

# Big Data: Big Challenges to Privacy and Data Protection

Abu Bakar Munir, Siti Hajar Mohd Yasin, Firdaus Muhammad-Sukki

**Abstract**—This paper seeks to analyse the benefits of big data and more importantly the challenges it pose to the subject of privacy and data protection. First, the nature of big data will be briefly deliberated before presenting the potential of big data in the present days. Afterwards, the issue of privacy and data protection is highlighted before discussing the challenges of implementing this issue in big data. In conclusion, the paper will put forward the debate on the adequacy of the existing legal framework in protecting personal data in the era of big data.

**Keywords**—Big data, data protection, information, privacy.

## I. INTRODUCTION

WE are living in the age of “big data”. Data has become the raw material of production and a new source of immense economic and social values. The advances in data mining and analytics and massive increase in the computing power and data storage capacity have expanded, by orders of magnitude, the scope of information available to businesses, governments, and individuals. The volume of data stored and generated in the world is growing so fast that scientists have had to create new terms, including zettabyte and yottabyte, to describe the flood of data [1]. Anderson and Raini have indicated that [2]:

“We swim in a sea of data... and the sea level is rising rapidly. 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 by-product of other activities such as their internet browsing and searching or moving around with their smart phone in their pocket [2]”.

Big data present threats and opportunities for organisations as well as individuals. The organisations are not only exploring different ways to analyse, exploit and monetise the information contained within it but also have to grapple with the cost and risk and storing that data. With regard to individuals, on the one hand, most people now have instant access to vast amounts of information, which provides a wide range of benefits, including spurring innovation,

Abu Bakar Munir is a Professor of Law with the Faculty of Law, University of Malaya, 50603, Kuala Lumpur, Malaysia. He was the Advisor to the Malaysian Government on Data Protection. (phone: +603-796-6526; fax: +603-795-3239; e-mail: abmunir@um.edu.my).

Siti Hajar Mohd Yasin is an Associate Professor of Law with the Faculty of Law, Universiti Teknologi MARA, 40450, Shah Alam, Selangor DE, Malaysia (e-mail: sitihajar425@salam.uitm.edu.my).

Firdaus Muhammad-Sukki is a Lecturer with the Faculty of Engineering, Multimedia University, 63100, Cyberjaya, Selangor DE, Malaysia (e-mail: firdaus.sukki@mmu.edu.my, firdaus.sukki@gmail.com).

communication and freedom of expression. On the other hand, these new pools of data also include information about individuals and the use of big data tools to combine and analyse that information could result in privacy infringement on a massive scale [3]. Big data brings big benefits and promises. Nonetheless, the potential threats to privacy and data protection are too great to ignore.

## II. WHAT IS BIG DATA?

The world is experiencing a data revolution. Previously, a relatively small volume of analog data was produced and made available through limited number of channels. Today, a massive amount of data is regularly being generated and flowing from various sources, through different channels, every minute in today’s Digital Age. Research has shown that the amount of information stored each year grew to 161 exabytes per year, up from only 5 exabytes in 2003, roughly equal to the amount of information stored in 37,000 libraries, the size of the U.S. Library of Congress [4]. An exabyte is a quintillion bytes that are rapidly becoming passé[1]. Someone has calculated that if we loaded an exabyte of data on to the DVDs in slim line jewel cases, and loaded them into Boeing 747 aircraft, it would take 13,513 planes to transport one exabyte of data [1]. Therefore, using DVDs to move the data collected globally in 2010 would require a fleet of more than 16 million jumbo jets [1].

The McKinsey Global Institute (MGI) [5] estimates that the data volume has been growing by 40% per year, and will grow 44 times this rate between 2009 and 2020. MGI estimates that enterprises, globally, stored more than 7 exabytes of new data on disk drives in 2010, while consumers stored more than 6 exabytes of new data on devices such as the PCs and notebooks [5]. Every day, we create 2.5 quintillion bytes of data – so much that 90 per cent of the data in the world today has been created in the last two years alone [6].

This data comes from everywhere: sensors used to gather climate information, posts to social media sites, digital pictures and videos, purchase transaction records, and cell phones GPS signals to name a few [6]. This data is big data. Big data is being used in a range of socially and economically powerful ways: to drive important research in medicine and health care delivery, to model climate change impacts like the sea level rise and to help private companies and government agencies to detect fraud. But as with any new technology, big data raises big questions concerning the protection of personal data.

There is no standard definition of big data. Initially, the idea was that the volume of information had grown so large that the quantity being examined no longer fit into the memory that the computers use for processing, so engineers needed to revamp

the tools they used for analyzing it all [7]. O'Reilly defines big data as, "data that exceeds the processing capacity of the conventional database systems. The data is too big, moves so fast, or doesn't fit the structure of your database architectures." David Kellogg, meanwhile, simply defines it as being "too big to be reasonably handled by current/traditional technologies" [8]. A consulting and research firm, Mc Kinsey Global Institute (MGI) agrees with Kellogg's concept of big data and defines it as "datasets whose size is beyond the ability of typical database tools to capture, store, manage, and analyse" [5].

The EU Data Protection Working Party describes big data as the exponential growth in the availability and automated use of information: it refers to the gigantic digital datasets held by corporations, governments and other large corporations, which are then extensively analysed using computer algorithms [9]. Seeing at it from the technology's perspective, IDC defines big data in this manner [10]:

"Big data technologies describe a new generation of technologies and architectures, designed to economically extract value from very large volumes of a wide variety of data, by enabling high-velocity capture, discovery, and/or analysis".

According to the OECD, these definitions and many others are in continuous flux, as they depend on the evolving performance of available technologies. Volume is not the only important element of big data, apart from volume; the other characteristics of big data are velocity and variety. Gartner defines it as high-volume, high-velocity and/or high-variety information assets that demand cost-effective, innovative forms of information processing for enhanced insight, decision making, and process optimization. Thus, the three properties are – volume, velocity and variety, referred to as the three Vs [11].

Big data are big in all those aspects. IBM adds another V, which is veracity, to this interpretation. Considering it from the business standpoint, the IBM states [12], "a big data is more than simply a matter of size; it is an opportunity to find insights in new and emerging types of data and content, to make your business more agile and to answer questions that were previously considered beyond reach." The volume and velocity refer to the sheer quantity of big data available and the speed at which the data must be stored and/or analysed. The variety refers to the huge variation in the types and sources of big data, and veracity refers to the level of quality and trustworthiness that can be ascribed to a dataset.

Big data, according to the PWC encompasses of structured, semi-structured and unstructured information created inside a company or available for sale by the commercial data aggregators and for free by the governments – from demographic information about consumers to product reviews and commentary; blogs; content on social media sites; and data streamed 24/7 from mobile devices, sensors and tech-enabled devices [13].

The Networked European Software and Services Initiative (NESSI) is of the view that big data is a notion covering several aspects by one term, ranging from the technology base

to a set of economic models. The European technology platform adopted the following definition of big data [14]:

"A term encompassing the use of techniques to capture, process, analyze and visualize potentially large datasets in a reasonable time frame not accessible to the standard IT technologies. By extension, the platform, tools and software used for this purpose are collectively called big data technologies".

Simply put, big data is a term that describes large volumes of high velocity, complex and variable data that require advanced techniques and technologies to enable the capture, storage, distribution, management, and analysis of the information.

Viktor Mayer-Schonberger and Kenneth Cukier [7] state that big data is about three major shifts of mindset that are interlinked and hence reinforce one another. The first is the ability to analyze vast amounts of data about a topic rather than be forced to settle for smaller sets. The second is a willingness to embrace data's real-world messiness rather than privilege exactitude. The third is a growing respect for correlations rather than a continuing quest for elusive causality [7]. Big data relies on the increasing ability of technology to support the collection, storage, analyze, understand and take advantage of the full value of data.

Hence, there are many definitions of "big data" which may differ depending on whether you are a computer scientist, a financial analyst, or an entrepreneur pitching an idea to a venture capitalist. Most definitions reflect the growing technological ability to capture, aggregate, and process an ever-greater volume, velocity, and variety of data. In other words, "data is now available faster, has greater coverage and scope, and includes new types of observations and measurements that previously were not available." More precisely, big datasets are "large, diverse, complex, longitudinal, and/or distributed datasets generated from instruments, sensors, Internet transactions, email, video, click streams, and/or all other digital sources available today and in the future. More precisely, big datasets are "large, diverse, complex, longitudinal, and/or distributed datasets generated from instruments, sensors, Internet transactions, email, video, click streams, and/or all other digital sources available today and in the future [15].

### III. BIG DATA: THE NEXT BIG THING

According to [5], big data is "the next frontier for innovation, competition, and productivity." Big data, according to the firm, now has reached every sector in the global economy and it creates value in several ways: Firstly, creating transparency in the organizational activities that can be used to increase efficiency; Secondly, enabling experimentation to discover needs, expose variability, and improve performance; Thirdly, segmenting populations in order to customize actions; Fourthly, replacing/supporting human decision making with automated algorithms; and Fifthly, innovating new business models, products, and services. The 2011 report states [2], "Our research finds that data can create significant value for the world economy,

enhancing the productivity and competitiveness of companies and the public sector and creating substantial economic surplus for consumers.”

The United Nations states [16], “Big Data is a sea change that, like nanotechnology and quantum computing, will shape twenty-first century. ... Constitutes an historic opportunity to advance our common ability to support and protect human communities by understanding the information they increasingly produce in digital forms. Big data will affect the development works somewhere between significantly and radically, but the exact nature and magnitude of the change is difficult to project.”

The United Nations report [16] affirms that if properly analysed, big data offers the opportunity for an improved understanding of human behaviour that can support the field of global development in three main ways: firstly, early warning - early detection anomalies in how population use digital devices and services can enable faster response in time of crisis; secondly, real-time awareness – big data can paint a fine-grained and current representation of reality which can inform the design and targeting of programs and policies; and thirdly, real-time feedback – the ability to monitor a population in real time makes it possible to understand where policies and program are failing and make the necessary adjustments.

Big data benefits governments, companies, individuals and society at large. A report by the Tech America Foundation Big Data Commission [17] which focuses on the governments recognizes that the manner in which big data can be used to create value across the government and in the global economy is broad and far reaching. The Commission asserts [17], “We are at the cusp of a tremendous wave of innovation, productivity, and growth – all driven by the big data as citizens, companies, and government exploit its potential. ...Big data enables government organizations to be smarter to improve the productivity of the enterprise, and serve the needs of their stakeholders by improving decision-making in individual agencies and across the government eco-system.”

Governments around the world are beginning to appreciate the importance of big data for various purposes. The Obama Administration in 2012 [18] announced a new, multi-agency big data research and development initiative aimed at advancing the core scientific and technological means of managing, analyzing, visualizing and extracting information from large, diverse, distributed, and heterogeneous data sets. Under this initiative, six U.S. government agencies would spend more than \$200 million to help the government better organize and analyse large volumes of digital data. This project is designed to focus on building technologies to collect, store and manage huge quantities of data.

The Digital Agenda for Europe has the overall aim to create a sustainable and economic European Digital single market with a number of measures directed at the use of data sources in that region. It emphasizes on the importance of maximizing the benefits of public data and specifically the need for opening up public data resources for re-use. The EU Open Data Strategy encourages more openness and re-use of public

sector data. Neelie Kroes, Vice – President of the European Commission, responsible for the Digital Agenda at a conference on big data recently says [19]:

“Knowledge is the engine of our economy. And data is its fuel. For the public sector, better data allows services that are more efficient, transparent and personalized. For scientist, open results and open data allow new ways to share, compare, and discover: permitting whole new fields of research. For citizens, data is the key to more information and empowerment and to new services and applications”.

The Australian Government has issued a document, Big Data Strategy – Issues Paper [20] in March 2013 and acknowledges that, “The opportunity that big data presents to government agencies is in the potential to unlock the value and insight contained in the data agencies already hold via the transformation, facts, relationships and indicators. ...Of interest more broadly to agencies, big data analysis may provide profound insights into a number of key areas of society including health care, medical and other sciences, transport and infrastructure, education, communication, meteorology and social sciences.”

The benefits of big data to the society at large will be myriad, as big data becomes part of the solution to pressing global problems like addressing climate change, eradicating disease, and fostering good governance and economic development [7]. Big data has already been used for the economic development and conflict prevention. It has revealed the areas of the African slums that are vibrant communities of economic activity by analyzing the movements of cell phone users [7]. It has uncovered the areas that are ripe for ethnic clashes and indicated how refugees might unfold [7].

Big data applications are numerous and in various sectors, including healthcare, mobile communications, smart grid, traffic management, fraud detection, marketing, etc. The MGI demonstrates that the transformative effect that big data has had on the entire sectors ranging from healthcare to retail to manufacturing to political campaigns [5]. Big data helped the re-election of Obama as the U.S President [21]. During the final 18 months of campaign, a sprawling team of data and software experts sifted, collated and combined dozens of pieces of information on each registered U.S voter to discover patterns that let them target fund-raising appeals and ads to those most likely to respond [21].

The greater the amount of data that becomes available, the more informative the data gets. In fact, with enough data it is even possible to discover about a person’s future. Last year Adam Sadilek, a University of Rochester researcher and John Krumm, an engineer at Microsoft’s research lab, showed they could predict a person’s approximate location up to 80 weeks into the future, at an accuracy of above 80 per cent [21]. To get there, the pair mined what they described as a “massive data set” collecting 32,000 days of GPS readings taken from 307 people and 396 vehicles. Then they imagined the commercial applications, like ads that say “Need a haircut? In four days, you will be within 100 meters of a salon that will have a \$5 special at that time.” Sadilek and Krumm called

their system “Far Out” [21].

According to the Software and Information Industry Association (SIIA), there is one reason for all of the excitement about big data right now, which is, “Data-Driven Innovation (DDI) [4] presents tremendous economic and social value, capable of transforming the way we work, communicate, learn and live our lives.” Firms like Google, eBay, LinkedIn, Facebook and many more were built around big data from the beginning because they had massive amounts of data in the new and less structured formats.

Unsurprisingly, big data creates big market for the IT technology and services. The IDC [10] expects that the big data technology and services market to grow from \$3.2 billion in 2010 to \$16.9 billion in 2015. This represents a compound annual growth rate (CAGR) of 39.4% or about seven times that of the overall information and communication technology (ICT) market. Opportunities for vendors will exist at all levels of the Big Data technology stack including infrastructure, software, and services. Organizations that have begun to embrace Big Data technology and approaches are demonstrating that they can gain competitive advantage by being able to take actions based on timely, relevant, complete, and accurate information, rather than guesswork.

Meanwhile, Gartner projects that such data analytics and related capabilities will drive \$34 billion of IT spending in 2013 [4]. Further, these technologies are becoming an engine of job creation as businesses discover ways to turn data into revenue. By 2015, innovation around data is projected to help create 4.4 million IT jobs globally, of which 1.9 million will be in the U.S. Further, applying an economic multiplier to those jobs, Gartner expects that each big data IT job added to the economy will create employment for three more people outside the tech industry in the U.S., adding six million jobs to the economy [4]. The Centre for Economic and Business Research (CEBR) estimates that the value of big data to the UK economy alone amounted to 216 billion pounds and will create 58,000 jobs in the next five years.

Big Data is big business and will continue to be one of the more predominant areas of focus in the coming years from small start-ups to large scale corporations to government departments. Big data is the future and as briefly mentioned above, big data can be used to predict the future. The White House report on big data concludes[15]:

“Big data tools offer astonishing and powerful opportunities to unlock previously inaccessible insights from new and existing data sets. Big data can fuel developments and discoveries in health care and education, in agriculture and energy use, and in how businesses organize their supply chains and monitor their equipment. Big data holds the potential to streamline the provision of public services, increase the efficient use of taxpayer dollars at every level of government, and substantially strengthen national security.”

From a different perspective, if we want proof that big data is a big thing and the future, just look at the job pages. Inside we’ll find desperate appeals for statistics-savvy data scientists, and the salaries are often eye-watering [22]. The Former NSA

Director, General Keith Alexander said [23], “We’re living in the age of big data and we have to figure out how to harness it...That’s what the future is going to be about.”The National Association of Software and Services Companies (NASSCOM) of India [24] regarded big data as the next big thing acknowledging that the global Big Data market is expected to grow about 46 per cent to more than USD 25 billion by 2015. According to NASSCOM [24], the IT and IT-enabled services, including analytics, are expected to grow the fastest, at a rate of more than 60 per cent, with their share in the total Big Data market expected to increase to 45 per cent in 2015 from 31 per cent in 2011. Like others, NASSCOM is optimistic that Big Data is likely to continue to grow as an area which can deliver substantial benefits [24].

The World Economic Forum (WEF) seemed to suggest that Big Data is the next big thing. The 2014 of the WEF states [25], “But despite the sometimes exaggerated hype surrounding “big data,” the fundamental assertion is true: data—and the decisions driven by those data—now represent the next frontier of innovation and productivity.” The IDC [10] predicts that the market for big data technology and services will hit \$32.4 billion in 2017, nearly doubling the predicted size for 2015 and an impressive 10 times the size it was in 2010. That estimates includes infrastructure software such as security and data centre management, and high performance data analysis.

#### IV. IS PRIVACY AND DATA PROTECTION DEAD?

Some, especially technologists, argue that privacy is dead because of the big data, social networks, ICT technology, etc. In 1999, Scott Mc Nealy, the then CEO of Sun Microsystems, uttered his now famous remark to a group of reporters: “You have zero privacy anyway. Get over it.” Meanwhile, Michael Froomkin and Simon Garfinkel both spoke of the “death of privacy”. In the movie of *Enemy of the State* (1998), there is a dialogue “Privacy’s been dead for 30 years. . . The only privacy that’s left is the inside of your head. Maybe that’s enough?” In January 2010, Facebook Chief Executive, Mark Zuckerberg, declared the age of privacy to be over. He said that the rise of social networking online means that people no longer have an expectation of privacy and it was no longer a “social norm.” A month earlier, Google Chief Eric Schmidt expressed a similar sentiment. In November 2011, a documentary film, *Terms and Conditions May Apply*, asked the question: is privacy dead? And answers: if it’s not dead yet, it’s surely on life-support.

We are shocked when the founder of whistle blowing website WikiLeaks Julian Assange in a press conference warned us, the iPhone, BlackBerry, Gmail users:

“You’re all screwed. The reality is (that) intelligence contractors are selling, right now, to countries across the world mass surveillance systems for all those products... There is an international corporatised mass surveillance industry... This industry is, in practice, unregulated. Intelligence agencies, military forces and police authorities are able to silently, and en masse, to secretly intercept calls and take over computers without the help

or knowledge of the telecommunication providers.”

When the anguish of Julian Assange is reiterated by the UN Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, Frank La Rue, we have sufficient logical grounds to be worried enough. On 16 May 2011, he in his annual report stated that:

“States have used popular social networking sites, such as Facebook, to identify and to track the activities of human rights defenders and opposition members, and in some cases have collected usernames and passwords to access private communications of Facebook users.”

The world was awoken by Edward Snowden who leaked an estimated 200,000 files that exposed the extensive and intrusive nature of phone and internet surveillance and intelligence gathering by the US and its western allies. In reality, despite all these reports and declarations about its death, privacy is still alive and well. People, including the younger generation, still care about privacy. More and more countries either have enacted or modernised or enacting data protection laws around the world. Privacy and data protection is high on the agenda worldwide. President Barack Obama said [26];

“The open information platforms of the 21st century can also tempt institutions to violate the privacy of citizens. Dramatic increases in computing power, decreases in storage costs and huge flows of information that characterize the digital age bring enormous benefits, but also create risk of abuse. We need sensible safeguards that protect privacy in this dynamic new world.”

The President gave his commitment to strengthen privacy protections for the digital age and will harness the power of technology to hold government and business accountable for the violations of personal privacy. Recently, he reiterates this by saying [27], “One thing should be clear, even though we live in a world in which we share personal information more freely than in the past, we must reject the conclusion that privacy is an outmoded value. It has been at the heart of our democracy from its inception, and we need it now more than ever.”

Brandon Lynch, the Chief Privacy Officer of Microsoft, in the similar vein says, “So, privacy matters as much now as ever, arguably even more so in the future. Privacy is changing and we have to update how we think about it, and protect it in the digital age”. Disagreeing that privacy is dead, he further said [28], “And after all, some technology and internet companies today take the position that privacy is dead, or at least that privacy is an outdated concept that people need to get over so technology companies can help them reap the benefits of sharing as much information as possible. But we disagree that privacy is not relevant or desirable, in this sensor-driven, social everywhere, big data world that we are heading towards. People today expect strong privacy protections because they are increasingly aware of, and concerned about, the digital trails they leave behind online and indeed there's plenty of evidence that people still care deeply about privacy.”

Edward Snowden in his alternative Christmas Message on 25 December 2013 states, “A child born today will grow up with no conception of privacy at all. They'll never know what it means to have a private moment to themselves... And that's a problem because privacy matters, privacy is what allows us to determine who we are and who we want to be.”

In December 2013, the U.N. General Assembly unanimously adopted a resolution aimed at protecting the right to privacy against unlawful surveillance in the digital age in the most vocal global criticism of U.S. eavesdropping. The resolution affirms that the same rights that people have offline must also be protected online, including the right to privacy. It calls on the 193 U.N. member states to respect and protect the right to privacy, including in the context of digital communication, to take measures to end violations of those rights and to prevent such violations including by ensuring that national legislation complies with international human rights law. The resolution also calls on all countries to review their procedures, practices and legislation regarding the surveillance of communications, their interception and collection of personal data, including mass surveillance, interception and collection, with a view to upholding the right to privacy of all their obligations under international human rights law. The resolution calls on the U.N. members to establish or maintain independent and effective oversight methods to ensure transparency, when appropriate, and accountability for state surveillance of communications, their interception and collection of personal data.

Another indicator that privacy is still alive and well is that more and more countries around the world are either enacting or modernising their existing laws to cope with technology to better protect privacy and personal data. As at 30th May 2014, 103 countries have done so with the major addition of South Africa and Brazil which have enacted their laws last year. Over 20 countries currently have official Bills. All these indicate that privacy is still alive and imperative. However, undeniably, big data poses great challenges to privacy and data protection as well as the legal framework.

It is too simplistic to say that privacy is dead and people are unconcerned about the use of their personal data. Research commissioned by the International Institute of Communications [29] shows that people's willingness to give personal data, and their attitude to how that data will be used, is context-specific. That context depends on a number of variables, e.g. how far an individual trust the organisation, what information is being asked for, etc. Furthermore, The Boston Consulting Group [30] found that for 75% of consumers in most countries, the privacy of personal data remains a top issue, and that young people aged 18-24 are only slightly less cautious about the use of personal online data than older age groups.

#### V. CHALLENGES TO PRIVACY AND DATA PROTECTION

Although privacy and data protection is still alive and well, undoubtedly big data presents an immense challenge to the privacy and data protection law due to three defining features of big data. The first is the availability of data at a massive

scale collected not only online but through the use of mobile devices with location tracking capabilities and thousands of 'apps' that share data with multiple parties. Secondly, the use of high speed, high transfer rate computers, coupled with petabytes of storage capacity, resulting in cheap and efficient data processing, based on the cloud-computing model. The third feature is the use of new computational frameworks (such as Apache Hadoop) for storing and analyzing this huge volume of data [31]. As a result, big data analytics often has three main characteristics which are different to those of traditional processing: use of algorithms, using 'all the data' and repurposing data [32].

Big data poses significant risks for the protection of personal data and the right to privacy. With the advent of big data, systematic collection, storage and analysis of personal data has dramatically increased [14]. From internet logs, user information can be extracted that is accessible for surveillance and marketing purposes; identity management tools are now used on the Internet to track the identity of users; in the physical world cameras are used for surveillance; mobile phones send location information to the network providers; debit and credit card payment systems reveal the amounts spent and stores visited [14]. Store loyalty cards allow analysing consumer behaviour; and social media allow user to contact and access to pictures, videos and movies [14].

An increasing number of companies, established global players, SME's and start-ups, built their business models on using and selling user profiles generated from these data sources. Data mining tools sift through the data to find patterns in large collections of personal data, to identify individuals and to predict preferences and interests. These patterns and predictions are stored in company databases and combined with new data. Governments are also analyzing and exchanging information about their citizens [14]. Thus, big data raises concerns about the tracking and profiling of people and consumers. In this respect, the European Data Protection Supervisor states [33]:

"Big data promises big benefits for society in sectors ranging from entertainment and transport to health and energy conservation; but where it involves personal data it also implies big risks for the individual to whom the information relates. While many consumers may be becoming more and more 'tech savvy', most appear unaware of or unconcerned by the degree of the intrusiveness into their searches and emails as information on their online activities is logged, analysed and converted into revenue by service providers."

Each stage of big data lifecycle – collection, combination, analysis and use – has changed in recent years in a way that could present serious risks to individual privacy. The great data collection, great data combination, greater analysis, greater insights, and great use of information are the underlying challenges. Great data collection of information about individuals allows the digital foot prints be created, tracked and monitored. If multiple pools of data are combined, a more detailed profile of individuals can be built. Put simply, for example, Google knows almost everything about you –

what you want to know, what you are watching and where you live [15]. Greater analyses allow data to be combined and mined which can provide an insight into an individual's life. Finally, data collected for specific use or purpose may be used for other uses or purposes without the consent or knowledge of the individuals. As the White House Report puts it [34]:

"Big data drives big benefits, from innovative businesses to new ways to treat diseases. The challenges to privacy arise because technologies collect so much data (e.g., from sensors in everything from phones to parking lots) and analyze them so efficiently (e.g., through data mining and other kinds of analytics) that it is possible to learn far more than most people had anticipated or can anticipate given continuing progress."

The international instruments on data protection such as the OECD Guidelines, European Directive, APEC Privacy Framework and the national data protection laws around the world are based on a number of fundamental principles: (i) restrictions on further use – personal data must not be excessive and can only be processed for specific purpose, cannot generally be used for other purposes without the consent of the individuals), (ii) notice – individuals must be informed about any processing of their personal data, (iii) choice and legitimate purpose – personal data can only generally be used for certain specified situations, such as where the individual has given consent or where required by law, (iv) accuracy – organisations are required to ensure personal data are accurate, and (v) retention – personal data must not be kept longer than necessary and must be destroyed when the data has served the purposes of collection.

Viktor Mayor and Cukier [7] argue that with big data, the value of information no longer resides solely in its primary purpose but in its secondary purposes and uses. They further argue that in this context, the concept of notice and consent underlying the data protection laws around the world is no longer suitable as is often either too restrictive to unearth data's latent value or too empty to protect individuals' privacy.

In the similar vein, [34] states that, "notice and consent is defeated by exactly the positive benefits that big data enables: new, non-obvious, unexpectedly powerful uses of data. It is simply too complicated for the individual to make fine-grained choices for every new situation or app [34]. The Executive Office of the U.S President in its report (Podesta Report) [15] echoed this sentiment and states that in a technological context of structural over-collection, in which re-identification is becoming more powerful than de-identification, focusing on controlling the collection and retention of personal data, while important, may no longer be sufficient to protect personal privacy.

In the big data context, the businesses or organisations are not aware of the full potential uses of the data at the time of collection. Some argue that requiring the organisations to go back to the individuals to obtain their consents or to re-notify them of the new uses of the data may not be practical. Furthermore, according to [34], notice and consent requirement fundamentally places the burden of privacy

protection on the individual – exactly the opposite of what is usually meant by a “right.” Worse yet, if it is hidden in such a notice that the provider has the right to share personal data, the user normally does not get any notice from the next company, much less the opportunity to consent, even though the use of the data may be different. Furthermore, if the provider changes its privacy notice for the worse, the user is typically not notified in a useful way[14].

On a more fundamental issue, [31] questions the three longstanding regulatory assumptions of the data protection laws. The first is whether the personal data/non-personal data distinction remains viable. The second is whether anonymisation remains effective in protecting users against tracking and profiling. The third is whether data minimisation can survive the onslaught of big data. He argues that the idea that personal data processing must be restricted to the minimum amount necessary is inimical to the underlying thrust of big data, which discovers new correlations by applying sophisticated analytics techniques to massive data collection, and seek to do so free of any ex ante restrictions.

Referring to the situation in the USA, the Electronic Privacy Information Center (EPIC)[35], a public interest research center in Washington argues that the current big data environment poses enormous risk to the Americans. The center asserts that the on-going collection of personal information in the United States without sufficient privacy safeguards has led to staggering increases in identity theft, security breaches and financial fraud. Additionally, the use of personal information to make automated decisions and segregate individuals based on secret, imprecise and oftentimes impermissible factors presents clear risks to fairness and due process. The EPIC states that far too many organisations collect detailed information and use it with too little regard for the consequences. The current big data environment is plagued by data breaches and discriminatory uses of predictive analytics. According to the centre, the scenario in the US is dominated by three factors: firstly, commercial institutions collecting data have insufficient data security to protect the Americans’ privacy, secondly, students are particularly vulnerable to big data risks, and thirdly, government’s collection of big data is particularly problematic (the government has also abused big data).

## VI. CAN THE EXISTING LAW COPE WITH BIG DATA?

The debate is on-going in Europe and the US as to whether the existing privacy and data protection laws are able to cope with the technological changes particularly big data and address all the issues emerge as discussed above. There are conflicting views. The European Commission recognises that the dramatic technological changes have occurred since the Data Protection Directive (DPD) was first proposed and the Commission is very concerned with problems raised by the profiling and data mining [36]. However, the Commission believes that the current framework of the DPD remains sound as far as its objective and principles are concerned [36]. Recently on 11 June 2014, the EU Working Party reiterating this stance states [37], “At this stage, the Working party has no

reason to believe that the EU data protection principles, as they are principally enshrined in the Directive 95/46/EC, are fundamentally challenged by the development of big data.” Nevertheless, the Working party intends to carry out its own assessment of the development of big data on the basis of the EU legal framework.

Interestingly, the United Kingdom Information Commissioner’s Office (ICO) recently published a comprehensive report on “Big Data and Data Protection” [32] and concludes that the DPD and the UK law are still fit and adequate in today’s big data world. The ICO states [32]:

“Our view is that the basic data protection principles already established in the UK and EU law are still fit for the purpose in the big data world. The view that current data protection principles are not adequate underestimates their inherent flexibility. Applying those principles involves assessing the impact of the processing on the individuals and whether it is proportionate to the aim being pursued in any particular case. It is true that the current European data protection law was drawn up in the early days of the internet and it is right to look to update it to take into account of how personal data is processed now. However, this does not mean that the basic data protection principles are no longer fit for the purpose in the big data world, or that a new data protection paradigm is required.”

Remarkably, the ICO categorically states that [32], “We do not accept the argument that the data protection principles are not fit for the purpose in the context of big data. Big data is not a game that is played by different rules. There is some flexibility inherent in the data protection principles. They should not be seen as a barrier to progress, but as the framework to promote privacy rights and as a stimulus to developing innovative approaches to informing and engaging the public.”

Across the Atlantic, in the US, the Obama Administration, through the two reports (PCAST and Podesta) is of the view that the technology advances have made privacy and data protection laws obsolete. The Podesta report recommends regulations that encourage the responsible use of data and privacy protection, no matter what technology is used to collect the information. One of its specific recommendations is that for the Department of Commerce to take appropriate consultative steps to seek stakeholder and public comment on big data developments and how they impact the Consumer Privacy Bill of Rights (CPBR) and then devise the draft legislative text for consideration by the stakeholders and submission by the President to Congress.

The proposal has received mixed reactions. Microsoft in support of this proposal, calls for the Congress to enact a strong and comprehensive privacy legislation noting that although big data holds great promise for the technology, economy and people, such legislation can and should be part of unlocking that promise. The company states;

“The nation should move forward on privacy legislation now. A strong and comprehensive federal privacy legislation could establish a framework that

enables all the players to harness the potential of big data while respecting the privacy rights of those whose information contributes to the data. Microsoft has supported the privacy legislation at the federal level for many years and the rise of big data only increases the need for action.”

Similarly, the EPIC is of the view that the existing legal framework in the US is inadequate to protect individual privacy. The centre states [35], “Although the Privacy Act of 1974 anticipated many of the challenges that Big Data present, the current legal frameworks fail to safeguard the individual privacy by adequately implementing Fair Information Practices (“FIPs”) and adhering to the privacy enhancing techniques. Because Big Data has threatened individual privacy for many years, and the risks to Americans increase daily, it is imperative that this Administration confronts Big Data problems expeditiously. Congress should swiftly enact the Consumer Privacy Bill of Rights.”

On the other side, the Internet Association, which represents the companies like Google, Facebook and Yahoo advocated for a flexible and balanced self-regulatory responsible use framework[38]. The Association is of the view that whole changes to the United States’ existing privacy framework are unnecessary as existing regulations provide a strong yet flexible framework subject to enforcement by state and federal bodies [38]. According to the Association, any legislative proposal to address big data may create a precautionary principle problem that hinders the advancement of technologies and innovative services before they even develop [38].

## VII. CONCLUSION

We live in an exciting time, when the scale and scope of value that data can bring is coming to an inflection point, set to expand greatly as the availability of Big Data converges with the ability to affordably harness it. Hidden in the immense volume, variety and velocity of data that is produced today is new information – facts, relationships, indicators and pointers - that either could not be practically discovered in the past, or simply did not exist before [17]. This new information, effectively captured, managed, and analysed, has the power to change every industry including cyber security, healthcare, transportation, education and the sciences.

Big data with all the benefits, prospects and promises, at the same time, poses challenges to privacy and data protection. The legal frameworks that protect personal data and privacy of individuals have been put into question – whether or not they are able to address all the issues created by big data. The debates are on-going, particularly in the US (which has adopted sectoral legislation to protect data privacy) and Europe (which has adopted comprehensive legislation).

Fortunately or unfortunately, nothing much have been/being done to deliberate on this issue in other parts of the world. No or lesser talk or attention is given on the issue of privacy and data protection. The focus is very much on the big data projects, applications and projections. The non-existence of the legal framework to protect the personal data in many

countries contributes to this almost non-existence debate on privacy and data protection in the context of big data. In Malaysia, for example, the country is very keen to become the hub for big data and big data analytics for the region. The government recognises that big data and big data analytics can play an important role towards improving the nation’s growth and to drive efficiency and efficacy of the government’s plan towards achieving developed nation status by the year 2020. The big data framework is being developed and supposed to be ready and announced soon. It is imperative that the issue of privacy and data protection be taken into account based on the Personal Data Protection Act which has been put into effect early this year.

## REFERENCES

- [1] C. Kuner, F. H. Cate, C. Millard, and D. J. B. Svantesson, “The challenge of ‘big data’ for data protection,” *International Data Privacy Law*, vol. 2, no. 2, pp. 47–49, May 2012.
- [2] J. Q. Anderson and L. Rainie, “The future of the internet,” Washington DC, 2012.
- [3] R. Cumbley and P. Church, “Is ‘Big Data’ creepy?,” *Computer Law & Security Review*, vol. 29, no. 5, pp. 601–609, Oct. 2013.
- [4] Software & Information Industry Association, “Data-Driven Innovation,” 2013.
- [5] 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,” 2011.
- [6] IBM, “IBM - What is big data?,” 28-Nov-2014. (Online). Available: <http://www-01.ibm.com/software/data/bigdata/what-is-big-data.html>. (Accessed: 12-Dec-2014).
- [7] V. Mayer-Schönberger and K. Cukier, *Big Data: A Revolution That Will Transform How We Live, Work, and Think*. John Murray, 2013, pp. 1–242.
- [8] Datastax, “Big Data: Beyond the Hype Why Big Data Matters to You,” 2013.
- [9] European Commission, “Article 29 Data Protection Working Party Opinion 03 / 2013 on Purpose Limitation,” 2013.
- [10] D. Vesset, H. D. Morris, G. Little, L. Borovick, S. Feldman, M. Eastwood, B. Woo, R. L. Villars, J. S. Bozman, C. W. Olofson, S. Conway, and N. Yezhkova, “Worldwide Big Data Technology and Services 2012 – 2015 Forecast,” USA, 2012.
- [11] OECD, “Exploring Data-Driven Innovation as a New Source of Growth: Mapping the Policy Issues Raised by ‘Big Data,’” OECD Publishing, 222, 2013.
- [12] IBM, “Big Data at the Speed of Business | The Big Data Hub,” 2014. (Online). Available: <http://www.ibmbigdatahub.com/video/big-data-speed-business>. (Accessed: 12-Dec-2014).
- [13] PWC, “Capitalizing on the promise of Big Data: How a buzzword morphed into a lasting trend that will transform the way you do business,” 2013.
- [14] NESSI, “Big Data - A New World of Opportunities,” 2012.
- [15] Executive Office of the President, “Big Data - Seizing Opportunities, Preserving Values,” USA, 2014.
- [16] UN Global Pulse, “Big Data for Development: Challenges & Opportunities,” 2012.
- [17] TechAmerica Foundation, “Demystifying Big Data: A Practical Guide To Transforming The Business of Government Microsoft,” pp. 1–40, 2012.
- [18] Office of Science and Technology Policy - Executive Office of the President, “Obama Administration Unveils ‘Big Data’ Initiative: Announces \$200 Million In New R&D Investments,” 2012.
- [19] European Commission, “The Economic and social benefits of big data,” 2013.
- [20] Department of Finance and Deregulation - Australian Government, “Big Data Strategy – Issues Paper,” 2013.
- [21] MIT Technology Review, “Big Data Gets Personal,” *MIT Technology Review*, pp. 1–29, 2013.
- [22] J. Carter, “How important is big data?,” *TechRadarPro*, 2014.

- [23] F. Konkel, "Former NSA Director: Big Data Is the Future," *Nextgov*, 2014. (Online). Available: <http://www.nextgov.com/big-data/2014/05/former-nsa-director-big-data-future/84712/>. (Accessed: 15-Dec-2014).
- [24] NASSCOM, "Big Data: The Next Big thing," 2012.
- [25] World Economic Forum, "The Global Information Technology Report 2014 - Rewards and Risks of Big Data," 2014.
- [26] B. Obama, "Barack Obama: Connecting and Empowering All Americans," 2007. .
- [27] Q. Palfrey, "Internet Privacy: Protecting Consumers, Building Trust, Creating Jobs," *White House Office of Science & Technology Policy*, 2012. (Online). Available: <http://www.whitehouse.gov/blog/2012/02/24/internet-privacy-protecting-consumers-building-trust-creating-jobs>. (Accessed: 15-Dec-2014).
- [28] M. Smith, "Digital privacy in the big data era: Microsoft's data protection keynote," *Network World*, 2012. (Online). Available: <http://www.networkworld.com/article/2223607/microsoft-subnet/digital-privacy-in-the-big-data-era--microsoft-s-data-protection-keynote.html>. (Accessed: 15-Dec-2014).
- [29] International Institute of Communications, "Personal Data Management: The User's Perspective," 2012.
- [30] J. Rose, C. Barton, R. Souza, and J. Platt, "The Trust Advantage - How to Win with Big Data," 2012.
- [31] I. S. Rubinstein, "Big Data: The End of Privacy or a New Beginning?," *International Data Privacy Law*, vol. 3, no. 2, pp. 74–87, Jan. 2013.
- [32] Information Commissioner's Office, "Big data and data protection," 2014.
- [33] European Data Protection Supervisor, "The Preliminary Opinion of the European Data Protection Supervisor -Privacy and competitiveness in the age of big data: The Interplay between data protection, competition law and consumer protection in the Digital Economy," Brussels, 2014.
- [34] Executive Office of the President - President's Council of Advisors on Science and Technology, "Big Data And Privacy: A Technological Perspective," no. May, 2014.
- [35] Electronic Privacy Information Centre, "Comments of the Electronic Privacy Information Centre to The Office of Science and Technology Policy - Request for Information: Big Data and The Future of Privacy," 2014.
- [36] European Commission, "Impact Assessment," 2012.
- [37] I. Falque-Pierrotin, "ARTICLE 29 Data Protection Working Party - Letter to Mr John Pedosta," no. June. 2014.
- [38] The Internet Association, "Comments Concerning Big Data and the Consumer Privacy Bill of Rights." pp. 1–15, 2014.

optical solar concentrators and renewable energy policy. He has a number of papers in high impact factor journals, and has presented in a number of conferences related to his area. He carried out a number of technical research, market trend and financial analysis related to solar technologies for countries such as Malaysia, Japan and United Kingdom. He has excellent track record in collaborating with research universities in Malaysia Dubai, Saudi Arabia and the United Kingdom. Prior to joining the academia, he was a communication engineer in Malaysia's largest telecommunication company. He is now a Visiting Research Fellow in Glasgow Caledonian University.

**Professor Abu Bakar Munir (LLB, LLM)**, the former dean of the faculty of law at the University of Malaya, Malaysia is an internationally renowned scholar, expert and consultant on ICT Law and data protection law. He is also an associate fellow at the University of Malaya Malaysian Centre of Regulatory Studies and a visiting professor at several universities in Asia, Australia, New Zealand, the Middle East and Europe. Munir has been consulted by the governments and private entities in Malaysia and around the world. He was appointed the adviser to the government of Malaysia on data protection law in 2007 and was instrumental in crafting and the passing of the PDPA 2010. His other area of interests include nanotechnology law and policy, air and space law and renewable energy law and policy.

**Associate Professor Siti Hajar Mohd Yasin (LLB, LLM)** is an Associate Professor of Law, at the Faculty of Law, University Technology MARA, Shah Alam, Selangor D.E, Malaysia. She has been a full time lecturer at the university for more than twenty years, teaching various fields of laws. She graduated with a degree of Law, LLB (Hons) from the University of Malaya in 1987, received her post-graduate degree in Master in Comparative Laws (MCL) from the International Islamic University Malaysia (IIUM) in 1995. She also holds a post-graduate diploma in the Shariah Law and Practice from the IIUM. Associate Professor Mohd Yasin has researched, published and spoke at conferences, locally and internationally, and extensively. Her research interests include the Information and Communication Technology (ICT) Law, Nanotechnology Law, Constitutional Law and Environmental Law.

**Dr. Firdaus Muhammad-Sukki (MEng, PhD, MIET, ACGI)** has been an active researcher and is currently working at Multimedia University, Malaysia. His research interest is in solar energy, particularly in terms of