Big Data: Road Ahead for India

Madhukar Dayal, Sachin Garg and Rubaina Shrivastava

Abstract

Advancements in computing technologies make new platforms and large volumes of data available to businesses and governments to discover hidden underlying patterns in the data and creating new knowledge. While businesses need to embrace these technologies in order to stay ahead of competition, governments can reap great benefits in cost effectively delivering social services and bring about improvement in social development indices. However, before any new technology can become a powerful resource (for business or for government), there exists a fundamental need for extensive planning, such that one can chalk out a future trajectory, prepare for the changes to come, and invest prudently. Exploitation of Big Data platforms and technologies requires both corporate strategies and government policies to be in place much before the results would start pouring in. In this paper, we investigate the potential of available Big Data platforms and technologies, their current use by various governments, and their potential for use by the central and state Governments in India.

Keywords: big data, business strategy, government policy, social welfare.

1. Introduction

The world is changing: earlier understanding of the historical chain of events was viewed as knowledge but now its meaning has turned into being a capability to predict and influence the future, including the ability to diminish negative future outcomes and enhance positive ones. In one of its emerging forms, this science is known as Big Data.

There is no rigorous definition of Big Data. As pointed out by, Mayer-Schönberger & Cukier (2013, pp 7): "…the real revolution is not in the machines that calculate data but in data itself and how we use it". According to Gartner: "Big Data is high volume, high velocity and high variety information assets that demand cost-effective, innovative forms of information processing for enhanced insight and decision making". Morton, Runciman & Gordon (2014) reflecting on the characteristics of Big Data list them as: Volume, Velocity, Variety, Veracity and Value. Such voluminous data first accumulated from the astronomical and weather data (for example, collected from various satellites). Today, such data additionally comes from a wide variety of sources, such as, sensor data, web logs, data streams on the Internet, social media, customer transactions, etc.

Big Data is perceived as comprising structured, unstructured and semi-structured data. Amongst them the unstructured data lead, with an estimated share of over 95% in Big Data. Structured data are those that are systematically stored for retrieval, manipulation and analysis, for example, as in relational databases. Semi-structured data do not reside in relational databases but have some organisational properties making them easier to analyse. With a few alterations semi-structured data can often be reorganised in relational databases. XML is an example of semi-structured data. Unstructured data, on the other hand, do not follow any specified format and are largely void of meta-data, such as data from social media, emails, videos, photos and audio files data.

1.1 Understanding Big Data

To better understand and appreciate "Big Data", we should go back to what Diebold (2012) talks about Big Data being three things: the term ("firmly entrenched"), the phenomenon ("continuing unabated") and an "emerging" discipline. Thus, Big Data is many things to different people and it is imperative to understand it deeper before it can be put to use. Towards this end, we look at how Big Data is changing the paradigms of social science research (and thus the lenses through which we perceive the world) and follow this with how Big Data tools and techniques are being used to make smart policy and business decisions. This is followed by a deeper look at Big Data in the Indian perspective and the road ahead.

Big Data and Social Science Research Paradigms: The Debate: One of the big controversies about Big Data and its use for science was started by Chris Anderson, Editor in Chief of Wired magazine when he claimed "the end of scientific theory building and hypothesis testing" was near - "faced with massive data, this approach to science-hypothesize, model, test-is becoming obsolete" (Anderson, 2008). His contention was that Google had "conquered the advertising world" without knowing anything about the "culture and conventions" of advertising, merely on the assumption that better data, analyzed better would win. A similar theme has been evoked by Mayer-Schönberger & Cukier's (2013) claim that data analysis has now shifted from using a sample size of N (where N is a subset of the population (N << all)) to the entire population (N = all). As it is infeasible to collect data from the entire population, it is the accepted research practice to survey a statistically significant sample and extrapolate the findings to inform decisions applicable to the entire population (Agresti & Finlay, 2009; Hutcheson & Sofroniou, 1999; Salsburg, 2001; Velleman, 1997). However, Mayer-Schönberger & Cukier (2013) contend that with fixed sample sizes, one could not be sure that all population characteristics were accounted for, and extremely small groups might not even find a place in the sample. As $N \rightarrow all$, this is no longer the case and there is greater freedom in asking relevant questions. These claims have sparked off a vociferous debate on the role of theory in social science research, Big Data's contribution(s) to social science research, and even more on what big data detracts from social science research. One of the most infiuential participants in this debate have been boyd and Crawford. They rebutted Anderson's claim(s) by providing a bold new definition of big data and provoking a conversation around it.

1.2 boyd and Crawford's Definition

boyd and Crawford (2012) defined "Big Data" as a cultural, technological, and scholarly phenomenon that rests on the interplay of:

Technology: maximizing computation power and algorithmic accuracy to gather, analyze, link, and compare large data sets;

Analysis: drawing on large data sets to identify patterns in order to make economic, social, technical, and legal claims;

Mythology: the widespread belief that large datasets offer a higher form of intelligence and knowledge that can generate insights that were previously impossible, with the aura of truth, objectivity, and accuracy.

The idea that "Big Data" rests on "mythology" needs deeper digging into. Similar to the ideas put forth by Gitelman (2013) and Bollier and Firestone (2010), boyd and Crawford posit that "all researchers are interpreters of data". Data is not a "given" Gitelman (2013, ch 1), but is rather subject to a "cleaning" and "interpretive" process. Thus, their contention - just because we have more data, it is a myth to presume that our insights will be truthful, accurate or more objective. They argue that in the case of "Big Data", the four forces that regulate social systems - market, law, social norms and architecture (code in case of technology) are frequently at odds and it is necessary to provoke conversations around what it all means. To accomplish that, they list out six provocations that we use to review the current state of the "correlation-causation" debate.

Provocation 1: Big Data Changes the Meaning of Knowledge: boyd and Crawford posit that Big data profoundly changes our thinking at the epistemological levels by reframing key questions about knowledge and research processes. Numbers don't speak of themselves and others methods of studying phenomena can get blown away by the sheer force of numbers. This talks to the qualitative/quantitative divide also talked about by Manovich (2012) who talks about "deep data"-about a few people and "surface data" about lots of people. There are different and distinct questions that can be asked and answered using the two types of data, and researchers should be cognizant not to prefer one to the detriment of the other. They also point out that the tools often used to study Big Data phenomenon, for example Twitter and Facebook come with their own limitations and restrictions. They treat the study of society using the Big Data tools as analogous to "accounting tools" that "shape the reality they measure". The idea that tools limit what can be collected resonates with Vis (2013) who points out that the Application Programming

Interfaces (APIs) that are used by researchers to collect data limit what can be collected. Tufekci (2013) points out how the practice of using "hashtags" to filter tweets biases the data towards a particular demographic-those who use a particular hash tag are more "wedded" to the issue, and are thus different from the rest. Manovich (2012) cautions us against taking what is spoken on social media as authentic by telling about his personal experiences growing up in the erstwhile Soviet Union and how what was spoken out was very different from what was actually meant.

Provocation 2: Claims to Accuracy and Objectivity are Misleading: boyd and Crawford claim "all researchers are interpreters of data". Data is not a "given", existing in and of itself. Data is actively collected and sourced. As Desouza & Jacob (2014) point out that it is difficult to "recognize data in its unstructured form and then to understand how to 'connect it' to more conventional forms of data". Bollier & Firestone (2010, pg. 13) ask if the data represents an "objective truth" or are interpretations biased due to the way the data is "cleaned"? Visualizations also have judgments embedded within them (Bollier & Firestone, 2010, pp 11-12). Khoury & Ioannidis (2014) point out that "Big error" is another challenge with "Big Data". They say, "big data's strength is in finding associations, not in showing whether these associations have meaning. Finding a signal is only the first step". Thus, it is important to understand the biases and the limitations of the data.

Provocation 3: Bigger Data are not Always Better Data: Using the example of Twitter, boyd and Crawford point out that we cannot assume that Twitter users provide an appropriate sample as not everyone is on Twitter and "bots" also inhabit it. Also, results are not transferable between social networks due to the network's unique demographies (Ruths & Pfeffer, 2014). Sometimes, "smaller" data might be more relevant. This is similar to Manovich (2012)'s concept of "shallow" and "deep" knowledge. Lagoze (2014) has called out Mayer-Schoenberger and K. Cukier (2013) on their concept of N = all. It is an unachievable mathematical ideal as not everything can ever be measured. Sometimes, sampling the data is better as pointed out by Hal Varian who mentioned that for the economic studies Google undertakes, a random sample was good enough (Bollier and Firestone, 2010).

Provocation 4: Taken Out of Context, Big Data Loses its Meaning: Contextual integrity of the data is extremely important to gain value from the data. boyd and Crawford contend that people's real-world 'personal' networks are different from their 'articulated' and 'behavioral' networks traced out through data. Taylor & Schroeder (2014) provide an example of wrong inferences due to lack of context. Practical Big Data analysis requires the data to pass through multiple stages through the pipeline (Jagadish et al., 2014) and at each stage the data gets "repurposed, reprocessed, retrofitted, and reinterpreted" thus losing context on the way (Schintler & Kulkarni, 2014).

Provocation 5: Just Because it is Accessible Does not make it Ethical: Big Data is changing the perception of ethics. Earlier, individual decisions had specific and knowable outcomes. With the advent of Big Data, many can take actions without realizing how their actions impact others (Zwitter, 2014). Crawford & Finn (2014) talk about the use of social media (Twitter) sourced data in the context of crisis, where people share location data and other personal information because they want help. One of the bigger challenges is anonymity and re-identification. boyd and Crawford (2012) say that researchers should focus on "accountability", which is a much broader concept than "privacy". Privacy and anonymity continue to be the biggest challenges regarding the use of Big Data in social science and public policy.

Provocation 6: Limited Access to Big Data Creates New Digital Divides: The "digital divide" regarding Big Data can be thought to exist along three axes- (a) who creates the data, (b) who accesses the data, and (c) who has the resources to analyse said data. Hilbert (2013) talks about these axes as necessary, but not sufficient conditions to harness "Big Data" and the need to have institutional mechanisms (or appropriate policies) in place. Manovich (2012) also talks about the challenges of accessing data, as much of the "social media" data is only accessible to few, trained people within the companies. A similar point was made by Taylor, Schroeder, & Meyer (2014) in the context of income and other "sensitive" data that is only accessible to a few researchers. The challenge of who creates the data is important in a policy context to maintain equity and ensure that policy makers do not divert a disproportionate share of resources to those segments on whom data is available, at the expense of the data-poor.

Once we consider these research and perspective aspects, other questions that are needed to be asked are: *What is the data? Where is the data? Who owns it? Who has access to it? Is it technically, legally, ethically, politically feasible to run analytics on it? What does it cost to analyse this data? What are the expected benefits of this analysis? Do the benefits outweigh the costs?*

Thus, it is imperative to consider that all data is collected, stored and analysed in an environment circumscribed by policy. It is policy which sets the boundaries of what is acceptable or unacceptable in particular contexts. Various regulations like HIPPA, Children Privacy Act, FERPA etc. determine the limitations of data collection and analysis. Thus, we have an interesting paradox between data and public policy: Public Policy simultaneously bounds and is bounded by the data collection and analysis framework(s).

2. Big Data in the Public Sector

The public sector is a ripe area for applying the tools and techniques of Big Data to increase the efficiencies in the sector. This can happen in two ways: by using Big Data to improve programmatic outcomes, and to improve decision making.

Improving Programmatic Outcomes: Desouza & Jacob (2014) contend that most public sector data is of low complexity and hence organizations can improve their "programmatic outcomes". They provide an example of New York city where the mayor's office is linking together otherwise unconnected databases and mining them to identify areas of focus. As an example, the office is using predictive analytics to identify potential zoning violations and target inspections towards such potential violations. The key insight here is the idea of linking together disparate databases to get a more complex database, which may not possess large volume or velocity. However, predictive policing in the Los Angeles Police Department is an example of using "true" big data. The data in this case is a fusion of historical and real-time data (that includes real-time city and traffic camera feeds). This data is used to identify areas where crime may occur and concentrate resources in such areas. They point out that possibly one of the biggest insights of this project was the realization that video streams are just another form of data that can be connected to other data via geocoding. The public sector in US is deploying big data technologies in Postal services, health care and human services, and internal revenue services, to name a few.

Improving Decision Making: The underlying premise behind using big data to enhance decision making is providing feedback loops that allow citizens to engage with government and thus reveal preferences that are not revealed through the traditional polling process. Desouza & Jacob (2014) identify two different mechanisms that have been proposed to assess the peoples' will: prediction markets and sentiment analysis. However, these methods require new types of data, richer data and existing data cannot be shoe-horned to fit these paradigms, and thus need greater investment.

Prediction markets, designed to take advantage of the "wisdom of the crowds" work akin to a futures market, where the commodity being traded on is an event. Though such markets are part of well developed stocks and commodities markets, they are still akin to wagers. While wagering on stocks and commodities is accepted, wagering on politically sensitive issues is not and leads to ethical concerns.

Sentiment analysis draws on messages posted on social media like Twitter and Facebook as a means to understand the populace. The use of social media has been criticized by scholars as not being representative of the society (see Tufekci, 2013; boyd and Crawford, 2012; Ruths & Pfeffer, 2014). Another aspect is that social media can be manipulated as noted by Desouza & Jacob (2014). It should also be noted that sentiment analysis depends heavily on natural language processing, which is not as well developed for languages other than English. Thus the use of this technique in other countries and regions may not be as successful.

However, despite these shortcomings, we should still attempt to use these mechanisms as "another" input into the decision-making process. Some examples of how big data is being used as input into the decision making process follow.

Boston Street Bump is a project of Boston's Mayor's Office of New Urban Mechanics. It crowd sources road condition data using the accelerometer and GPS sensors of smartphones. This data is aggregated across users and used to fix short-term problems like potholes. The application found mention in Podesta et al., (2014). It was also reported that initially the app sent repair crews to wealthier neighbour hoods where people were more likely to carry smartphones, and this bias was fixed by first deploying it to city-road inspectors, who service all parts of the city equally; the public now provides additional supporting data (Podesta et al., 2014).

Flowminder used cell phones to track people's movement during the 2010 Haiti cholera outbreak (Taylor & Schroeder, 2014) and identify areas outside the capital (Port-au-Prince) at risk of cholera. This technique used near real-time (upto 12 hour) location data of cell phones (based on the phone-tower communication) from phone companies and extrapolated the results (considering that multiple people share a phone). These results were validated by comparing with on the ground data from local and UN agencies.

Billion Prices Project at MIT has created an "inflation index" by tracking online prices (Taylor & Schroeder, 2014). The initial idea behind this was to use actual on the ground prices in Argentina and compare them with officially released figures that could be potentially affected by political demands. Decuyper et al., (2014) have attempted to use indicators derived from mobile phone data (call detail records and airtime purchases) as food security indices in an African country.

2.1 Big Data and Public Policy: Key Challenges

From the foregoing, we note that big data can be used to analyze, formulate and monitor public policy in myriad ways. However, its use continues to be fraught with many challenges, chiefly related to privacy, discrimination and liability (Schintler & Kulkarni, 2014). Of these, however, privacy has the largest mindshare, owing in part to the Snowden affair. The primary fallout of Snowden's disclosures on the NSA's bulk collection of telecommunication metadata was that privacy concerns with big data came to the forefront in the eyes of the public. This also prompted a review of US Signals Intelligence and a mandate to look closely at the "challenges inherent in big data":"Look at how the challenges inherent in big data are being confronted by both the public and private sectors; whether we can forge international norms on how to manage this data; and how we can continue to promote the free flow of information in ways that are consistent with both privacy and security." - The White House, Review of US Signals Intelligence, Jan 17, 2014.

2.2 Privacy, Anonymity and Big Data

The researcher community's concerns over privacy issues are echoed by Lane & Stodden (2013): "privacy issues could stop bona-fide data collection and statistical research in its tracks". This does not imply that the researcher community is not concerned about privacy, but rather wants a "sensible structure for data access that ensures the goal of good science is attained while protecting confidentiality and respecting individual agency" (Lane & Stodden, 2013).

The major challenges impacting the usage of Big Data for Public Policy and Social sciences are:

Lack of data collection and management infrastructure. This challenge is even more pronounced in the developing world. The developing world does not have enough ears on the ground to collect high quality data. Most of the data that Big Data and development experts talk about is essentially digital exhaust of a very specific type-mobile phone call data records (UN Global Pulse, 2012). Though analysing these records has shown utility, there are inherent challenges in accessing these records. These records are all part of private businesses owned data systems and sharing them is fraught with competition and other issues. Though the records are collected in near real time, they are actually released after a significant amount of time (cf. D4D challenge

(Taylor & Schroeder, 2014)). Even when released, they are only accessible to a small number of researchers. Hence, even plucking the low hanging fruit in this case is non-trivial (Prydz, 2014).

Lack of Institutional mechanisms to curate the data and mediate access to it. Again, while in the US, researchers have access to resources curated and mediated by the ICPSR, such institutional arrangements are largely lacking in the developing world. Even, in the USA, presence of legislation overly focused on privacy aspects has prevented the linking together of administrative record data both across agencies and across states (Lane & Stodden, 2013; Lane & Schur, 2009; Lane & Schur, 2010).

Absence of infrastructure to support privacy preserving data mining, wherein the researchers do not need access to the raw data per se. Data enclaves (Abowd & Lane, 2004) are a possible solution, but have concerns regarding accessibility to a select few. Thus, some virtual solutions are needed.

Crisis of Reproducibility is illustrated by Google Flu Trends (Lazer et al., 2014). Both science and social science are passing through a crisis of reproducibility, partly because of entrenched notions on sharing and the incentive mechanisms, and partly because it is difficult or impossible to share. Studies like the Facebook contagion study (Kramer, Guillory, & Hancock, 2014) cannot be replicated (and thus criticized) outside the platform for which they were designed and same is the case with multiple other studies, many of which depend on data sourced from commercial entities mediated through Application Programming Interfaces (Vis, 2013; Ruths & Pfeffer, 2014).

Privacy and anonymity as Daries et al., (2014) say, the two are intimately linked and sides of the same coin. We need to clearly understand what needs to be protected and build strong policy foundations for the same. The consent framework that does not take care of end-use only instills a false sense of privacy, as it does not really protect end-use (Barocas & Nissenbaum, 2014; Mundie, 2014; Podesta et al., 2014).

Data Brokers or "omnibus information providers" are largely unregulated and hold detailed profiles on almost all citizens (Nissenbaum, 2010; Podesta et al., 2014). Whereas the government is not able to link together its own administrative databases together because of the "big brother is watching attitude", multiple data brokers acquire data from federal, state and county governments, and link them together to form extremely rich datasets (Washington, 2014). Ansolabehere & Hersh (2012) provide an illustration when they detail how the private firms they engaged "Catalist" and "Polimetrix" shared data amongst themselves to link various voting records, and shared de-identified records with the researchers.

Lack of Data Integrity or provenance which is a "cornerstone of credible science" (Lagoze, 2014) is a major challenge to the reproducibility and applicability of results. A lot of big data lacks provenance as (i) it has not been designed for research (social media data) (Lazer et al., 2014), (ii) it has been stripped of key identifiers (Podesta et al., 2014), or (iii) it has been "repurposed, reprocessed, retrofitted, and reinterpreted" (Schintler & Kulkarni, 2014) multiple times.

Till now we have looked at the various aspects of Big Data from the data and domain perspective. Now, we turn our gaze to the tools and techniques used to analyse this data and gain value.

3. Big Data Technologies and Challenges

The open source project Hadoop (by Apache Software Foundation) is a primary Big Data analytics platform which is built to operate on large distributed (high performance) compute clusters. *MapReduce*, the most popular function, is essentially a two stage fault tolerant analytical routine which distributes the data and task at hand, first, to various compute nodes, and integrates the results obtained later. This is done using the Hadoop Distributed File System (HDFS) (adopted from Google (distributed) File System or GFS). IBM's InfoSphere BigInsights and InfoSphere Streams are commercial platforms for analysis of big data at rest and in streams respectively. A survey on Big Data describing technologies, platforms, applications, and challenges with suggestions on designing Big Data systems is presented by Chen & Zhang (2014). For more details on the platform, we refer the interested readers to *Understanding Big Data: Analytics for Enterprise Class*

Hadoop and Streaming Data by Zikopoulos, Eaton, Deroos, Deutsch & Lapis (2015). Further developments for new analytical routines to add to the Hadoop family are also underway by several organisations including the open source community.

3.1 Challenges

Kambatla, Kollias, Kumar, & Grama (2014) point out that due to the scattered nature of Big Data it is difficult to store, process and analyse it at one place. Hence, it needs to be segregated and processed over different servers. But with such distributed databases there arises the complexity of privacy, fault-tolerance, security and access controls. Chen & Zhang (2014) highlight that the lack of awareness pertaining to Big Data poses serious threats to the nation's cyber security and is also a barrier to country's socio-economic development.

Big Data poses a serious challenge in regard to data complexity, large scale data integration, sheer volume and lack of availability of supporting high performance computing cluster (HPCC) hardware and software platforms to tackle the aforesaid challenges. For more detailed discussion on involved issues, we refer the interested readers to Big Data: Opportunities and Challenges by Morton, Runciman & Keith (2014).

3.2 A Techno-Legal Perspective

***Big Data: Information Security Panorama**:* Secure cyberspace has become an indisputable need. In the context of Big Data, all organisations involved in its life cycle must have robust information security frameworks, incorporating at least:

1. Limiting access through segregation and separation of duties with defined access rights restrictions and strict authentication and authorization parameters.
2. Use of data anonymization and a control on de-anonymization techniques while storing personal identifiable information (PID) or other sensitive information.
3. Establishing a trust boundary between data owners and data storage owners.
4. Implementation of sound access control policies and customized firewall configurations in parlance to the value and sensitivity of the information/data.
5. To conduct periodic internal and external security audits.
6. Real-time security monitoring to detect and respond to any alarming event.
7. Use of Fully Homomorphic Encryption (FHE) method, in order to keep a balance between need to perform operations on encrypted data packets and also keeping it secure while in transmission.
8. Cyber threat intelligence mechanisms.
9. Hosting critical information only in hardened host servers.

Mayer-Schönberger & Cukier (2013, pp 27) cite an example of 'Xoom', which is a firm holding a big name in the context of Big Data. Xoom analyses its transactions in totality and triggers an alarm if any suspicious behaviour is detected. Usually to detect malicious behaviour it works on pattern based detection techniques, which implies that whenever any suspicious behaviour is detected which appears to be against the 'normal' behaviour pattern of the firm, the software will raise an alert/alarm. Xoom provides 128 bit encryption protection for securing transactions on its website whether the user is logged in or not. Xoom is a Verisign certified site and a certified licensee of TRUSTe (www.xoom.com).

Big Data: Legal Panorama: There exist variety of sensitive information, such as, confidential organisational information, intellectual property (e.g., trade secrets), healthcare information (e.g., patient records or insurance information), personal financial information (e.g., employee salary details, social security details) which need to be protected from unauthorized disclosure, access, alteration or damage. Several nations have enacted laws to protect personally identifiable information, for example, European Union Data Protection Directive, Enhancing Privacy Protection Act, Health Insurance Portability and Accountability Act (HIPAA), and for protecting personal financial information the Gram-Leach-Bliley Act (GLBA). In India, there is a pressing need to frame and enact suitable data protection legislation and incorporate compliance mechanisms. As of now, personal data protection is covered by provisions in the Information Technology (Amendment) Act, 2008 under Sections 43, 66, & 72 and/or under the provisions of Indian Penal Code, 1860.

Big Data legal requirements and its sector wise applicability: The multiple layers of regulations can be implemented jointly or independently depending upon the case facts. In the context of Big Data, the applicability of the relevant law is ascertained on the basis of various factors, such as, the type of data (personal health, financial or corporate information). The applicable laws to select from are:

(1) Health Information Technology for Economic and Clinical Health Act (HITECH Act) - applicable to health care providers, health care clearing houses storing, processing, and exchanging electronically protected health information (e-PHI). HITECH Act widens the scope of privacy and security available under HIPAA. It further increases the potential liability in case of non-compliance and bestows better enforcement.

(2) Children's Online Privacy Protection Act - applicable to organizations collecting personal information of children (the age limit varies from country to country).

(3) CAN-SPAM Act 2003 - protects customers from targeted marketing campaigns of companies which results into unsolicited bulk emails.

Google's 'Usage of Big Data: A Strategic Business Purpose' example (in book Big Data: *A Business and Legal Guide* by Kalyvas & Overly): The analyses of Big Data is often for a purpose different than the one for which it was collected. Although, the commercial use of Big Data is apparent but organisations need to be transparent regarding its business purpose while using Big Data and ensuring that it does not exceed the defined purpose(s). Google narrowly escaped in a litigation, Authors Guild, Inc. v. Google Inc., 770 F.Supp.2d 666 (S.D.N.Y. 2011) in which Google had successfully avoided legal liabilities by clearly defining well in advance a business purpose for use of Big Data. In this case, Google was sued for violating copyright by creating a copy of authors books in the form of e-books (using optical character recognition technology) and then responding to the users queries/searches on the basis of matching keywords and thereby increasing Google books sales. The Honourable court favoured Google by stating that Google's usage does not fall into the category of 'massive copyright infringement' (as claimed by the plaintiff) or adversely impacts the rights of copyright holder, as it had followed the required due diligence and also is eligible to fall under the 'fair use' category. Google through its security measures didn't allow the users to have a complete view of the books but only snippet views, thus, giving respectful consideration to the author's rights and creativity. Not only that, the court held that Google had incorporated better research tool (in form of data mining), ease of access, efficient mechanism for identifying and locating books and quick search results for end users. On the whole, we may say that an organisation can mitigate the risk of litigation arising out of using Big Data with well-defined business purpose (inclusive of transparency to users regarding usage of data collected, protections from any competing commercial interests that may arise, and above all serving the public good).

Big Data in Healthcare Industry: Mckinsey Global Institute estimated that healthcare analytics will generate more than $300 billion in business value per year. Big Data can make significant changes and developments in reshaping public health. Google published a paper in the scientific journal 'Nature' estimating the likelihood of rapid spread of the H1N1 virus, just few weeks before the virus actually hit (Mayer-Schönberger & Cukier, 2013). As pointed out by Bill Hamilton (2012), "If a group of patients is discussing quality of care about a provider, there will likely never be 100% consensus. Patient experiences will be different, and there will be biases based on accidents, misunderstandings and other factors. The challenge will be to create useful information out of this collection of data to provide information such as provider ratings and improvement guidance" (Hamilton, 2012).

For more examples and a detailed discussion on legal aspects we refer the reader to *Big Data: A Business and Legal Guide* by Kalyvas & Overly (2014). Criticism of real life Big Data application has also surfaced (Lazer et al., 2014). However, we observe that Big Data applications and development are at a nascent stage and we envisage that over a period of time the technology, platforms, and applications will mature proving their utility.

4. Big Data Opportunities

Big Data being voluminous allows us to explore new information avenues with better granularity and without the risk of blurriness. Immense volumes of data lie around us needing to be collected and processed to extract value. One of the major benefits of creating and using Big Data is that it highlights and spots such points of concern which otherwise may be entirely undetectable when using sample data (Mayer-Schönberger & Cukier, 2013).

Governments and PSUs: Governments are increasingly adopting digital technologies. USA.gov and 'Digital India' are notable examples of this trend. The 2012 presidential election campaign in U.S has seen one of the remarkable uses of Big Data for better decision making. President Barack Obama's campaign team conducted Big Data analysis to target voters and identify the most responsive regions for campaigning and then allocating the resources to the destined areas. The winning of Obama and his getting re-elected as president of U.S.A demonstrated and unfolded a new strategic step in making sense of Big Data (Jin, Wah, Cheng & Wang, 2015).

Law Enforcement Agencies: Big Data can be used by Law enforcement agencies in order to analyse voluminious data and impede crime and terrorist attacks. The case of a notorious Chinese serial killer 'Zhou Kehua' is an example of usage and summary analysis of various information obtained from Big Data. The Big Data consisted of video data, photographs and some other related content and on the basis of it Zhou Kehua was tracked, investigated and captured. In this case, Big Data analysis played a decisive role for the law enforcement agency. Big Data may also prove to be applicable in identifying potential criminals (Jin et al., 2015).

Business and Economic Systems: Big Data studies can be applied to raise the economic value and to bring significant societal and scientific impact. Farecast, the air fare predictions website for best buy price helps consumers based on the Big Data analyses on earlier air fare data and thus giving substantial economic benefit to passengers (Mayer-Schönberger & Cukier, 2013, pp 4-5).

4.1 International Big Data Initiatives

United States of America: In September 1993, the 'Information Highway' program was launched in USA. Similarly, in March 2012 the 'Big Data Research and Development Initiative' was launched. The project envisions to improve and facilitate use of Big Data by extracting valuable information insights for better development. It primarily focuses on healthcare, emergency response and disaster recovery, cybersecurity, education and employability, transportation and energy sector (Jin et al., 2015).

United Kingdom: COSMOS (see *What is COSMOS?*) aims to be an open platform for social data analysis that can harvest, archive, analyze and visualize social media streams. In due course, the platform is expected to link to other social data and is currently linked to the UK Police API, harvesting crime statistics. Collectively, the European Union has also started partnering through the program 'Horizon 2020'.

Japan: Aspires to be the World's Most Advanced IT Nation by year 2020. 'The Integrated ICT Strategy for 2020' has already been launched with a mission to develop Japan as a leader of Information Technology with Big Data at its centre stage. The aforesaid IT strategy focuses at implanting the highest level of standards in Big Data technology and IT infrastructure (see *Declaration to be the World's Most Advanced IT Nation*).

Germany: The German Government has announced a Big Data research initiative namely 'production intelligence'. The aim is to perform real time analytics on all manufacturing data. This Big Data analysis will help to evaluate, improve, and enhance the manufacturing capacities and processes, to automate, and in effective decision making, and to achieve optimal manufacturing scenarios (see German government announces "Production Intelligence": funding for Jedox's Big Data project).

Australia: The Australian Public Service ICT Strategy 2012-2015 aims to use Big Data for better service delivery, efficient and effective mechanisms for e-governance, preserve national information assets, improve health service offerings and better emergency response mechanisms. Australian government uses Patient Admission Prediction Tool (PAPT) (in collaboration with Australian e-Health Research Centre Queensland Health, Griffith University and Queensland University of Technology) software for Big Data analytics in health industry. PAPT aims to achieve predictions for number of patients that hospital may expect in the near days, emergency cases, hospital staff's case(s) handling capacities, available and required labor pool as and when need arises, and balanced workload. These predictions can achieve timely service delivery, better disaster resilience and a far better quality care offering.

United Nations: UN recently launched a project 'Global Pulse: Harnessing Big Data for Development and Humanitarian Action'. Global Pulse is intended to ascertain and predict the societal issues like unemployment, disease outbreaks, and likewise. It aspires to achieve proactive approach in handling alarming events arising out of humanitarian grounds. It works for creating awareness and development in regard to Big Data opportunities and its value addition for society (http://www.unglobalpulse.org/).

5. Big Data: The Road ahead in India

Substantial Big Data is being generated (and stored) by Government departments in India already. Department of Science and Technology, GoI has announced plans to take Big Data research forward in the Indian context, including financial support for teams taking up such projects (http://dst.gov.in/scientific-programme/bigdatainitiative.html). However, continuous effort shall be needed for a long period of time before some success stories of big data studies and their results are visible.

More efforts to tap the potential of big data analytics, especially in the social welfare sphere, are needed. The bottlenecks needed to be overcome are: (i) not much (big) data is being collected and stored in India (leaving a few segments, such as, scientific community with space and weather data), (ii) accessibility to expensive platforms (hardware and software needed, though open source can be deployed) is limited, (iii) efforts are needed in the direction of preparing policies and legal frameworks covering issues such as responsibility for collection, storage, and preservation, protection from illegal use, ownership of the data and (extent of) freedom to share with others, etc.

Prime Minister's farmer soil health card is an initiative which could provide extremely valuable data in future contributing to the nation's food security. Similar schemes are also needed for our other national natural resources, such as: (i) monitoring (underground and surface) water availability, usage, and its preservation in India, (ii) rainfall harvesting activities and potential, (iii) land and its (current and possible) usage across the country, (iv) forest areas monitoring, (v) wildlife data, (vi) air quality data from cities (recent media reports indicate Delhi as the most polluted city in the World, now since a few years), (vii) wind farming potential, and so on.

On the human development index front, monitoring of diet and health data (including disease spread and control, vaccinations, etc), levels and adequacy of nutrition intakes in society, education (availability and usage), transport (needs, trends, and consumption), electricity (generation, distribution, shortages, losses, etc), provide scope for exploiting big data applications for big gains.

The central and state Governments in India stand to gain a lot by joint planning, collection, sharing, and analysis of big data to develop appropriate talent development plans for future, planned farming to avoid over and under production in a season leading to excess or shortage, and similar other schemes.

Effort is needed to tap the potential in big data starting with: to identify, support (such as through fully funded academic scholarships), develop, and employ special talent to tap the potential of big data. Simultaneously, to set up big data analytics centres with necessary infrastructure, and accessibility to the scientific and academic community, supported by series of funding for incorporating future developments in these technologies and ensuring their immediate availability to the talent pool for productive deployment.

As the currently available big data talent pool is small, it may not yield the critical mass necessary to push a series of big data projects ahead rapidly. Hiring talent from other countries, if available, may also be an expensive alternative. Hence, progressively expanding Government support appears to be the good road ahead. The first among these steps could be identifying and supporting scholars and students in acquiring the necessary knowledge and skills.

Being a very large nation with equally large variety of data generated (and additionally, challenges of different languages, notations, formats used etc), it is desirable for India to develop and follow a schema designed and suggested by scholars, researchers, various data users, and Governments collectively. Though, Big Data technology is designed to address analysis of humongous volumes and variety of data, bringing even a partial order to its collection and storage by pursuing an established policy can ease analytical loads (by somewhat reducing data's dimensionality), help in ensuring that it meets legal frameworks and manage changes brought to it from time to time.

Organisations who share their data with public should do so with an explanatory data dictionary, including Government organisations. Development of a data storage standard (such as Data Collection and Storage ISI standard) is also recommended, though both, its compliance by all involved agencies and retrofitting old data can become extremely challenging (if not impossible) undertakings. However, this is expected to be helpful in better collection of data in future, including its integration among several parties for meaningful analysis and insights.

Additionally, a central repository for organisations to contribute semi-structured and unstructured data (with its description, even if brief) should be sponsored by central and state Governments (a data-history-pedia). Such data may prove to be useful for hitherto unforeseeable analysis.

6. Conclusion

Big Data platforms and technology have crossed the chasm of mere interest. Across the world, scientific, academic, research, business, as well as, government communities are aggressively charting plans and paths to benefit from developments in the big data field. The issues pertaining to policy and existing frameworks developed over the last few years in some advanced countries have been identified and critiqued to identify unresolved issues. We anticipate much action in the business and government domains in the years to come, and one such potential arena would lie in big data which spans national boundaries.

Summarizing learnings from recent applications of big data across the world, and identifying the initiatives embraced by governments of prosperous nations, we underscore the huge potential that big data holds and can unfold for the central and state governments in India. We also propose steps that can be taken in India during policy formulation, legal frameworks enactment, infrastructure upgradation and talent pool creation, for the nation to benefit from Big Data platforms and technology.

References

Abowd, John M. and Lane, Julia. (2004). New Approaches to Confidentiality Protection: Synthetic Data, Remote Access and Research Data Centers. English. In: *Privacy in Statistical Databases.* Ed. by Josep Domingo-Ferrer and Vicenc¸ Torra. Vol.3050.Lecture Notes in Computer Science. Springer Berlin Heidelberg, pp.282- 289. ISBN: 978-3-540-22118-0. DOI: 10.1007/978-3-540-25955-8_22. URL: http://dx.doi.org/10.1007/978-3-540-25955-8_22.

Agresti, A. and Finlay, B. (2009). *Statistical Methods for the Social Sciences.* Pearson Education. Pearson Prentice Hall, New Jersey.

Anderson, Chris. (2008). The end of theory: The data deluge makes the scientific method obsolete. In: *Wired* 16.7, pp. 106-129. URL: http://www.wired.com/science/discoveries/magazine/16-07/pb_theory.

Ansolabehere, Stephen and Hersh, Eitan. (2012). Validation: What Big Data Reveal About Survey Misreporting and the Real Electorate. In: *Political Analysis* 20.4, pp. 437-459. DOI: 10.1093/pan/mps023. eprint: http://pan.oxfordjournals.org/content/20/4/437.full.pdf+html. URL: http://pan.oxfordjournals.org/content/20/4/437.abstract.

Australian Public Service Better Practice Guide for Big Data. (2015). Retreived from: http://www.finance.gov.au/sites/default/files/APS-Better-Practice-Guide-for-Big-Data.pdf

Barocas, Solon and Nissenbaum, Helen. (2014). Big Data' a End Run around Anonymity and Consent. In: *Privacy, Big Data, and the Public Good.* Cambridge University, 44-75.

Bollier, David and Firestone, Charles M. (2010). *The promise and peril of big data.* Aspen Institute, Communications and Society Program Washington, DC, USA.

boyd, Danah and Crawford, Kate. (2012). Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. In: *Information, Communication & Society* 15.5, pp. 662-679. DOI: 10.1080/1369118X.2012. 678878. eprint: http://www.tandfonline.com/doi/pdf/10.1080/1369118X.2012.678878. URL: http://www.tandfonline.com/doi/abs/10.1080/1369118X.2012.678878.

Bryant, Antony and Raja, Uzma. (2014). In the realm of Big Data .... In: *First Monday* 19.2. ISSN : 13960466. URL : http://firstmonday.org/ojs/index.php/fm/article/view/4991.

Chen, C. P., & Zhang, C. Y. (2014). Data-intensive applications, challenges, techniques and technologies: A survey on Big Data. *Information Sciences,* 275, 314-347.

Crawford, Kate and Finn, Megan. (2014). The limits of crisis data: analytical and ethical chalenges of using social and mobile data to understand disasters. English. In: *GeoJournal,* pp. 1-12. ISSN : 0343-2521. DOI : 10.1007/s10708-014-9597-z. URL : http://dx.doi.org/10.1007/s10708-014-9597-z.

Daries, J. P., Reich, J., Waldo, J., Young, E. M., Whittinghill, J., Ho, A. D., ... & Chuang, I. (2014). Privacy, Anonymity, and Big Data in the Social Sciences. In: *Commun. ACM* 57.9, pp. 56-63. ISSN: 0001-0782. DOI: 10.1145/2643132. URL: http://doi.acm.org/10.1145/2643132.

Declaration to be the World's Most Advanced IT Nation. (June 14, 2013). Retrieved on 20-Apr-2015 from: http://japan.kantei.go.jp/policy/it/2013/0614_declaration.pdf

Decuyper, A., Rutherford, A., Wadhwa, A., Bauer, J. M., Krings, G., Gutierrez, T., ... & Luengo-Oroz, M. A. (2014). Estimating Food Consumption and Poverty Indices with Mobile Phone Data. Version 1. In: preprint arXiv:1305.3212. arXiv: http://arxiv.org/abs/1412.2595v1 [cs.CY, physics.soc-ph].

Desouza, Kevin C. & Jacob, Benoy. (2014). Big Data in the Public Sector: Lessons for Practitioners and Scholars. In: *Administration & Society.* DOI: 10.1177/ 009539971 4555751.eprint:http://aas.sagepub.com/content/early/2014/11/06/0095399714555751.full.pdf+html. URL: http://aas. sagepub.com/content/early/2014/11/06/0095399714555751. abstract.

Diebold, Francis X. (2012). On the Origin(s) and Development of the Term 'Big Data'. PIER Working Paper No. 12-037. Available at SSRN: http://dx.doi.org/10.2139/ssrn.2152421

German government announces "Production Intelligence": funding for Jedox's Big Data project. (April 13, 2015). Retrieved from: http://www.protext.cz/english/zprava.php?id=22879

Gitelman, Lisa. (2013). "Raw Data" is an Oxymoron. *MIT Press.*

Halevy, A., Norvig, P. and Pereira, Fernando. (2009). The Unreasonable Effectiveness of Data. In: *Intelligent Systems,* IEEE24.2, 8-12. ISSN:1541-1672. DOI: 10.1109/MIS.2009.36.

Hamilton B. (2012). Big Data is the Future of Health Care. *Cognizant 20-20 insights.* pp-5. Retrieved from: www.cognizant.com/.../Big-Data-is-the-Future-of-Healthcare.pdf

Hilbert, Martin. (2013). Big Data for Development: From Information-to-Knowledge Societies. Available at SSRN 2205145. DOI: 10.2139/ssrn.2205145. URL: http://ssrn.com/abstract=2205145.

Hutcheson, Graeme D and Sofroniou, Nick. (1999). *The Multivariate Social Scientist: Introductory Statistics Using Generalized Linear Models.* Statistics Series. SAGE Publications.

Jagadish, H. V., Gehrke, J., Labrinidis, A., Papakonstantinou, Y., Patel, J. M., Ramakrishnan, R., & Shahabi, C. (2014). Big Data and Its Technical Challenges. In: *Commun. ACM* 57.7, pp. 86-94. ISSN : 0001-0782. DOI : 10.1145/2611567. URL : http://doi.acm.org/10.1145/2611567.

Jin, X., Wah, B. W., Cheng, X., & Wang, Y. (2015). Significance and challenges of big data research. *Big Data Research* (In press).

Kalyvas, J. R., & Overly, M. R. (2014). *Big Data: A Business and Legal Guide.* CRC Press.

Kambatla, K., Kollias, G., Kumar, V., & Grama, A. (2014). Trends in big data analytics. *Journal of Parallel and Distributed Computing,* 74(7), 2561-2573.

Khoury, Muin J. and Ioannidis, John P. A. (2014). Big data meets public health. In: Science 346.6213, pp. 1054-1055. DOI : 10.1126/science.aaa2709. eprint:http://www.sciencemag.org/content/346/6213/1054.full.pdf. URL : http://www.sciencemag.org/content/346/6213/1054.short.

Kramer, Adam D. I., Guillory, Jamie E., & Hancock, Jeffrey T. (2014). Experimental evidence of massive-scale emotional contagion through social networks. In: *Proceedings of the National Academy of Sciences* 111.24, pp. 8788-8790. DOI: 10.1073/pnas.1320040111. eprint: http://www.pnas.org/content/111/24/8788.full.pdf+html. URL: http://www.pnas.org/content/ 111/24/8788.abstract.

Lagoze, Carl. (2014). Big Data, data integrity, and the fracturing of the control zone.In: *Big Data and Society* 1.2. DOI:10.1177/2053951714558281.eprint: http://bds.sagepub.com/content/1/2/2053951714558281.full.pdf+html.URL:http://bds.sagepub.com/content/1/2/2053951714558281. abstract.

Lane, Julia and Schur, Claudia. (2009). Balancing Access to Data And Privacy. A review of the issues and approaches for the future. Working Paper Series of the German Council for Social and Economic Data 113. German Council for Social and Economic Data (RatSWD). URL: http://ideas.repec.org/p/rsw/rswwps/ rswwps113.html.

Lane, Julia and Schur, Claudia. (2010). Balancing Access to Health Data and Privacy: A Review of the Issues and Approaches for the Future. In: *Health Services Research* 45.5p2, pp. 1456-1467. ISSN: 1475-6773. DOI: 10.1111/j.14756773.2010.01141.x. URL: http://dx.doi.org/10.1111/j.14756773.2010.01141.x.

Lane, Julia and Stodden, Victoria. (2013). What? Me Worry? What to Do About Privacy. *Big Data, and Statistical Research.* URL:http://magazine.amstat.org/ blog/2013/12/01/bigdatastatresearch/.

Lazer, D. M., Kennedy, R., King, G., & Vespignani, A. (2014). The Parable of Google Flu: Traps in Big Data Analysis. In: Science 343.6176,pp.1203-1205. DOI: 10.1126/science.1248506.eprint: http://www.sciencemag.org/content/343/6176/1203.full.pdf. URL: http://www.sciencemag.org/content/343/6176/1203.short.

Manovich, Lev. (2012). Trending: The Promises and the Challenges of Big Social Data. Debates in the Digital Humanities. In: *Debates in the Digital Humanities.* Ed. by M K Gold. University of Minnesota Press. Chap. 27, pp. 460-475.

Mayer-Schönberger, V., & Cukier, K. (2013). *Big Data: A revolution that will transform how we live, work, and think.* John Murray, London.

Morton, J., Runciman, B. & Gordon, K. (2014). *Big Data: Opportunities and challenges.* BCS Learning & Development Limited, Great Britain.

Mundie, Craig. (2014). Privacy Pragmatism; Focus on Data Use, Not Data Collection. In: *Foreign Aff.* 93, p. 28.

Nissenbaum, Helen. (2010). *Privacy in Context: technology, policy, and the integrity of social life.* Stanford, Calif: Stanford Law Books. ISBN: 978-0804752367.

Online Security and Privacy at Xoom. (n.d.). Retrieved on 20-Apr-2015 from: https://www.xoom.com/india/security-center

Podesta, J., Pritzker, P., Moniz, E.J., Holdren, J., & Zients, J. (2014). Big Data: Seizing Opportunities, Preserving Values. *Tech. rep. Executive Office of the President.* URL: http://www.whitehouse.gov/ sites/default/files/docs/big_data_privacy_report_may_1_ 2014.pdf.

Prydz, Espen Beer. (2014). Knowing in time: How technology innovations in statistical data collection can make a difference in development. *PARIS 21.* URL: http://www.paris21.org/sites/default/files/PARIS21-DiscussionPaper2Knowing.pdf.

Ruths, Derek and Jürgen Pfeffer. (2014). Social media for large studies of behavior.In: *Science* 346.6213, pp.1063-1064. DOI:10.1126/science.346.6213. 1063.eprint: http://www.sciencemag.org/content/346/6213/1063. full.pdf. URL: http://www.sciencemag.org/content/346/6213/1063.short.

Salsburg, David. (2001). *The Lady Tasting Tea: How Statistics Revolutionized Science in the Twentieth Century.* Macmillan.

Schintler, Laurie A. and Rajendra Kulkarni. (2014). Big Data for Policy Analysis: The Good, The Bad, and The Ugly. In: *Review of Policy Research* 31.4, pp. 343- 348. ISSN: 1541-1338. DOI: 10.1111/ropr.12079. URL: http://dx.doi. org/10.1111/ropr.12079.

Taylor, Linnet and Schroeder, Ralph. (2014). Is bigger better? The emergence of big data as a tool for international development policy. English. In: *Geo Journal,* pp. 1-16. ISSN: 0343-2521. DOI: 10.1007/s10708-014-9603-5. URL: http://dx.doi.org/10.1007/s10708-014-9603-5.

Taylor, Linnet, Schroeder, Ralph & Meyer, Eric. (2014). Emerging practices and perspectives on Big Data analysis in economics: Bigger and better or more of the same? In: *Big Data & Society* 1.2. DOI : 10.1177/2053951714536877. eprint: http://bds.sagepub.com/content/1/2/2053951714536877.full.pdf+html. URL: http://bds.sagepub.com/content/1/2/2053951714536877.abstract.

Tufekci, Zeynep. (2013). Big Data: Pitfalls, Methods and Concepts for an Emergent Field. In: *SSRN* (March 7 2013). DOI: 10.2139/ssrn.2229952. URL: http://ssrn.com/abstract=2229952.

United Nations Global Pulse White Paper. (2012). *Big Data for Development: Opportunities and Challenges.* Retrived on 15-Apr-2015 from: http://www.unglobalpulse.org/sites/default/files/BigDataforDevelopment-UNGlobalPulseJune2012.pdf

Velleman, P F. (1997). The Philosophical Past and the Digital Future of Data Analysis: 375 Years of Philosophical Guidance for Software Design on the occasion of John W. Tukey's 80th Birthday. In: *The practice of Data Analysis: Essays in Honor of John W. Tukey.* Princeton Legacy Library, pp. 317-337.

Vis, Farida. (2013). A critical reflection on Big Data: Considering APIs, researchers and tools as data makers. In: First Monday 18.10. ISSN : 13960466.

Washington, Anne L. (2014). Government Information Policy in the Era of Big Data. In: *Review of Policy Research* 31.4, pp. 319-325. DOI: 10.1111/ropr. 12081.

What is COSMOS? (n.d.). Retrieved on 12-Apr-2015 from http://www.cs.cf.ac.uk/cosmos/

Zikopoulos, P., Eaton, C., Deeros, D., Deutsch, T. & Lapis, G. (2015). *Understanding big data: Analytics for enterprise class hadoop and streaming data.* McGraw-Hill Osborne Media.

Zwitter, Andrej. (2014). Big Data ethics. In: *Big Data & Society* 1.2. DOI : 10.1177 / 2053951714559253. eprint: http://bds.sagepub.com/cgi/reprint/1/2/2053951714559253. URL : http://bds.sagepub.com/cgi/content/abstract/1/2/2053951714559253.

Madhukar Dayal has served as a Gazetted officer in Indian Railways (Indian Railways Service of Mechanical Engineers) for over twenty years. His experience spans Railway operations, information technology projects, R&D, as well as, teaching. He has travelled widely across the country conducting seminars, delivering speeches, advising IR in IT projects, and in applications of emerging technologies in Indian Railways. He was also the editor of IRIMEE, Jamalpur newsletter and conference proceedings. He earned his Fellow (Computers & Information Systems) degree from IIM Ahmedabad. He has taught C, C++ (OOP), Industrial Management, computer hardware, software, and networking courses at the graduate level and in Management Development Programs. Currently he teaches Spreadsheet Modelling, Information Systems for Managers, and Modern Computing Applications for Businesses in the post-graduate programs, and DBMS & OLTP in the Fellow (PhD) program. His research interests include high performance compute cluster algorithms and business and government applications of technology.

Sachin Garg is a Ph.D. student at George Mason University's School of Policy, Government & International Affairs. He is currently researching how Big Data can be used to analyse and inform in Public Policy, and how policies impact the generation, collection and dissemination of such Data. Prior to joining the doctoral programme, Sachin was working as a Software Architect in Yahoo!. He has extensive and varied experience with Open Source Software, especially Linux. He holds a Masters in Computer Science from the University of Allahabad, India.

Rubaina Shrivastava is an Academic Associate (Information Systems) at Indian Institute of Management, Indore (IIMI). Earlier, she worked as a Research Fellow (Information Security) at National Law Institute University (NLIU, Bhopal). She has received Post Graduate from NLIU, Bhopal in the stream of Cyber Laws and Information Security. Also, she holds two globally respected certification viz. CCNA and RHCE. She is a compassionate enthusiast for addressing cybersecurity techno-legal challenges and to inform others about the good than can be done in line.