of data. But how do you lock down data while granting access to people who need to see it? Ideal for both technical and non-technical decision makers, group leaders, developers, and data scientists, this paper will actually tell that what and how of everything in context to Big Data Sharing. Providing lock-down security while safely sharing data is a significant challenge for a growing number of organizations. This paper discovers new options and ways in which Big Data is constantly affecting our life. Big data is a term that describes the large volume of data – both structured and unstructured – that inundates a business on a day-to-day basis. But it's not the amount of data that's important. It's what organizations do with the data that matters. Big data can be analyzed for insights that lead to better decisions and strategic business moves. Apart from the enormous splash Big Data has made on development scene, I believe leveraging this predictive analysis can help our networking and sharing system effectively instruct and direct a new generation of data processing.
Keywords – Big Data, Ethical, Data Analysis, Sharing

1. INTRODUCTION

While there's nothing particularly new about the analytics conducted in big data, the scale and ease with which it can all be done today changes the ethical framework of data analysis. Developers today can tap into remarkably varied and far-flung data sources[1]. Just a few years ago, this kind of access would have been hard to imagine. The problem is that our ability to reveal patterns and new knowledge from previously unexamined troves of data is moving faster than our current legal and ethical guidelines can manage. Now it is possible to do things that were impossible a few years ago, and have driven off the existing ethical and legal maps. If fail to preserve the values, it care about in new digital society, then big data capabilities risk abandoning these values for the sake of innovation and expediency[2].

The rest of the paper is organized as follows. Introduction of the big data is covered in section I. History of big data and current considerations are explained in section II. Concluding remarks are given in section III.

2. BIG DATA HISTORY AND CURRENT CONSIDERATIONS

Big data is an evolving term that describes any voluminous amount of structured, semistructured and unstructured data that has the potential to be mined for information. Big data is often characterized by 3Vs: the extreme volume of data, the wide variety of data types and the velocity at which the data must be processed. Although big data doesn't equate to any specific volume of data, the term is often used to describe terabytes, petabytes and even exabytes of data captured over time[3].

2.1 Everything you need to know about the Big Data :

Big data is an elusive concept. It represents an amount of digital information, which is uncomfortable to store, transport, or analyze. Big data is so voluminous that it overwhelms the technologies of the day and challenges us to create the next generation of data storage tools and techniques. So, big data isn't new. In fact, physicists at CERN have been wrangling with the challenge of their ever-expanding big data for decades. Fifty years ago, CERN's data could be stored in a single computer[4]. So it wasn't usual computer, this was a mainframe computer that filled an entire building. To analyze the data, physicists from around the world traveled to CERN to connect to the enormous machine. In the 1970's, ever-growing big data was distributed across different sets of computers, which mushroomed at CERN. Each set was joined together in dedicated, homegrown networks. But physicists collaborated without regard for the boundaries between sets, hence needed to access data on all of these. So, bridged the independent networks together in CERNET[5].

[1] Student, Department of Computer Science and Engineering, Integral University, Lucknow, UP, India
[2] Department of Computer Science and Engineering, Integral University, Lucknow, UP, India

In the 1980's, islands of similar networks speaking different dialects sprung up all over Europe and the States, making remote access possible but torturous. To make it easy for physicists across the world to access the ever-expanding big data stored at CERN without traveling, the networks needed to be talking with the same language. It is adopted the fledgling internetworking standard from the States, followed by the rest of Europe, and established the principal link at CERN between Europe and the States in 1989, and the truly global internet took off! Physicists could easily then access the terabytes of big data remotely from around the world, generate results, and write papers in their home institutes. Then, wanted to share their findings with all their colleagues. To make this information sharing easy, we created the web in the early 1990's. Physicists no longer needed to know where the information was stored in order to find it and access it on the web, an idea which caught on across the world and has transformed the way we communicate in our daily lives[6].

During the early 2000's, the continued growth of our big data outstripped our capability to analyze it at CERN, despite having buildings full of computers. It was to start distributing the petabytes of data to collaborating partners in order to employ local computing and storage at hundreds of different institutes. In order to orchestrate these interconnected resources with their diverse technologies, developed a computing grid, enabling the seamless sharing of computing resources around the globe. This relies on trust relationships and mutual exchange. But this grid model could not be transferred out of the community so easily, where not everyone has resources to share nor could companies be expected to have the same level of trust. Instead, an alternative, more business-like approach for accessing on-demand resources has been flourishing recently, called cloud computing, which other communities are now exploiting to analyzing their big data[7].

It might seem paradoxical for a place like CERN, a lab focused on the study of the unimaginably small building blocks of matter, to be the source of something as big as big data. But the way we study the fundamental particles, as well as the forces by which they interact, involves creating them fleetingly, colliding protons in accelerators and capturing a trace of them as they zoom off near light speed. To see those traces, detector, with 150 million sensors, acts like a really massive 3-D camera, taking a picture of each collision event - that's up to 14 million times per second. That makes a lot of data. But if big data has been around for so long, why do suddenly keep hearing about it now? Well, as the old metaphor explains, the whole is greater than the sum of its parts, and this is no longer just science that is exploiting this[8].

The fact that can derive more knowledge by joining related information together and spotting correlations can inform and enrich numerous aspects of everyday life, either in real time, such as traffic or financial conditions, in short-term evolutions, such as medical or meteorological, or in predictive situations, such as business, crime, or disease trends. Virtually every field is turning to gathering big data, with mobile sensor networks spanning the globe, cameras on the ground and in the air, archives storing information published on the web, and loggers capturing the activities of Internet citizens the world over. The challenge is on to invent new tools and techniques to mine these vast stores, to inform decision making, to improve medical diagnosis, and otherwise to answer needs and desires of tomorrow's society in ways that are unimagined today.

2.2 How sharing Big Data has managed to change the world so far:

When have a large body of data, it can fundamentally do things that we couldn't do when only had smaller amounts. Big data is important, and big data is new, and when think about it, the only way this planet is going to deal with its global challenges — to feed people, supply them with medical care, supply them with energy, electricity, and to make sure they're not burnt to a crisp because of global warming — is because of the effective use of data. So what is new about big data? What is the big deal? Well, to answer that question, let's think about what information looked like, physically looked like in the past. In 1908, on the island of Crete, archaeologists discovered a clay disc. They dated it from 2000 B.C., so it's 4,000 years old.

Now, there's inscriptions on this disc, but we actually don't know what it means. It's a complete mystery, but the point is that this is what information used to look like 4,000 years ago. This is how society stored and transmitted information. Now, society hasn't advanced all that much. It still store information on discs, but now can store a lot more information, more than ever before. Searching it is easier. Copying it easier. Sharing it is easier. Processing it is easier. And what can do is can reuse this information for uses that never even imagined when first collected the data. In this respect, the data has gone from a stock to a flow, from something that is stationary and static to something that is fluid and dynamic. There is, if will, a liquidity to information. The disc that was discovered off of Crete that's 4,000 years old, is heavy, it doesn't store a lot of information, and that information is unchangeable. By contrast, all of the files that Edward Snowden took from the National Security Agency in the United States fits on a memory stick the size of a fingernail, and it can be shared at the speed of light. More data. More. Now, one reason why have so much data in the world today is we are collecting things that we've always collected information on, but another reason why is we're taking things that have always been informational but have never been rendered into a data format and we are putting it into data. Think, for example, the question of location. Take, for example, Martin Luther. If we wanted to know in the 1500s where Martin Luther was, we would have to follow him at all times, maybe with a feathery quill and an inkwell, and record it, but now think about what it looks like today. You know that somewhere, probably in a telecommunications carrier's database, there is a spreadsheet or at least a database entry that records your information of where you've been at all times. If you have a cell phone, and that cell phone has GPS, but even if it doesn't have GPS, it can record your information. In this respect, location has been data field.

Now think, for example, of the issue of posture, the way that you are all sitting right now, the way that you sit, the way that you sit, the way that you sit. It's all different, and it's a function of your leg length and your back and the contours of your back, and if I were to put sensors, maybe 100 sensors into all of your chairs right now, I could create an index that's fairly

unique to you, sort of like a fingerprint, but it's not your finger. So what could we do with this? Researchers in Tokyo are using it as a potential anti-theft device in cars. The idea is that the carjacker sits behind the wheel, tries to stream off, but the car recognizes that a non-approved driver is behind the wheel, and maybe the engine just stops, unless you type in a password into the dashboard to say, "Hey, I have authorization to drive." Great. What if every single car in Europe had this technology in it? What could we do then? Maybe, if we aggregated the data, maybe we could identify telltale signs that best predict that a car accident is going to take place in the next five seconds. And then what we will have datafied is driver fatigue, and the service would be when the car senses that the person slumps into that position, automatically knows, hey, set an internal alarm that would vibrate the steering wheel, honk inside to say, "Hey, wake up, pay more attention to the road." These are the sorts of things we can do when we datafy more aspects of our lives. So what is the value of big data? Well, think about it. You have more information. You can do things that you couldn't do before. One of the most impressive areas where this concept is taking place is in the area of machine learning. Machine learning is a branch of artificial intelligence, which itself is a branch of computer science.

The general idea is that instead of instructing a computer what do, we are going to simply throw data at the problem and tell the computer to figure it out for itself. And it will help you understand it by seeing its origins. In the 1950s, a computer scientist at IBM named Arthur Samuel liked to play checkers, so he wrote a computer program so he could play against the computer. He played. He won. He played. He won. He played. He won, because the computer only knew what a legal move was. Arthur Samuel knew something else. Arthur Samuel knew strategy. So he wrote a small sub-program alongside it operating in the background, and all it did was score the probability that a given board configuration would likely lead to a winning board versus a losing board after every move. He plays the computer. He wins. He plays the computer. He wins. He plays the computer. He wins. And then Arthur Samuel leaves the computer to play itself. It plays itself. It collects more data. It collects more data. It increases the accuracy of its prediction. And then Arthur Samuel goes back to the computer and he plays it, and he loses, and he plays it, and he loses, and he plays it, and he loses, and Arthur Samuel has created a machine that surpasses his ability in a task that he taught it. And this idea of machine learning is going everywhere. How do you think we have self-driving cars? Are we any better off as a society enshrining all the rules of the road into software? No. Memory is cheaper. No. Algorithms are faster. No. Processors are better. No. All of those things matter, but that's not why. It's because we changed the nature of the problem. We changed the nature of the problem from one in which we tried to overtly and explicitly explain to the computer how to drive to one in which we say, "Here's a lot of data around the vehicle. You figure it out. You figure it out that that is a traffic light, that that traffic light is red and not green, that that means that you need to stop and not go forward."

Machine learning is at the basis of many of the things that we do online: search engines, Amazon's personalization algorithm, computer translation, voice recognition systems. Researchers recently have looked at the question of biopsies, cancerous biopsies, and they've asked the computer to identify by looking at the data and survival rates to determine whether cells are actually cancerous or not, and sure enough, when you throw the data at it, through a machine-learning algorithm, the machine was able to identify the 12 telltale signs that best predict that this biopsy of the breast cancer cells are indeed cancerous. The problem: The medical literature only knew nine of them. Three of the traits were ones that people didn't need to look for, but that the machine spotted. Now, there are dark sides to big data as well. It will improve our lives, but there are problems that we need to be conscious of, and the first one is the idea that we may be punished for predictions, that the police may use big data for their purposes, a little bit like "Minority Report." Now, it's a term called predictive policing, or algorithmic criminology, and the idea is that if we take a lot of data, for example where past crimes have been, we know where to send the patrols. That makes sense, but the problem, of course, is that it's not simply going to stop on location data, it's going to go down to the level of the individual.

Why don't we use data about the person's high school transcript? Maybe we should use the fact that they're unemployed or not, their credit score, their web-surfing behavior, whether they're up late at night. Their Fitbit, when it's able to identify biochemistries, will show that they have aggressive thoughts. We may have algorithms that are likely to predict what we are about to do, and we may be held accountable before we've actually acted. Privacy was the central challenge in a small data era. In the big data age, the challenge will be safeguarding free will, moral choice, human volition, human agency. There is another problem: Big data is going to steal our jobs. Big data and algorithms are going to challenge white collar, professional knowledge work in the 21st century in the same way that factory automation and the assembly line challenged blue collar labor in the 20th century. Think about a lab technician who is looking through a microscope at a cancer biopsy and determining whether it's cancerous or not. The person went to university. The person buys property. He or she votes. He or she is a stakeholder in society. And that person's job, as well as an entire fleet of professionals like that person, is going to find that their jobs are radically changed or actually completely eliminated.

Now, we like to think that technology creates jobs over a period of time after a short, temporary period of dislocation, and that is true for the frame of reference with which we all live, the Industrial Revolution, because that's precisely what happened. But we forget something in that analysis: There are some categories of jobs that simply get eliminated and never come back. The Industrial Revolution wasn't very good if you were a horse. So we're going to need to be careful and take big data and adjust it for our needs, our very human needs. We have to be the master of this technology, not its servant. We are just at the outset of the big data era, and honestly, we are not very good at handling all the data that we can now collect. It's not just a problem for the National Security Agency. Businesses collect lots of data, and they misuse it too, and we need to

get better at this, and this will take time. It's a little bit like the challenge that was faced by primitive man and fire. This is a tool, but this is a tool that, unless we're careful, will burn us. Big data is going to transform how we live, how we work and how we think. It is going to help us manage our careers and lead lives of satisfaction and hope and happiness and health, but in the past, we've often looked at information technology and our eyes have only seen the T, the technology, the hardware, because that's what was physical. We now need to recast our gaze at the I, the information, which is less apparent, but in some ways a lot more important. Humanity can finally learn from the information that it can collect, as part of our timeless quest to understand the world and our place in it, and that's why big data is a big deal.

Looking Forward- The future of Big Data Sharing Well what we are looking for are the insights and when you see new insights in big data it creates a couple differentiations one is a customer experience differentiation great examples google and google maps. The Volume of generated data will continue to rise One of the most reassuring things when it comes to big data Hadoop future is that the amount of data generated every day will only continue to grow. As of now, we generate approximately 2.3 trillion gigabytes of data every day, and this will only grow in the future. If you notice, there are smartwatches, smart televisions, smart wearable techs in the market that further collect data from consumers, leaving the scope for only massive generation of data. Big Data Sharing and its integration with other technologies. As described earlier, the future of big data is clear and unshakeable. If you have noticed, technologies like IoT, Machine Learning, artificial intelligence and more are making their ways into our everyday lives. Behind all of these is Big Data sitting strong in an authoritative position. There are devices talking to each other over a connected network sharing and generating data you feed, and there are algorithms learning patterns and processing information from the generated data. A simple example of the Internet of Things is your smart television that is connected to your home network and generating data on your viewing patterns, interests and more. With social apps installed, it is also taking into considerations your personal tastes and preferences and cumulatively working on personas like yours to deliver better online content and streaming options. You would be amazed to know that the massive blockbuster House of Cards was the result of Big Data analytics!

Together, they are designed to offer the best of convenience and support to consumers and industries globally. A warehouse of Amazon is mostly automated, and there are tech companies that have replaced manpower with a simple code for monotonous jobs. As much as redundancy is killed by Big Data and analytics, newer opportunities are equally arising on the other side as well.To aid and guide the search for new and improved materials, computational materials science is increasingly employing “high-throughput screening” calculations. In practice, this means that computational material scientists produce a huge amount of data on their local workstations, computer clusters, and supercomputers using a variety of very different computer programs, in this domain usually called “codes”. Though being extremely valuable, this information is hardly available to the community, since most of the data is stored locally. In publications, typically, only a small subset of the results is reported, namely that which is directly relevant for the specific topic addressed in the actual manuscript. Although several repositories have been created and maintained in the past for domain-specific applications, these typically do not store the full inputs and outputs of all calculations.

3. DEMANDS IN NEAR FUTURE

While enthusiasts see great potential for using Big Data, privacy advocates are worried as more and more data is collected about people—both as they knowingly disclose things in such things as their postings through social media and as they unknowingly share digital details about themselves as they march through life. Not only do the advocates worry about profiling, they also worry that those who crunch Big Data with algorithms might draw the wrong conclusions about who someone is, how she might behave in the future, and how to apply the correlations that will emerge in the data analysis. There are also plenty of technical problems. Much of the data being generated now is “unstructured” and sloppily organized. Getting it into shape for analysis is no tiny task. Imagine where we might be in 2020. The Pew Research Center’s Internet & American Life Project and Elon University’s Imagining the Internet Center asked digital stakeholders to weigh two scenarios for 2020, select the one most likely to evolve, and elaborate on the choice. One sketched out a relatively positive future where Big Data are drawn together in ways that will improve social, political, and economic intelligence. The other expressed the view that Big Data could cause more problems than it solves between now and 2020. Respondents to our query rendered a decidedly split verdict.

53% agreed with the first statement:

Thanks to many changes, including the building of “the Internet of Things,” human and machine analysis of large data sets will improve social, political, and economic intelligence by 2020. The rise of what is known as “Big Data” will facilitate things like “nowcasting” (real-time “forecasting” of events); the development of “inferential software” that assesses data patterns to project outcomes; and the creation of algorithms for advanced correlations that enable new understanding of the world. Overall, the rise of Big Data is a huge positive for society in nearly all respects .

39% agreed with the second statement, which posited:

Thanks to many changes, including the building of “the Internet of Things,” human and machine analysis of Big Data will cause more problems than it solves by 2020. The existence of huge data sets for analysis will engender false confidence in our predictive powers and will lead many to make significant and hurtful mistakes. Moreover, analysis of Big Data will be misused by powerful people and institutions with selfish agendas who manipulate findings to make the case for what they want. And the advent of Big Data has a harmful impact because it serves the majority (at times inaccurately) while

diminishing the minority and ignoring important outliers. Overall, the rise of Big Data is a big negative for society in nearly all respects. Respondents were not allowed to select both scenarios; the question was framed this way in order to encourage a spirited and deeply considered written elaboration about the potential of a future with unimaginable amounts of data available to people and organizations. While about half agreed with the statement that Big Data will yield a positive future, many who chose that view observed that this choice is their hope more than their prediction. A significant number of the survey participants said while they chose the positive or the negative result they expect the true outcome in 2020 will be a little bit of both scenarios. Respondents were asked to read the alternative visions and give narrative explanations for their answers using the following guideline questions, "What impact will Big Data have in 2020? What are the positives, negatives, and shades of grey in the likely future you anticipate? How will use of Big Data change analysis of the world, change the way business decisions are made, change the way that people are understood?"

4. CONCLUSION

With big-data driven materials research, the new paradigm of materials science, sharing and wide accessibility of data are becoming crucial aspects. Obviously, a prerequisite for data exchange and big-data analytics is standardization, which means using consistent and unique conventions. The projected growth of data from all kinds of sources is staggering—to the point where some worry that in the foreseeable future our digital systems of storage and dissemination will not be able to keep up with the simple act of finding places to keep the data and move it around to all those who are interested in it. Tens of millions of connected people, billions of sensors, trillions of transactions now work to create unimaginable amounts of information. An equivalent amount of data is generated by people simply going about their lives, creating what the McKinsey Global Institute calls "digital exhaust"—data given off as a byproduct of other activities such as their Internet browsing and searching or moving around with their smartphone in their pocket. Human-created information is only part of the story, a relatively shrinking part. Machines and implanted sensors in oceans, in the soil, in pallets of products, in gambling casino chips, in pet collars, and countless other places are generating data and sharing it directly with data "readers" and other machines that do not involve human intervention.

Overall, the growth of the 'Internet of Things' and 'Big Data Sharing' will feed the development of new capabilities in sensing, understanding, and manipulating the world. However, the underlying analytic machinery (like Bruce Sterling's Engines of Meaning) will still require human cognition and curation to connect dots and see the big picture. And there will be dark episodes, too, since the brightest light casts the darkest shadow. There are opportunities for terrible applications, like the growth of the surveillance society, where the authorities watch everything and analyze our actions, behavior, and movements looking for patterns of illegality, something like a real-time Minority Report. On the other side, access to more large data can also be a blessing, so social advocacy groups may be able to amass information at a low- or zero-cost that would be unaffordable today. For example, consider the bottom-up creation of an alternative food system, outside the control of multinational agribusiness, and connecting local and regional food producers and consumers. Such a system, what I and others call Food Tech, might come together based on open data about people's consumption, farmers' production plans, and regional, cooperative logistics tools. So it will be a mixed bag, like most human technological advances. The good of Big Data will outweigh the bad. User innovation could lead the way, with "do-it-yourself analytics and sharing."

5. REFERENCES

[1] Bellamy, Bojana et al , "A Risk-based Approach to Privacy: Improving Effectiveness in Practice", The Centre for Information Policy Leadership,2014.

[2] C. Fan, F. Xiao, H. Madsen, D. Wang, "Temporal Knowledge Discovery in Big BAS Data for Building Energy Management", Energy and Buildings 190 (2015) 75–89.

[3] P.A. Mathew, L.N. Dunn, M.D. Sohn, "Big-data for building energy performance: Lessons from assembling a very large national database of building energy use", Applied Energy 140 (2015) 85–93.

[4] K. Zhou, C. Fu, S. Yang, "Big Data Driven Smart Energy Management: From Big Data to Big Insights", Renewable and Sustainable Energy Reviews 56 (2015) 215–225.

[5] L. Mashayekhy, M.M. Nejad, D. Grosu, "Energy-Aware Scheduling of MapReduce Jobs for Big Data Applications", Ieee Transactions on Parallel and Distributed Systems 26(10) (2015) 2720–2733.

[6] J. Cooper, M. Noon, C. Jones," Big Data in Life Cycle Assessment", Journal of Industrial Ecology 17(6) (2013) 796–799.

[7] G. Bello-Orgaz, J. Jung, D. Camacho, "Social Big Data: Recent Achievements and New Challenges", Information Fusion 28 (2016) 45–59.

[8] H. Fang, Z. Zhang, C.J. Wang, "A Survey of Big Data Research", Ieee Network 29(5) (2015) 6–9.

[9] B.T. Hazen, C.A. Boone, J.D. Ezell, "Data quality for data science, predictive analytics, and big data in supply chain management: An introduction to the problem and suggestions for research and applications", International Journal of Production Economics 154 (2014) 72–80.

[10] W.Q. Meeker, Y. Hong, "Reliability Meets Big Data: Opportunities and Challenges", Quality Engineering 26(1) (2014) 102–116.

[11] H. Ozkose, E. Ari, C. Gencerb, Yesterday, "Today and Tomorrow of Big Data", Procedia - Social and Behavioral Sciences 195 (2015) 1042 – 1050.

[12] J. Jose Camargo-Vega, J. Felipe Camargo-Ortega, L. Joyanes-Aguilar, "Knowing the Big Data", Revista Facultad de Ingenieria 24(38) (2015) 63–77.

[13] H. Li, K. Lu, S. Meng, "Big Provision: A Provisioning Framework for Big Data Analytics", Ieee Network 29(5) (2015) 50–56.

[14] LaValle, S., and Lesser, E. 2013. "Big data, analytics and the path from insights to value," MIT Sloan Management Review

[15] Markus, M. L., and Topi, H. 2015. "Big Data, Big Decisions for Science , Society, and Business.," Martin, K. E. 2015. "Ethical Issues in the Big Data Industry," MIS Quarterly Executive (14:2), pp. 67– 85 (doi: 1540-1960).