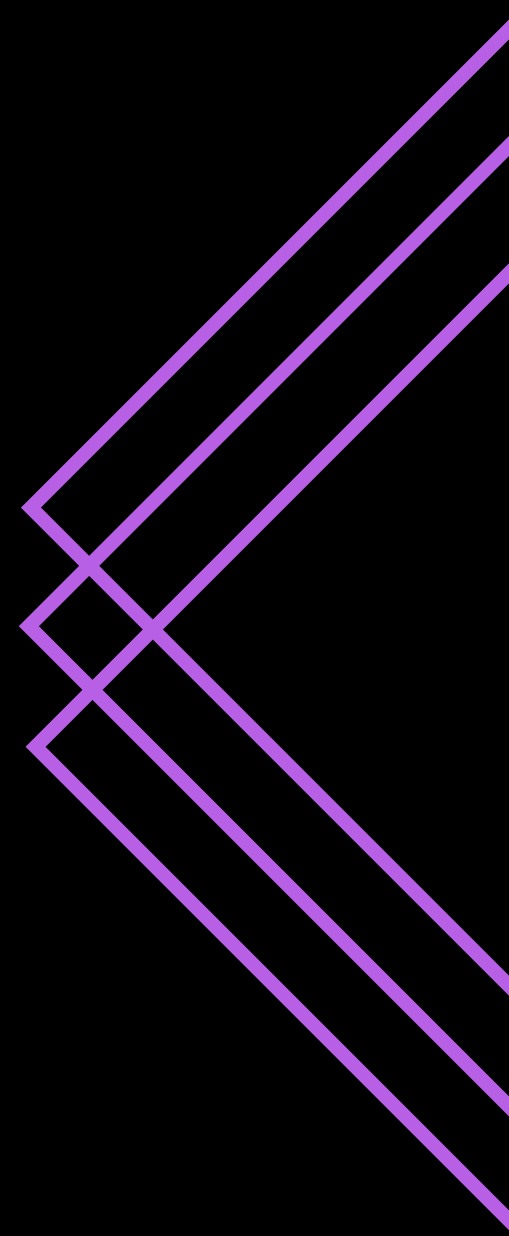July 2018

BIG DATA AND DATA PRIVACY

FOUNDATION FOR MEDIA ALTERNATIVES

Jamael Jacob
Maristela Miranda
Jessamine Pacis
Cecilia Soria

About FMA

The Foundation for Media Alternatives (FMA) is a non-profit service institution whose mission is to assist citizens and communities – especially civil society organizations (CSOs) and other development stakeholders – in their strategic and appropriate use of the various information and communications media for democratization and popular empowerment.

Since its formation in 1987, FMA has sought to enhance the popularization and social marketing of development-oriented issues and campaigns through media-related interventions, social communication projects and cultural work. In 1996, FMA streamlined its programs and services in both traditional and new media, with a major focus on information and communications technologies (ICTs), to enable communities to assert their communication rights and defend their rights to information and access to knowledge, towards progressive social transformation.

FMA seeks to develop programs and projects that strategically address the questions of access to and equity of disadvantaged sectors in the area of information and communications – and in locating the so-called digital divide within existing socio-political divides, including gender. These involve:

- Promoting equitable partnerships for innovating connectivity and community access alternatives to assert the agenda of disadvantaged communities;
- Facilitating capacity-building sessions for CSOs in the area of ICT literacy, ICT management, online collaboration or advocacy, and secure online communications;
- Helping CSOs manage development content through appropriate tools and technologies towards building vibrant online communities and knowledge networks; and
- Enhancing multi-stakeholder consensus-building on strategic information and communication agendas, via action-oriented research, constituency-building and public advocacy.
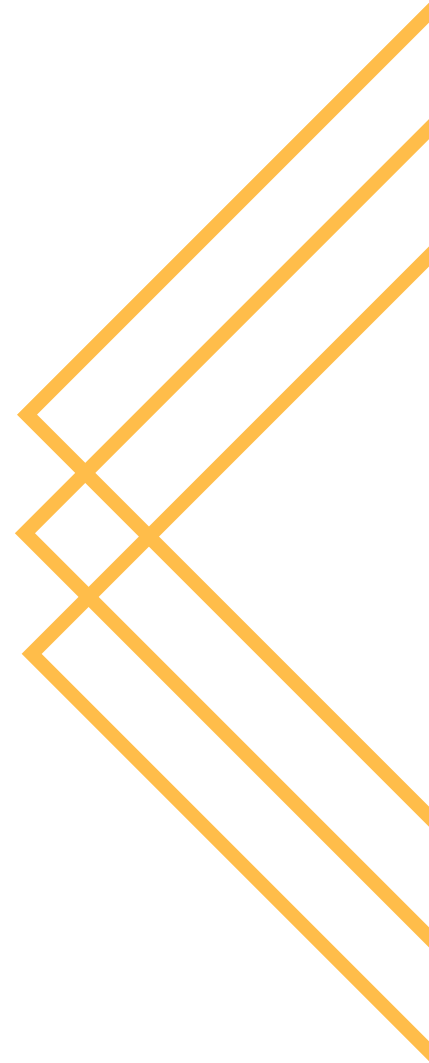
Introduction

What is big data? Given its name, does one define it simply in terms of size or quantity? If that's the case, much of the excitement that surrounds it is overblown. History, after all, is teeming with evidence of humans dealing with generous amounts of data. That also means the concept is a dated one, merely resuscitated from relative obscurity.

No. Big data goes beyond that—though volume remains a major distinguishing mark.The consensus seems to be that it is characterized by the three"Vs": volume (because of the massive datasets involved), velocity (because of near-real-time accumulation, generation, or processing), and variety (because of the many different sources it is culled from)[1].[2]

This is how one gets to say that big data is at play when a company is able to cut the inventory flow in its national distribution network (saving billion dollars)by mining product, location, and transport data, including those sourced from scanners and sensors, using special software, in order to identify potential time-saving and cost-cutting opportunities.[3] It's why people say it is

right in the thick of things when a search engine company is able to predict flu outbreaks by processing the 50 million most common search terms and comparing them with existing government data on said outbreaks over a given period, using software and 450 million different mathematical models.[4] And it's not, by any means, limited to statistics and aggregate data either.[5] For instance, one big data program involves analyzing the data streams from medical devices monitoring patients in an intensive care unit (which typically generates 160,000 data points per second) in order to spot early warning signals of a patient's worsening condition.[6]

More than anything else, those 3 characteristics define what big data means to most people today. There is no universally recognized definition for the term, even though that hasn't stopped people from developing their own. In the U.S., for instance, a paper produced under the Obama administration was dedicated to big data. The paper referred to it as consisting of "...large, diverse, complex, longitudinal, and/or distributed datasets generated from

---

[1] Information Commissioner's Office. *Big data, artificial intelligence, machine learning and data protection,* version 2.2 (p.6) https://ico.org.uk/media/for-organisations/documents/2013559/big-data-ai-ml-and-data-protection.pdf

[2] Executive Office of the President (2014 May). *Big data: seizing opportunities, preserving values.* The White House (p.4) https://obamawhitehouse.archives.gov/sites/default/files/docs/big_data_privacy_report_may_1_2014.pdf

[3] Lohr, S. (2015). *Data-ism.* London: Oneworld Publications.

[4] Mayer—Schonberger, V. & Cukier, K. (2013). *Big data: A revolution that will transform how we live, work and think.* London: John Murray..

[5] Executive Office of the President (2014 May). Op. cit.

[6] Lohr, S. (2015). Op. cit.
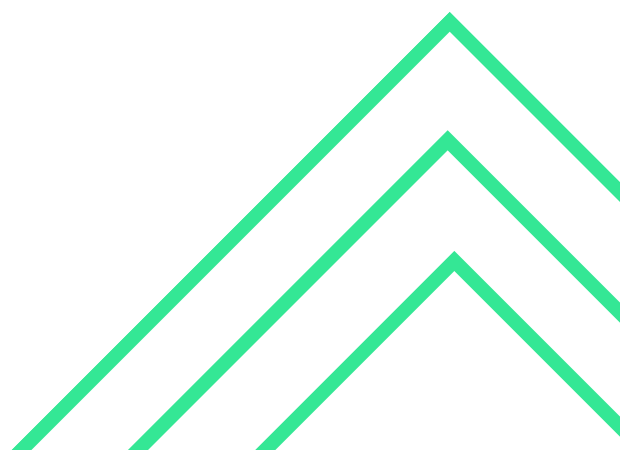
instruments, sensors, Internet transactions, email, video, click streams, and/or other digital sources available today and in the future".[7] Authors Mayer-Schonberger and Cukier, meanwhile, describe it as "the ability of society to harness information in novel ways to produce useful insights and services of significant value.[8]

Widespread fascination with the subject may be traced, in part, to studies that describe big data as critical in improving economic productivity. In 2011, for instance, the McKinsey Global Institute noted that the world is in the "cusp of a tremendous wave of innovation, productivity, and growth, as well as new modes of competition and value capture", all thanks to this newest technological marvel. For this reason, it went on, consumers and businesses alike should take full advantage of its potential.[9] A survey conducted around the same time among global business leaders also showed that 90% of the respondents considered data as a new factor of production whose importance to businesses is no different from physical assets, labor, or capital.[10] Author Steve Lohr, perhaps describes the optimism best when he says, "big data technology is ushering in a revolution in measurement that promises to be the basis for the next wave of efficiency and innovation across the economy."[11]

Today, many accounts appear to validate all those predictions. Some confirm that there have been substantial improvements in productivity among companies that engage in data-driven innovations. In the EU, it was predicted that there will be a 1.9% increase in Gross Domestic Product (GDP) if the region applies big data analytics (BDA) from 2014 to 2020. The revenue for big data and business analytics has also been forecast to grow from almost US$122 billion in 2015 to more than US$187 billion in 2019—more than a 50% increase.[12] Even the United Nations (UN) has joined the bandwagon, having proclaimed big data as a means to achieve the Sustainable Development Goals.[13]

---

[7] National Science Foundation (2012). Solicitation 12-499: *Core techniques and technologies for advancing big data science &engineering (BIGDATA).*
https://www.nsf.gov/pubs/2012/nsf12499/nsf12499.htm
[8] Mayer—Schonberger, V. &Cukier, K. (2013). Op. cit.
[9] Manyika, J, et. al (2011 May). Big data: The next frontier for innovation, competition and productivity. *McKinsey Global Institute*, (p. 2)
https://www.mckinsey.com/~/media/McKinsey/Business%20Functions/McKinsey%20Digital/Our%20Insights/Big%20data%20The%20next%20frontier%20for%20innovation/MGI_big_data_exec_summary.ashx
[10] European Parliamentary Research Service (2016 September). *Big data and data analytics: The potential for innovation and growth.* (p. 3)
http://www.europarl.europa.eu/RegData/etudes/BRIE/2016/589801/EPRS_BRI(2016)589801_EN.pdf

[11] Lohr, S. (2015). Op. cit.
[12] Davis, J. (2016 May 24). Big data, analytics sales will reach $187 billion by 2019. *Information Week.*
https://www.informationweek.com/big-data/big-data-analytics/big-data-analytics-sales-will-reach-$187-billion-by-2019/d/d-id/1325631
[13] http://www.un.org/en/sections/issues-depth/big-data-sustainable-development/index.html

Working with Big Data

In 2016, a research and advisory firm based in the U.S. calculated that around forty percent (40%) of firms were using big data technology.[14] Last year, the 2017 Big Data Analytics Market Study revealed that fifty-three (53%) of the respondent-organizations—Asia-Pacific, Europe, Middle East, Africa and North America)—now use BDA. Among the early adopters were the telecommunications and financial services sectors,[15] although other industries like media and entertainment, healthcare, education, manufacturing, insurance, retail, transportation, utilities, and even government, are now in the mix too.[16]

Meanwhile, a growing list of technologies have been developed to handle the processing demands of BDA, owing to the inability of traditional data techniques and platforms to rise up to the challenge. The more prominent ones include:

- *Predictive analytics* — analyzes historical data[17] to develop predictive models to improve business performance or mitigate risks[18] (e.g., fraud detection, credit scoring, marketing and finance).

- *Prescriptive analytics*— advises users on what to do to achieve a desired result,[19] based on "any combination of analytics, math, experiments, simulation, and/or artificial intelligence to improve the effectiveness of decisions made by humans or by decision logic embedded in applications".[20]

- *Stream analytics* — allows for the real-time filtration, aggregation, and analysis of data in motion, in any format.[21]

- *Data Lakes*— huge depositories that collect data from different sources and store these data in their raw forms.[22]

- *Search and knowledge discovery tools and technologies* — allow for self-service extraction of structured and unstructured data from large depositories of multiple sources.[23]

- *Data virtualization*— integrates data from varying sources, locations, and formats to create a unified database, without having to replicate any data.[24]

[14] Harvey, C. (2018 January). Big data trends. https://www.datamation.com/big-data/big-data-trends.html

[15] Columbus, L. (2017, 24 December). 53% of companies are adopting big data analytics. *Forbes.* https://www.forbes.com/sites/louiscolumbus/2017/12/24/53-of-companies-are-adopting-big-data-analytics/#7fe0fff39a19

[16] Gaitho, M. (2017, 20 December). How applications of big data drive industries. *Simplilearn.* https://www.simplilearn.com/big-data-applications-in-industries-article

[17] Harvey, C. (2017, 2 August). Big data technologies.https://www.datamation.com/big-data/big-data-technologies.html

[18] Press, G. (2016, 14 March). Top 10 hot big data technologies. *Forbes.* https://www.forbes.com/sites/gilpress/2016/03/14/top-10-hot-big-data-technologies/#62f7b52b65d7

[19] Harvey, C. (2017, 2 August). Op. cit.

[20] Gualtieri, M. (2017, 20 February). What exactly the heck are prescriptive analytics? *Forrester.* https://go.forrester.com/blogs/17-02-20-what_exactly_the_heck_are_prescriptive_analytics/

[21] https://www.techopedia.com/definition/31746/streaming-analytics
See also: Press, G. (2016, 14 March). Op. cit.

[22] Harvey, C. (2017, 2 August). Op. cit.

[23] Press, G. (2016, 14 March). Op. cit.

[24] https://www.denodo.com/en/data-virtualization/overview
See also: Press, G. (2016, 14 March). Op. cit.

- *Data integration* — combines technical and business processes to make sense and value of data combined from different sources.[25]

- *Blockchain technology* — stores identical blocks of information in no single location, and which cannot be controlled by a single entity.[26] Because of its security feature, it is popular among those in the finance, insurance, healthcare, and retail industries.[27]

In utilizing these tools, organizations have yielded impressive results, albeit with varying degrees of success. For instance, Germany's football federation was reported to have analyzed recorded videos of the national football team's previous games to improve performance, eventually helping them to win the 2014 FIFA World Cup.[28] Dr. Pepper Snapple Group's sales agents now use iPads instead of voluminous binders full of customer and sales data to make intelligent sales decisions.[29] Bristol-Myers Squibb reduced the number of its clinical trial subjects from sixty (60) to forty (40), and shorten the length of the study by more than a year using the cloud and big data.[30] And then, there's the availability of personalized medicine at significantly lower costs.[31]

[25] https://www.ibm.com/analytics/data-integration
[26] https://blockgeeks.com/guides/what-is-blockchain-technology/
[27] Harvey, C. (2017, 2 August). Op. cit.
[28] Barton, S. (15 September). Top 5 analytics success stories. *Innovation Enterprise.* https://channels.theinnovationenterprise.com/articles/80-top-5-analytics-success-stories
[29] Boulton, C. (2017, 5 Sept.). 6 data analytics success stories: An inside look. *CIO.* https://www.cio.com/article/3221621/analytics/6-data-analytics-success-stories-an-inside-look.html

[30] Morgan, L. (2015, 27 May). Big data: 6 real-life business cases. *Information Week.* https://www.informationweek.com/software/enterprise-applications/big-data-6-real-life-business-cases/d/d-id/1320590?image_number=2
[31] Panahiazar, M., et. al. (2014 October). Empowering personalized medicine with big data and semantic web technology: Promises, challenges and use cases. *US National Library of Medicine.* https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4333680/

Big Data Use in the Philippines

In the domestic context, the use of big data by government and the private sector is also growing steadily. Some notable examples include:

- *IBM in Davao City.* In 2012, IBM partnered with the local government unit (LGU) of Davao City to establish its Intelligent Operations Center (IOC). Enjoying a PhP 128 M fund support from the LGU, the IOC supposedly integrated all services and data from the LGU's different offices to facilitate better government response to crime, terrorism, traffic, and emergency responses and other public safety issues.[32]

  The IOC continues to operate today. It is equipped with multi-channel unified communication, video analytics software, and global positioning system (GPS) location tracking. The video analytics software gets input from the closed-circuit television (CCTV) cameras installed all around the city. The IOC also helps coordinate emergency services such as police, fire, anti-terrorism task force, and the K9 urban search and rescue services.

- *Measuring the Information Society Project.* The Department of Information and Communications Technology (DICT), assisted by the Philippine Statistics Authority,[33] partnered with the International Telecommunication Union (ITU) for the latter's "Measuring the Information Society" big data research project. The project, in general, aims to produce "new and existing ICT indicators to enhance data availability, benchmarks, and methodologies to measure the information society," using big data.[34] It hopes to do this using a number of data points sourced from telecommunications companies. The resulting indicators could then be used for policy and investment decisions. The Philippines joined five other countries as pilot area for the study. Smart Communications, Inc. (SMART) and Globe Telecom, Inc.—the country's only telecommunication firms—were also onboard. The research was completed in the third quarter of 2017.

- *Open Traffic Initiative.* This big data project of the Philippine government is being administered through the Department of Transportation (DOTr), in partnership with the World Bank and Grab Philippines.[35]

---

[32] MindaNews (2013, 7 June). Davao goes high-tech on public safety, security efforts. *MindaNews.* http://www.mindanews.com/top-stories/2013/06/davao-goes-high-tech-on-public-safety-security-efforts/

[33] From the presentation of Ramon Ibrahim titled "Big Data Analytics: Applications in Policy, Development, and Governance", at the Big Data conference, Asian Institute of Management, Makati, Philippines, 26 July 2017.

[34] From the presentation of Margus Tiru, Project Coordinator and ITU Consultant, "Big Data for Measuring the Information Society"

[35] Schnabel, C. (2016, 5 April). Grab, World Bank launch big data project to ease PH traffic. *Rappler.* https://www.rappler.com/business/industries/208-infrastructure/128327-open-traffic-initiative-grab-worldbank

Launched in 2016, it allows traffic management agencies and city planners to access real-time traffic data sourced from Grab's platform (i.e., GPS data submitted by Grab-registered vehicles). From this same dataset, the DOTr also gathers road incident reports, which are then used to develop traffic management interventions, and to improve the government's emergency response capability. For its pilot phase, the following studies are expected to be undertaken: (1) peak-hour analysis along key corridors; (2) travel time reliability; (3) corridor vulnerability to inclement weather or traffic incidents; and (4) identification of road incident blackspots.[36] Areas initially covered are Metro Manila and Cebu City. The World Bank and Grab plan to implement similar projects in other key Southeast Asian cities.

- *Combatting a Dengue Outbreak.* In 2015, dengue fever cases were rising all across the province of Pangasinan. Wilson Chua, a Filipino big data analyst based in Singapore, got in touch with the Department of Health (DOH) and requested data on dengue fever cases in the province. Using the data, he was able to pinpoint several barangays in Dagupan City as hotspots for the disease.[37]

More importantly, he was also able to isolate and identify ground-zero of the outbreak. Working together with city government personnel, he was able to determine the breeding areas for the Dengue fever-carrying mosquitos. With the help of the Bureau of Fisheries and Aquatic Resources, he was also able to help provide the solution: the release of fishes that feed on mosquito eggs in the affected areas.

---

[36] World Bank (2016, 5 April). Philippines: Real-time data can improve traffic management in major cities. http://www.worldbank.org/en/news/press-release/2016/04/05/philippines-real-time-data-can-improve-traffic-management-in

[37] The Philippine Statistics Authority has defined barangay as the smallest political unit in the country.

Emerging Concerns

Of course, no technology has ever been developed that does not have its less agreeable side. To date, concerns already abound regarding BDA, particularly when personal data is involved:

- *Perpetuation of Social Biases.*[38] Through BDA, evaluating correlations between data collected from different sources in order to determine an individual's presumably accurate identity (e.g., creditworthiness, likelihood of committing a crime, suitability for a job, preferences and others) is completed much faster.[39] This enhanced profiling ability is aided by the increasing ubiquity of so-called Internet of Things (IoT) devices, and could, if left unchecked, aggravate existing discriminatory practices, especially against vulnerable groups.

- *Gaming the System.* In theory, familiarity with the way BDA works could allow an individual to obtain favorable decisions or yield analyses that benefit him or her by providing or generating falsified data.[40] This could potentially undermine the reliability of BDA as a source of information from which organizations make and implement critical policies and decisions.

- *Privacy Concerns.* With BDA and other related technologies, it is now easier for governments and big businesses to collect and process personal data. This makes information more vulnerable to unauthorized and unrestricted use. Consider the recent reports of China initiating a mandatory social credit system for all its citizens, wherein the social score of each individual is essentially determined by the government and private companies.[41] The score either lets an individual access certain benefits or deprives him or her of others. Managing such a large and complex system would require some form of BDA infrastructure running in the background. For the Philippines, the impending passage of a national ID system law would enable the government to run its own information management system. With the system designed to monitor ID use, connecting to other data processing systems—government or private—is but a small step away. Once consolidation (or at least inter-operability) is achieved, BDA can be put to work for purposes only the State can control or rein in.

[38] Davis, E. (2017, 15 February). The problems of big data and what to do about them. *World Economic Forum.* https://www.weforum.org/agenda/2017/02/big-data-how-we-can-manage-the-risks
See also: https://obamawhitehouse.archives.gov/the-press-office/2014/05/01/fact-sheet-big-data-and-privacy-working-group-review
[39] Brown, M. (2017, 30 August). Math isn't biased, but big data is. *Forbes.* https://www.forbes.com/sites/metabrown/2017/08/30/math-isnt-biased-but-big-data-is/#5bb6e5214d56
See also: Davis, E. (2017, 15 February). Op. cit.
[40] Davis, E. (2017, 15 February). Op. cit.

[41] Ma, A. (2018, 8 April). China has started ranking citizens with a creepy 'social credit' system — here's what you can do wrong, and the embarrassing, demeaning ways they can punish you. *Business Insider.* http://www.businessinsider.com/china-social-credit-system-punishments-and-rewards-explained-2018-4?utm_source=feedly&amp%3Butm_medium=referral

- *Accountability Dilemma.* The way BDA is designed and implemented (with plenty of automated processes involved) is set to make traditional grievance mechanisms obsolete. Challenging results or decisions based on automated processing operations is still uncharted territory. That these processes are run by algorithms that organizations are unwilling to disclose (for proprietary reasons) only makes matters worse.[42]

- *Policies Playing Catch-Up.* A familiar problem when technology and policies intersect is the consistent inability of the latter to keep up with pace of the former. With BDA, many take comfort in the presence of data protection legislation. However, like other laws, they, too, need to be updated regularly to avoid becoming useless or irrelevant.

---

[42] Davis, E. (2017, 15 February). Op. cit.

Looking to the Future

The accumulation and use of data continues to evolve in terms of speed, breadth, and complexity. Things have reached that point when it is now possible to process data at speeds that allow some machines to act and respond like humans. Personalizing products for individual customers is now a reality; the same way predicting behavior no longer wallows in the realm of science fiction. All because of big data and the technologies developed to harness it.

Looking ahead, many offer predictions as to what the next chapters of the big data revolution will look like. Businesses are expected to increase investments in big data technologies and algorithms,[43] with many poised to join the market offering data-as-a-service.[44] There are those who proclaim that machine learning will be at the forefront of this movement,[45] with prescriptive analytics evolving into an integral part of business intelligence software.[46] These would all lead to organizations becoming more efficient and effective,[47] but only if the high demand for capable data protection officers and data scientists are adequately met.[48]

Naturally, addressing privacy and data security concerns will also continue to be a huge challenge.[49]

In the local scene, a pending legislation in Congress is worth noting. Senator Paolo Benigno Aquino IV filed Senate Bill No. 688 or the "Big Data Act," which seeks to establish a State-owned big data center. The institution is tasked with developing a research program, which shall facilitate access by the government and its development partners to various alternative real-time data sources by entering into partnerships with relevant government agencies and private entities.[50] Using the data gathered, it is supposed to come up with new and innovative solutions for government services.[51] The bill offers some form of data protection by explicitly recognizing the need to ensure protection and security of personal data,[52] including the use of anonymization.[53] A counterpart bill (HB 3056) was filed by Rep. Victor Yap at the House of Representatives, although both appear to be languishing at the committee level as of this writing.[54]

---

[43] Ahmed, I. (2017, 15 September). The future of big data: 10 predictions you should be aware of. Smart Data Collective. https://www.smartdatacollective.com/future-big-data-predictions/

[44] Marr, B. (2016, 15 March). 17 predictions about the future of big data everyone should read. Forbes. https://www.forbes.com/sites/bernardmarr/2016/03/15/17-predictions-about-the-future-of-big-data-everyone-should-read/#70583f281a32

[45] Ahmed, I. (2017, 15 September). Op. cit.
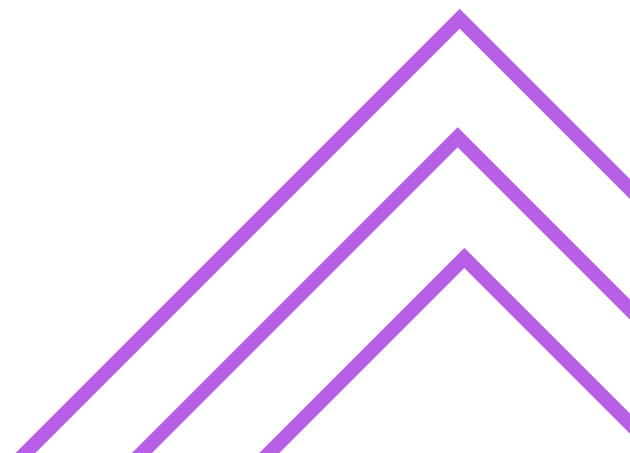
[46] Ibid.

[47] Ibid.

[48] Ibid.

[49] Ibid.

[50] §5.

[51] §10 (d)

[52] §13 (a)

[53] §13 (b)

[54] http://www.congress.gov.ph/legisdocs/basic_17/HB03056.pdf

Irrespective of all these developments, it has to be clear to everyone that big data and all related technologies are but tools meant to help achieve certain ends. Whether they will ultimately work to make society and life, in general, better will all boil down to ethics, the rule of law, and the intentions of their respective users. For now, while the pace of change continues to accelerate and the debate over the merits (and risks) of big data also persists, a few critical points are worth keeping in mind:

- Clear guidelines for the use of big data need to be developed. They must be consistent with the requirements of data protection laws.

- Now, more than ever, transparency is indispensable given how technologies are becoming more sophisticated by the day. People must be informed of the many ways their personal data are being collected and generated, including the rights they may exercise relative to their data.

- Human intervention in decision-making, especially in matters that affect other people's lives, must not be dispensed with. Beyond pop culture fascination with a dystopian future where it is the machines who already reign over people, relinquishing human prerogative in decision-making is a prospect full of unanswered questions that still require further studies and reflection. To move forward without resolving such questions would be nothing less than folly. ◼