BIG DATA AND THE PRACTICE OF LARGE SCALE DATA COLLECTION IN INDONESIA:

An Introduction to Understanding the Actual Challenges of the Protection of the Right to Privacy



ELSAM
PRIVACY INTERNATIONAL

BIG DATA AND THE PRACTICE OF LARGE SCALE DATA COLLECTION IN INDONESIA: An Introduction to Understanding the Actual Challenges of the Protection of the Right to Privacy

Institute for Policy Research and Advocacy (ELSAM)
Privacy International

BIG DATA AND THE PRACTICE OF LARGE SCALE DATA COLLECTION IN INDONESIA: An Introduction to Understanding the Actual Challenges of the Protection of the Right to Privacy

Author:
Wahyudi Djafar

Cover & Layout:
Dodi Sanjaya

First published in Indonesian by::
Institute for Policy Research and Advocacy (ELSAM) and Privacy International, 2017

All ELSAM publications are dedicated to victims of human rights violations in addition to being part of efforts to promote and protect human rights in Indonesia.



Except where otherwise noted, content on this report is licensed under a Creative Commons Attribution 3.0 License. Some rights reserved.

TABLE OF CONTENTS

TABLE OF CONTENTS.....i

A. INTRODUCTION 1

B. SUPPORTING FACTORS AND REASONS FOR LARGE SCALE COLLECTION OF DATA 4

C. FAKTOR PENDUKUNG DAN ALASAN PENGUMPULAN DATA SKALA BESAR 9

D. THE PRACTICE AND RISKS OF LARGE SCALE DATA COLLECTION IN INDONESIA.....22

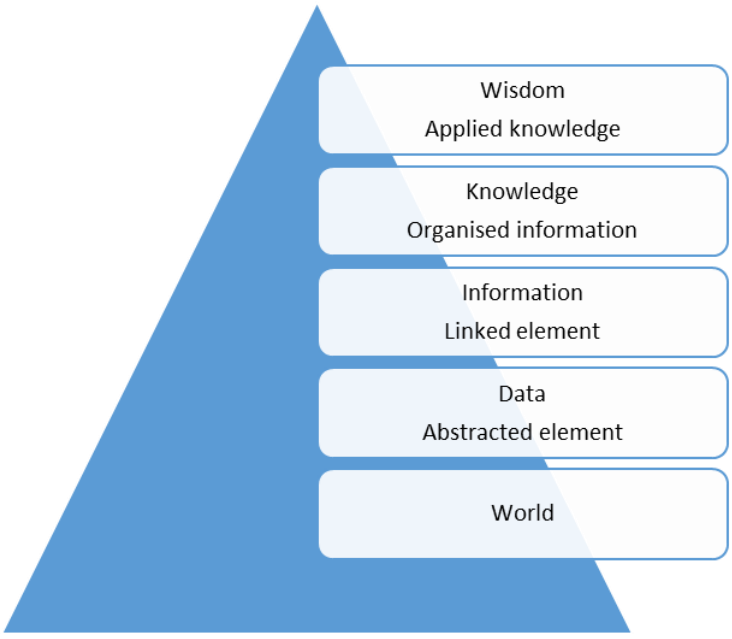
E. THE IMPORTANCE OF PERSONAL DATA PROTECTION REGULATION IN INDONESIA 38

F. CONCLUSION AND RECOMMENDATIONS 40

ORGANISATIONAL PROFILES 44

INTRODUCTION

Data have become a very valuable commodity, many institutions, both government/public and the private sector are vying to conduct data collection and mining on a massive scale. Data are generally generated through and by human activities, about humans, and for the purpose of understanding and ultimately managing and perhaps manipulating-engineering human and societal life. Data are constant sources of knowledge, and their uses have resulted in competitive advantages that are essential for anyone. Armed with the data collected, to be registered, sorted, verified, processed, and analyzed, which then produces a synthesis of information, the various kinds of information produced are then transformed into knowledge, which is then taken into consideration in decision making or policy making. This cycle is often known as the DIKW Pyramid (*data*, *information*, *knowledge*, *and wisdom*), which connects the process of data that is processed into information, analyzed into knowledge, and finally becomes a policy [1]



Source: David McCandless (2010), quoted from Mortimer J. Adler (1986)

1) See: Martin Frické, The Knowledge Pyramid: A Critique of the DIKW Hierarchy, in Journal of Information Science, XX (X) 2007, pp. 1–13.

Referring to the cycle, the concept of *data-driven decision making* or commonly known as evidence based policies is developed, particularly in the development policies of a state, including in Indonesia. In particular, the 2015-2019 National Medium-Term Development Plan (RPJMN) places the development of evidence-based policies to be one of its priorities. Even in the current political contestation, in the implementation of general elections in a number of countries, the *data driven political campaign* model has begun to be developed, so how about data-based political campaigns? This model then resulted in the concept of *voters micro targeting* in political campaigns. A form of political direct marketing by targeting personalized messages to individual voters, by applying predictive modeling techniques to data sets of voters on a massive scale.[2] The development of this technique has encouraged the exploitation of voter data, both sourced from the voter database, as well as other sources, such as personal data of social media users.

Improvement of decision and policy quality is often the reason for collecting various data from various lines of life and everyday human activities. Data become the main basis for crime prevention, with data a candidate can win a political battle, with data security forces can prevent social violence, and with data, a company can increase its business and profits. At this level lies the power of data so that data becomes an entity and commodity that is very valuable. What is the meaning of data? Data are collections of facts, observations or even raw elements that are not regulated or processed in a certain way, and that do not contain clear meaning. Data are basic units that records something, an event, activity or transaction. A piece of data is not significant or valuable by itself, until it is interpreted so that it contains a certain meaning.[3]

Data are collections of facts, observations or even raw elements that are not regulated or processed in a certain way, and that do not contain clear meaning. Data are basic units that records something, an event, activity or transaction.

2) Ira S. Rubinstein, Voter of Privacy in the Age of Big Data, dalam Wisconsin Law Review 861 (2014), pp 861-936.

3) Jean-Louis Monino dan Soraya Sedkaoui, Big Data, Open Data and Data Development, (London: John Wiley & Sons, Inc., 2016).

Data can now be collected on an unprecedented scale, and also from increasingly diverse sources. While in the past large-scale data might only be generated through a census or enumeration mechanism held by state authorities, in its development, large amounts of data can be produced through various human activities. This situation also encourages changes in the way of data processing and manipulation, given the large number of unstructured data sources. Volumes of data are no longer in measured in megabytes or gigabytes, but in terabytes, petabytes, and now even zettabytes. Nowadays, the process of gathering data no longer depends on what data are considered to be useful in the future? Instead, almost all data are collected, the government and the private sector compete to increase their data storage capacity, and increasingly rarely delete data. They find new values in the data, so data are treated as tangible assets. This new era of data management is commonly referred to as Big Data.[4]

The term 'Big Data' itself first appeared in the mid-1990s, when John Mashey, a former chief scientist at Silicon Graphics, presented about the management and analysis of massive datasets. Initially, this term lacked appeal, so it did not develop too much. Only in the next decade, around 2008, various groups, particularly the industry and academia began to use the term 'Big Data'. Currently, Big Data has become a buzzword, used in various circles, whether government, business, academia, or media. [5]

The varied practices of massive data collection that are supported by intensive data systems certainly have a significant influence on the enjoyment and management of human rights, especially one's privacy rights, because the data collected in this process are related to human beings, human behavior, and everyday human activities. Data are very closely related to one's dignity and private life. Although in the process of collecting data, the principle of anonymity is referred to, the risk of leakage or misuse, allowing one to be profiled, is also very large.

If the massive increase in data collection is not

4) Malik, P., 2013. Governing Big Data: Principles and practices. IBM Journal of Research and Develpoment 57, 1:1–1:13.
5) Francis X. Diebold, On the Origin(s) and Development of the Term “Big Data”, PIER Working Paper 12-037 (2012).

carried out within the framework of respect for rights, then inevitably the process and objectives will be used in a way that excludes privacy rights of the community. It is also illustrated from the early practices in massive data collection, which did not pay attention to the aspects of protection of privacy. Therefore, as the practice of collecting big data becomes more intensive, other laws and architectures must be framed in an effort to ensure that this practice also guarantees the protection of the rights to the privacy of someone whose data is collected. This paper is intended to observe in general the practices of large data collection in the regime of Big Data and intensive data systems, especially those that occur in Indonesia. From the practices, the vulnerability of each process is examined, and mechanisms or instruments of protection that must be prepared to minimize the risk of violations and misuse.

B. UNDERSTANDING BIG DATA AND INTENSIVE DATA SYSTEMS

The concept of Big Data does not come with a standard definition that is agreed by all experts. However, it is generally agreed that Big Data is different from traditional business analysis and small-scale data, even in large numbers. As a result, there are often doubts and misunderstandings in understanding Big Data, as a result of the breadth of its definition. Even at some point the definitions contradict each other.[6] From the perspective of computer science, Big Data or data revolution in general is often considered solely as the substance of technological innovation. This means that this concept is limited to the attributes or elements, which consist of newly discovered data and super sophisticated computing powers. This was stated by Manovich (2011), who stated that Big Data generally refer to a large enough data set that require a supercomputer. Currently though, the process can be analyzed through a desktop computer

Big Data is commonly used to explain the application of analytical techniques to search, collect and cross-reference datasets in large numbers to develop systems of intelligence and insight.

6) Ward, J., & Barker, A. (2013). Undefined by Data: A Survey of Big Data Definitions.

7) Manovich, L. (2011) 'Trending: the promises and the challenges of big social data', in Debates in the Digital Humanities, ed. M. K. Gold, The University of Minnesota Press, Minneapolis.8

using standard software.[7] Rob Kitchin (2014) stated that Big Data is the result of the development and convergence of various technological advancements.[8]

Indeed, in most cases, the term ' Big Data ' generally refers to a combination of volume, velocity, variety and veracity (4V) - a a concept developed by Gartner (2012), and then mostly appropriated by IBM.[9] This data-centric approach mainly emphasizes the magnitude of the increase in data, the increase in speed, the instant production of data, and the variety of formats of data being processed.[10] In fact, every second, the world produces more data than it stored 20 years ago, all from different data sets, from various sources and in various formats, ranging from audio, GPS signals, social media interactions, and various types of contents.[11]

The process of massive data collection becomes easier to do, with the world's population increasingly connected to the Internet, connected and interacting with each other through online services. The amount of data generated by each person connected to an Internet device are growing at an exponential rate. Financial institutions, companies and healthcare providers generate large amounts of data through their interactions with suppliers, patients, customers and employees. Beyond that interaction, large volumes are made through searches on the Internet, social networks, GPS, and stock market transactions. This widespread data production then produces a "data revolution" or the *Age of Big Data*, which some experts say is at the core of the fourth industrial revolution.

Massive data collection practices were initially performed by the Information Awareness Office (IAO) established by the Defense Research Agency (DARPA) of the United States in January 2002. This office was established

8) Rob Kitchin, Big Data, New epistemologies and paradigm shifts, Big Data & Society April–June 2014: 1–12.

9) Ward, J., & Barker, A., Op.Cit.

10) Goes, Paulo, B,. (2014). "Big Data and IS Research", MIS Quarterly Vol. 38 No. 3 pp. iii-viii.

11) Mcafee, A., & Brynjolfsson, E. (2012). Big Data: The management revolution. Harvard Business Review, 90(10), 60-6, 68, 128.

after the terrorist attacks on September 11, 2001, with the intention of bringing together several DARPA projects that focused on the application of surveillance technology and information to track and monitor terrorists and other asymmetric threats to US national security, by achieving *Total Information Awareness* (TIA) This objective was achieved by creating a huge computer database to collect and store personal information of everyone in the United States, including personal e-mails, social networks, credit card records, telephone calls, medical records, and many other sources, without the requirements of a court warrant. Information is then analyzed for suspicious activities, connections between individuals, and threats. In addition, the program included funding for biometric surveillance technology that can identify and track individuals using surveillance cameras, and other methods. However, following public criticism that the development and application of this technology has the potential to lead to a mass surveillance system, the IAO was disbanded by the Congress in 2003. However, several IAO projects continue to be funded and run under different names, as revealed by Edward Snowden (2013). The methods carried out by the IAO at that time, later became the backing and the basis for the work of mass data collection today, including the trigger for the data revolution, which in the end all people, institutions, and countries do.[12]

Big Data is commonly used to explain the application of analytical techniques to search, collect and cross-reference datasets in large numbers to develop systems of intelligence and insight. This large collection of data can be obtained from general sources, as well as a collection of customer data for a particular company. Big Data gradually includes not only general data, but also includes information collected by the private sector. However, until now there has not been a single definition of Big Data. Some experts, for example, define Big Data as the emergence of new data sets with large volumes that change rapidly, are very complex, and exceed the range of environmental analysis capabilities of common

12) See: Elliot D. Cohen, Mass Surveillance and State Control: The Total Information Awareness Project, Palgrave Macmillan, 2010.

hardware and software used for data management. In short, the volume of data becomes too large to be handled by conventional tools and methods.[13]

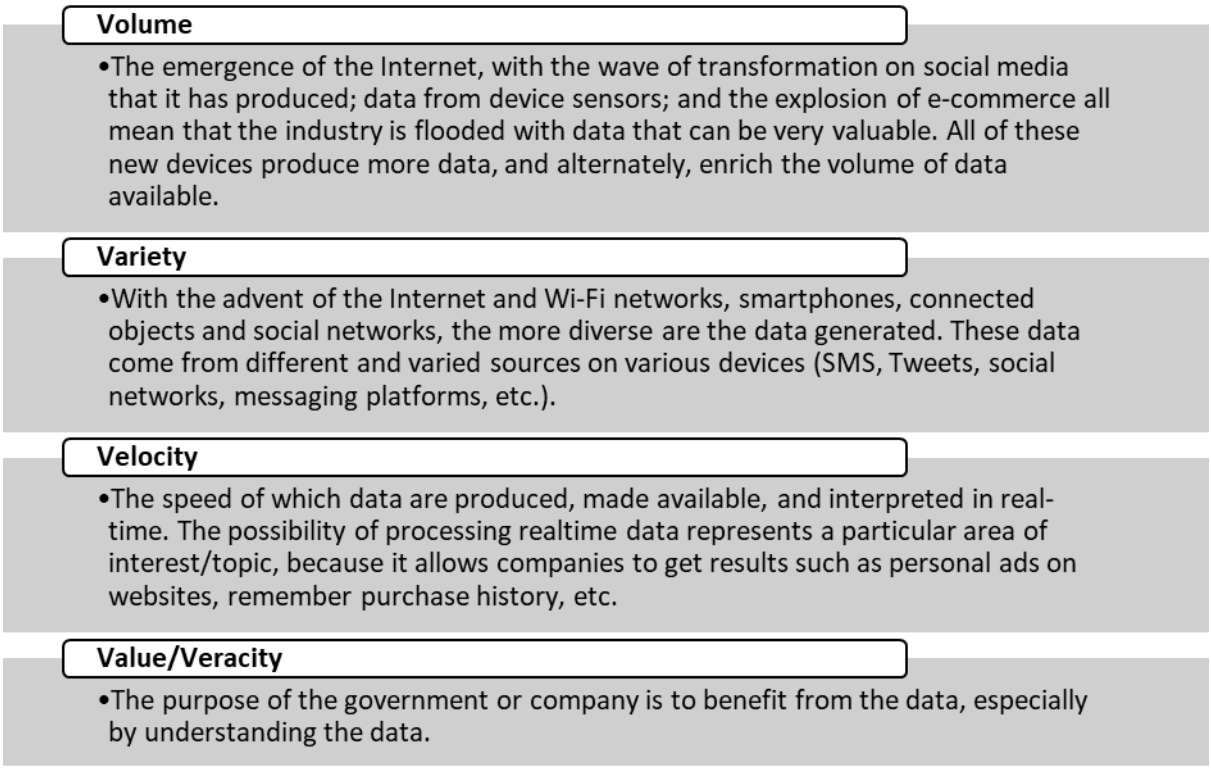
Big Data itself is intended to analyze, process, and store large data collections originating from different sources. Big Data is usually required when traditional data analysis, technology and processing and storage techniques are no longer sufficient. Big Data meets different requirements, such as combining several unrelated data sets, processing large amounts of unstructured data and harvesting hidden information in a time-sensitive manner. In addition to traditional analytical approaches based on statistics, a new aspect of Big Data is the addition of new techniques that utilize computational resources and the approach to run algorithmic analytics. This shift is important because datasets continue to grow to be larger, more diverse, more complex and streaming-centric (real time). The progress in computational science has enabled the entire processing of datasets, so that sampling from a population as before did not need to be done again.[14]

The data revolution has allowed a new generation of information technology designed to provide the speed of processing, analyzing and extracting values from large data sets, using special softwares. The Big Data phenomenon does not only refer to the explosion in the volume of data produced, which is made possible by developing information storage capacity and dissemination on all types of platforms, but also refers to phenomena involving new discoveries, related to data processing capabilities through data intensive system computation techniques. The growth of available data in terms of quantity, diversity, speed and access value is significant, giving way to the "Four Vs": "Volume", "Variety", "Velocity" and "Value", as mentioned above, which are the features of what is called Big Data:[15]

13) Babak Akhgar, et.al. (eds.), Application of Big Data for National Security: A Practitioner's Guide to Emerging Technologies, (Oxford: Butterworth-Heinemann, 2015).

14) Thomas Erl, Wajid Khattak, and Paul Buhler, Big Data Fundamentals *Concepts, Drivers & Techniques*, (Boston: Prentice Hall, 2016).

15) James R. Kalyvas and Michael R. Overly, Big Data A Businessand Legal Guide, (New York: CRC Press, 2015).



In addition to the attributions above, other key characteristics of Big Data are: (i) depth in terms of scope (data), by trying to capture the entire population or system, or at least a much larger sample size than conventionally used in studies of data in small sizes; (ii) a very high resolution, with the aim to provide as much detailed analysis as possible, unique and identical identification; (iii) relational in nature, containing common columns that allow combinations of different datasets; and (iv) flexible, has extensionality (can add new fields easily), and measurable (can expand in size quickly).[16]

The increasing development of softwares for processing large amounts of data (Big Data) has opened up the possibility of collecting, processing, and expanding the volume of data that is increasingly larger, which has also become the driving force in this century's innovation. This development also allows the presence of an intensive data system, which combines

16) Danah Boyd & Kate Crawford, Critical Questions For Big Data Provocations For A Cultural, Technological, And Scholarly Phenomenon, in Information, Communication & Society Vol. 15, No. 5, June 2012, pp. 662–679.

various systems at once, based on data. The term intensive data system refers to a combination of data systems, which among others consist of search engines, recommendation systems, biological systems, space exploration, and financial analysis. The basic characteristics or elements of intensive data systems require a connection between aggregated content, reference structures, algorithms, and data structures. The development of a computational system for processing data on a large scale is instrumental for the birth of the data intensive system concept, which puts Big Data as its main feature.[17]

C. SUPPORTING FACTORS AND REASONS FOR LARGE SCALE COLLECTION OF DATA

Big Data is a ubiquitous keyword, although it is not clearly defined. Often this innovation is considered a positive development for several sectors, such as the private industry, research, medicine, science, government, and humanitarian assistance. With a number of key characteristics possessed as explained above, Big Data is considered capable of solving various problems that arise, ranging from social, economic, to political. Supporters of Big Data often claim that the application of sophisticated algorithms to very large amounts of data will provide broader insight into any topic. Big Data often used as a justification to collect large amounts of information, process information for a variety of unforeseen reasons, beyond the willingness of individuals to share information (data), and use that information to get a profile of individuals, groups and even society as a whole.

This rapid growth can arise because of the simultaneous development of a number of elements, including technology, infrastructure, enabling techniques and processes, and rapid integration of these elements into business and everyday social practices. This new knowledge infrastructure, covering various information and communication technologies, especially the static and dynamic Internet; embedding software into all types of objects, machines and systems, the emergence of 'smart' technology, and the creation of pure digital devices and systems; development of computing and the ability to access networks

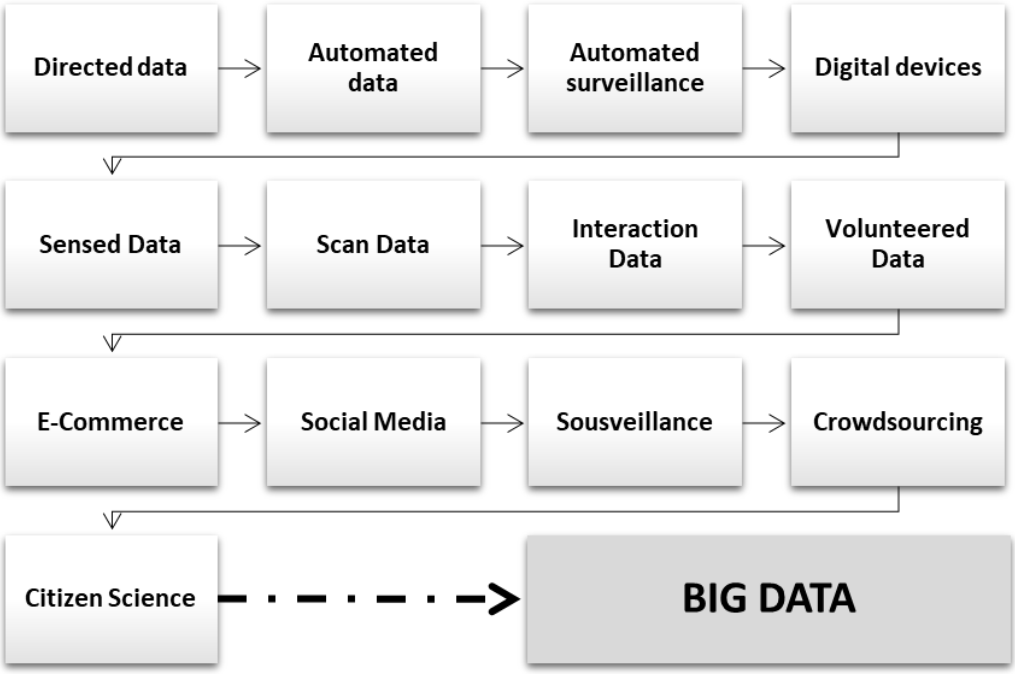
17) Chris A. Mattmann, et.al., Architecting Data-Intensive Software Systems, dalam B. Furht and A. Escalante (eds.), *Handbook of Data Intensive Computing*, Dordrecht: Springer, 2011, hal. 25.

and computing in many environments; progress in database design and information management systems; increasingly affordable data storage and distribution, and a long period of storage (forever); and new forms of data analytics designed to address the abundance of data compared to the scarcity of data.[18]

In addition to enabling access, sharing and storage of data, today's knowledge infrastructure also allows large-scale data production. In general, the rapid development of Big Data is made possible by a number of the following supporting environments: (i) development of computational techniques; (ii) development of networked equipment; (iii) devices that are connected to each other, and the ability to spread quickly or pervasively through digital systems; (iv) the ability of the machine to read and identify (indexical identification); and (v) the increasing capacity of storage and ease of exchange and distribution. The radical expansion and integration of computing technology, networking, development of digital devices and data storage have provided a powerful platform for the explosion of data on a large scale. The combination of a number of these factors has become the means used to generate big data, data processing, data distribution, and data analysis.

Further, in large-scale data collection, data are obtained at least through three approaches: (i) data generated from traditional forms of observation, by directing technology to a person or place by a human operator; (ii) data generated automatically via use of a device or system; and (iii) voluntarily generated data, when people submit data to the system. The details of how large-scale data collection are generally carried out are shown through the following methods:

18) Rob Kitchin, The Data Revolution Big Data, Open Data, Data Infrastructures and Their Consequences, (London: Sage Publishing, 2014.



1. Directed data

Data collection aimed at a specific purpose, carried out by state and private authorities, for various purposes with various methods, including all forms of surveillance conducted in an organized and structured manner by certain institutions, for example, law enforcement officials, intelligence officers, government bureaucrats, or service provider companies. This collection is intended to observe certain people or citizens in general, either directly or through technological devices, including surveys, censuses, filling out forms, tax receipts, inspections, camera observations, carried out by the state and private sector. The system is generally equipped with individual identification methods, such as fingerprints, photographs, insurance or social security numbers, passports and driving licenses, which makes it easy to do the tracking.[19]

19) David Lyon, Surveillance Studies: An Overview, (London: Polity Press, 2007). See also: Daniel J. Solove, *Understanding Privacy*, (Cambridge, MA: Harvard University Press, 2008).

2. Automated data

Through this system, data are generated automatically by various digital technologies with little human supervision. In addition, the data are also automatically processed, analyzed and followed up by algorithms to produce certain patterns or findings designed to regulate a phenomenon. There are several methods used to generate automatic data. Some of them are even byproducts of a system, which is not the main goal of the system.[20] The development of Internet of Things (IoT) technology in devices of household appliances, such as televisions, printers, refrigerators, and other electronic devices, is one example of data are collected automatically.

3. Automated surveillance

Current technological developments make it possible to automate various aspects of monitoring systems, including adding new techniques, so that monitoring is more effective and efficient in tracking different systems and places. In addition, it is also possible to perform real-time tracking, which has the use of matching demand with supply and finding faults/leakages in a system.[21] For example, in addition to being used to carry out commodity monitoring in the supply chain, this system is also used in monitoring traffic management or transportation systems, using smart cards equipped with RFID.

4. Digital devices

Digital devices have helped to facilitate, improve and manage many aspects of daily life. Devices not only process digital data and provide further instructions, but also actively produce digital data, both as a primary or a side function. For example, cameras, videos, telescopes, GPS, and various other forms of equipment, which are equipped with networked technology (Wi-Fi). Even mobile phones can yield data about how the tool is used, for example related to time, location, person referred to in the

20) Martin Dodge and Rob Kitchin, Code/Space: Software and Everyday Life, (Cambridge: the MIT Press, 2011).
21) V. C. Gungor, D.Sahin, TaskinKocak, SalihErgut, Concettina Buccella, Carlo Cecati, and Gerhard P. Hancke. A Survey on Smart Grid Potential Applications and Communication Requirements. IEEE Transactions on Industrial Informatics, Vol. 9, No. 1, pp.28-41, 2013.

conversation, and related data are logged and sent to a third party (such as device manufacturers and service providers). The explosion of data becomes increasingly massive along with the growing magnitude of the use of smart phones, which are capable of generating primary data, including recordings of notes, images, emails, comments on social media, and notes on the use of various applications.[22]

5. Sensed data

Certain types of digital devices can function both as sensors and actuators at once. This system can be embedded in devices or placed on different structures to measure certain outputs such as light levels, humidity, temperature, gas, chemicals, electrical resistance, air pressure, movement, speed, and so on. Even miniature sensors and actuators can create a continuous flow of data. They can work passively and be read by scanners, or can be active, broadcasting data periodically through limited networks or in large areas. In fact they may have near field communication abilities, as embedded on a number of mobile phone devices, which allows two-way communication.[23]

6. Scan data

In addition to digital devices and sensors that actively generate data and a part of the Internet, other objects that can be tracked digitally or through a specific identification code can also generate data automatically. The use of barcodes in trade and industry contexts and official forms provided by the government has enabled the collection of data on a massive scale.[24]

7. Interaction data

While some Internet users see the social media as a rather anonymous terrain, almost all interactions across the Internet

22) Martin Dodge and Rob Kitchin, Software, objects, and home space, in Journal Environment and Planning volume 41 (2009), pp. 1344-1365.

23) Gerhard P. Hancke, et.al., Op.Cit., See also: Viktor Mayer-Schonberger, Kenneth Cukier, Big Data: A Revolution That Will Transform How We Live, Work and Think, (Boston: Houghton Mifflin Harcourt, 2014).

24) See: Martin Dodge and Rob Kitchin, Code and the Transduction of Space, in Annals of the Association of American Geographers Volume 95, 2005.

network generate interaction data. In fact, the generation of data is inherently built into various forms of communication and uses facilitated by the Internet. For example, an Internet service provider can track the sites accessed by the individual user accounts and computers, and many websites, especially company websites, utilize cookies to generate clickstream data, tracking how a user navigates through a website and which features are clicked. Generally, the data are used to monitor the performance of a website, conduct market research and track employee productivity, or to be sold to third parties.

8. Volunteered data

In contrast to observations directed at people or objects by individuals and certain institutions, or captured automatically as an inherent feature of a device or system, much large-scale data are actively produced by the people (users/consumers) voluntarily. In such cases, individuals produce and enter data to use services (such as social media) or to take part in collective projects (such as Wikipedia). In this concept there is no longer a relationship between producers and consumers, individuals have taken on the role of producers of data and consumers of content services at the same time.[25]

9. E-commerce

In an online transaction, the buyer is required to enter their personal data, which is related to payment, address and other metadata needed, as part of the verification and buying and selling process, as well as ways to learn more about the consumer profile. As when someone fills out a form provided by the government voluntarily in their capacity as a citizen, in the context of online transactions, there is even a data exchange for a service that goes beyond cash payments. Additional data provided at the time of the transaction allow imperfect profiles to be developed from time to time or to be linked to other data sources, such as demographic data and location.[26]

25) George Ritzer and Nathan Jurgenson, Production, Consumption, Prosumption The nature of capitalism in the age of the digital 'prosumer', Journal of Consumer Culture Vol 10(1): 13–36 1469-5405 (2010).

26) In Lee, Transformations in E-Business Technologies and Commerce: Emerging Impacts, (Hershey, Business Science Reference, 2012).

When someone often shops at a store and uses customer cards, the company can use the customer database, based on customer card information and shopping habits of customers, and compare the data with census data containing income, property ownership, age and other information about certain aspects. The company can then identify which customers are most likely to respond to certain ads. In comparing and merging data sets, Big Data analysts can connect pieces of data, which, if standing alone, contain information that seems harmless, to compile a detailed picture of an individual. The creation of new personal information through the application of this predictive algorithm creates serious privacy problems: when joining a customer card scheme, consumers may be willing to share data, but may not like - or even disagree - the use of the data in the end.

10. Social media

The social media is one of the modern innovations that allows users to communicate around the world in real time. Social media includes all forms of electronic communication (such as websites for social networks and microblogging), in which users create online communities to share information, ideas, private messages, and other contents (such as pictures, videos). Messages are not only targeted at certain individuals, but are also open to the public.

Michael Mandiberg (2012) argues that the idea of "social media" has been linked to various concepts, especially the favorite corporate media of "user-generated content". Henry Jenkin (2009) focuses on the culture of convergence. Jay Rosen (2013) focuses on the phenomenon of people formerly known as the audience or bystanders, who then actively participate in politics. Yochai Benkler (2013) emphasizes the concept oriented to peer-production; and Tim O'Reilly (2005) identifies social media in Web 2.0-oriented computer programming.[27]

27) See: Christian Fuchs, Social Media: a Critical Introduction, (London: Sage Publications, 2014). See also: Jed Hallam, Social Media Manifesto, (New York: Palgrave Macmillan, 2013).

The use of social media or Web 2.0 sites relies heavily on the active participation of the public who are willing to share information about their lives and do work such as writing, editing, expanding, remixing, posting, sharing, tagging, communicating, and so on. By doing so, they not only do voluntary work, but also provide a lot of data about themselves, including user profiles, photos, locations, opinions, preferences and values, and their network of social contacts uploaded through the applications that they use. In general, these services are owned by newly established companies (startups), so the data collected can later be traded to other parties, or used to create new models of capital accumulation by way of extracting value from the data.[29]

11. Sousveillance (self surveillance)

Sousveillance is self-monitoring and management of one's personal health and life through digital technology, for example, fitness equipment equipped with computers that can record data about a person. Unlike surveillance or monitoring in general, when an individual is monitored from an external position by another entity, sousveillance is consciously used and controlled by the individual to fulfill their personal needs. Sousveillance has encouraged people to actively monitor and record their personal data. In general, individuals monitor the health aspects and fitness of their bodies, capture their consumption data (for example, food/calorie intake), physical conditions (such as blood pressure, pulse), emotional states (for example, mood, passion), and performance (for example, walking, running or cycling distance). The data produced do not need to be shared with the wider public, although it can be, but it is voluntarily shared with the service provider companies, in the form of very rich personal data.[30]

29) See: Andrea Ceron, Luigi Curini and Stefano M. Iacus, Politics And Big Data: Nowcasting And Forecasting Elections With Social Media, (New York: Rotledge, 2017).

30)See: Steve Mann, Jason Nolan and Barry Wellman, Sousveillance: Inventing and Using Wearable Computing Devices for Data Collection in Surveillance Environments, in Surveillance & Society 1(3): 331-355, (2003). See also: Steve Mann, Surveillance (oversight), Sousveillance (undersight), and Metaveillance (seeing sight itself), in IEEE Conference on Computer Vision and Pattern Recognition Workshops (2016).

12. Crowdsourcing

Crowdsourcing is a collective generation of media, ideas and data that is voluntarily carried out by many people to complete a particular task. While social media content can be said to be crowdsourced, in the sense that it comes from a large number of people, the objectives are diverse and unfocused. Instead, crowdsourcing focuses on the collective production of information and creating solutions to certain issues by utilizing the energy, knowledge, skills and consensual and collective action of a crowd of people.[31] There are three main types of crowdsourcing: those who collectively produce solutions, such as OpenStreetMap and Wikipedia; those who use the crowd of people to evaluate various solutions/products, such as Amazon and Trip Advisor reviews; and those who are looking for a single or multiple solutions from a crowd of people, such as InnoCentive and hackathons. To encourage this movement, companies usually organize contests to encourage people to get involved in the project to find the solution. The same thing is done by city governments by holding open data application contests. Such contests can take the form of hackathons, where people work together in teams to develop solutions and applications that are thought to be an applicable solution for a problem.[32]

13. Citizen Science

Citizen science is a special form of crowdsourcing where 'communities or citizen networks act as observers in several fields of science'. Here, people produce, prepare, and process empirical observations and detailed measurements of a phenomenon, which have crucially real value as data points for science. Citizen scientists are very useful in terms of observations because they can be distributed geographically and planted in place through time. Examples include amateur meteorological measurements made from independent weather stations,

31) Jeff Howe, Crowdsourcing: Why the Power of the Crowd Is Driving the Future of Business, (New York: Crown Publishing, 2011).

32) See: David Weinberger, Too Big to Know: Rethinking Knowledge Now That the Facts Aren't the Facts, Experts Are Everywhere, and the Smartest Person in the Room Is the Room, (New York: Basic Books, 2011).

and astronomy hobbyists who carefully observe the sky. From the results of personal observations, the data are exchanged and distributed voluntarily, which if collected can be large-scale data.[33]

Many experts state that the evolution of data does not take place non-ideologically and passively. As well as all revolutions, of course, it is driven by a series of strong arguments, passed on by those who believe on the benefits of the new ways or means, which will provide utility to private lives or group interests. In that context, the data revolution is developed, with various arguments and reasons, and offers for a new radical way of understanding that could manage all aspects of human life. In general, collecting large-scale data as the core of data revolution are at least based on four reasons: (i) the need to manage people; (ii) the need to manage an organization; (iii) utilizing value and generating capital; and (iv) creating a better place or situation (health, welfare, social and human development, ability to address social and ecological problems).[34]

1. Need to Manage People

In the effort to carry out its functions, from the beginning a state certainly strives to create a more systematic method to manage and regulate its population, including management of services to its citizens. One of the main ways that has been attempted to effectively and efficiently carry out these roles is through audit and quantification of society - creating detailed knowledge of its jurisdiction through the development of a widespread administrative system, supported by a variety of technological apparatus, which measures and tracks index data and attributes about individuals, companies, institutions and other elements.[35] In all government institutions, collection of data, management,

33) Michael F. Goodchild, Citizens As Sensors: The World Of Volunteered Geography (2007). See also: Linda See, et.al., Crowdsourcing, Citizen Science or Volunteered Geographic Information? The Current State of Crowdsourced Geographic Information, in (ISPRS International Journal of Geo-Information, 2016, 5, 55).

34) Viktor Mayer-Schönberger and Kenneth Cukier, Big Data: A Revolution That Will Transform How We Live, Work and Think, (Buena Vista, Mariner Books, 2014).

35) Martin Dodge and Rob Kitchin, Mapping Cyberspace, (London: Routledge, 2001).

storage and analysis of data, are most fundamental tasks, which will then be used to assess the obligations and rights of citizens of the state, as well as to detect non-compliance, in connection with the predetermined legal norms and obligations.[36]

In order to facilitate these processes, the state tries to build and utilize various systems to implement more sophisticated administrative methods. Included in this effort are the application of systems, technology and sophisticated infrastructure, which are designed to capture and process data effectively, including a stronger identification system (signatures, photos, fingerprints, social security and passport numbers, tax codes, licenses and account numbers, biometric information, usernames and passwords, postal codes, etc.), which can be operated real time and from a certain distance, even across the network.[37]

Therefore, Big Data is certainly not new to the state, they have long been dealing with large amounts of data in managing jurisdiction, some of which have been processed or acted upon in or near real time when agencies interact with the subject. However, Big Data is seen as the latest set of technologies that can expand and enhance the work of the state by increasing the timeliness and expansion of calculative practices. In some cases, the discursive drive to big data solutions is being driven by vendors who want to sell services to the state, in other cases, the state is actively leading the process. At present, there are two main discursive strategies used to encourage the use of large data. The first involves the improvement of public administration and cost savings through increasing 'operational efficiency, reducing costs due to maladministration, and increasing tax revenues by narrowing the tax gap.[38] The second is concerned about state security and handling of crime. All countries are involved

36) Wendy Hartanto, The 2010 Indonesia Population Census, Badan Pusat Statistik, dapat diakses di http://jmstat.com/publications/The%202010%20Indonesia%20Population%20Census-final.pdf.

37) Martin Dodge and Rob Kitchin, Code and the Transduction of Space, Journal of Annals of the Association of American Geographers Vol. 95 Issue 1 (2005).

38) J. Manyika, M. Chui, B. Brown, et al. Big Data: The Next Frontier for Innovation, Competition, and Productivity. McKinsey Global Institute (2011).

in surveillance for the purpose of security, safety and crime prevention, implemented by the police and intelligence agencies. Thus Big Data is also seen as a way to combat crime and protect national security in a more efficient manner.[39]

2. Need to Manage an Organization

Just as with any country and government institution, which continues to adopt newest management practices and technologies for the interests of their operations, businesses and other institutional entities, such as political parties and civil society organizations, also began to make similar moves. Data become a key aspect of how organizations understand themselves and their constituents – their markets, which will then serve as a basis for making decisions about their current and future operations. Data provide evidence, which is processed to manage organizations more effectively, efficiently, competitively and productively, and to identify products and the creation of new markets.

On a large scale, data are seen in a continuous process to find organizational efficiency and structures that benefit from competitive advantages, reducing risk, operational costs and losses, and enhancing customer experience. Big Data offers the possibility of complete, highly detailed, and real-time intelligence in all aspects of the organization. In other words, Big Data will provide a very detailed and very timely knowledge on all aspects of their organization, which can then be used to formulate appropriate actions (policies).[40]

3. Utilizing Profits and Producing Capital

In addition to making companies more efficient and effective in their management, Big Data is being promoted on a large scale as a means to increase profits from the activities of the company and its relationships with suppliers and consumers. Indeed, as noted by Gantz and Reinsel (2011), the main drivers of big data are not technology, but finance, the promise of

39). Martin Innes, Control Creep, in *Sociological Research Online* Vol. 6 Issue 3 (2001).

40) J. Manyika, et.al., Op. Cit.

greater efficiency and profits. The application of Big Data solutions is thus a strategy to realize capital accumulation that has not been used yet by offering a significant return on investment and providing a competitive advantage.[41]

4. Creating A Better Place

Throughout history people have found and adopted new technologies designed to create wealth and improve their daily lives. Many of these technologies greatly affect the nature, structure, regulation and urban and rural life, especially utilities, transportation, communication, and building infrastructure. However, these technologies often have conflicting and harmful effects (for example, factories that create wealth but also pollution). Similarly, information and communication technology (ICT), which is the latest wave of technological devices and infrastructure, is expected to provide solutions to problems created by the technologies of the previous period.

Based on this idea, in the past decade, cities have used ICT as a key component of their regional development strategies. The latest development of the vision of urban development controlled by ICT is the idea of a smart city. A smart city has powerful instruments that can be understood and regulated in real-time, with the entire system based on Big Data. In this context Big Data provides a comprehensive, dynamic, cohesive understanding of the city and input to the system that will create a place that is more livable, safe, functional, competitive and sustainable.[42]

This vision of the smart city is heavily promoted by a number of software and hardware companies, such as IBM, CISCO, Microsoft, Intel, Siemens, Oracle, SAP, and enthusiastically adopted by various cities in the world. All hope that, with smart city technologies, cities in the world will experience rapid socio-economic progress, and be able to renew the city center as a center for innovation and work. The main

41) Paul C. Zikopoulos, et.al., Harness the Power of Big Data The IBM Big Data Platform, (New York: McGrawHill, 2012).

42) Anthony M. Townsend, Smart Cities: Big Data, Civic Hackers, and the Quest for a New Utopia, (New York: W. W. Norton & Company, 2013).

function of big data in this technology is to provide real-time analytics to manage the functions of the city and manage cities. Such real-time surveillance and data analytics have been used for several years in several sectors, such as transportation and public service systems, which are often managed monopolistically by a single vendor.[43]

D. THE PRACTICE AND RISKS OF LARGE SCALE DATA COLLECTION IN INDONESIA

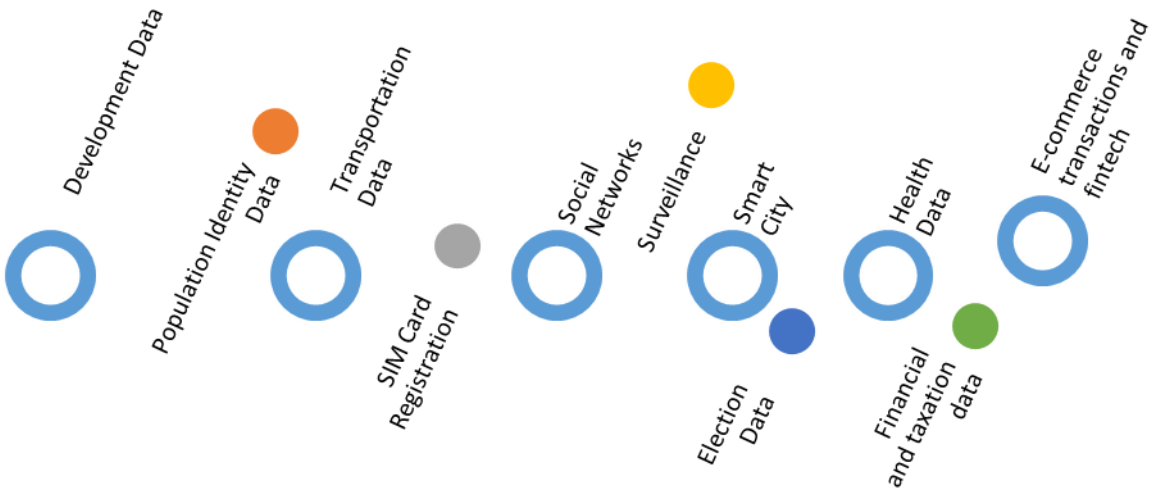
As is the case in other countries in the world, massive data collection practices are also carried out in Indonesia, through multiple methods and mechanisms, as well as for a variety of reasons and interests. Collection is carried out both by government institutions with various data recording methods, as well as private companies in various business sectors. In particular, UN Global Pulse, which developed the Big Data regime with a view to sustainable development and humanitarian actions, specifically established its laboratory in Jakarta, in addition to Kampala and its headquarters in New York. Indonesia is chosen on the grounds as the country with the largest population in ASEAN, Internet penetration reaching 50% of the total population (260 million people), and active social media users reaching 40% of the total population. In Indonesia, Pulse Lab mainly works on issues related to the information gap in development and humanitarian issues, by optimizing the implementation of Big Data, real-time analysis, and the development of artificial intelligence.[44] UN Global Pulse is one of the earliest bodies pioneering the use of Big Data, launched in 2009, in recognition of the need for more timely information to track and monitor the impact of socio-economic crises at the global and local levels. This initiative explores how digital data sources and real time analysis technologies can help policymakers understand community welfare and the emergence of vulnerabilities, in real time,

43) Karima Kourtit, Smart Cities in the Innovation Age, in Innovation: The European Journal of Social Science Research Vo. 25 Issue 2 (2012).

44) See: https://www.unglobalpulse.org/jakarta.

in order to better protect the public from shock.[45]

Broadly speaking, data collection practices in Indonesia, both by the government and private sector, can be illustrated by the following practices: (1) collection of development data, for example, poverty data, population census, economic census, disaster data, etc; (2) population identity data, especially electronics-based ID card (e-KTP); (3) SIM Card registration for users of mobile telephones; (4) communication surveillance and direct data access to databases, including maps; (5) smart city projects; (6) election data, collected through the voter registration process; (7) health records, both medical records and health insurance, and other social security; (8) financial and taxation data, whether collected by banking, financial services, insurance companies or tax offices; (9) transportation data, especially those collected by providers of online transportation platforms; (10) social networking, including the use of apps and social media; and (11) e-commerce and financial technology transactions



45) UN Global Pulse clearly identifies privacy issues related to the use of Big Data and the impact of privacy in "Big Data for Development: Challenges & Opportunities" and has adopted the Principles of Data Privacy and Protection. While these are positive steps in the right direction, there is still much to be done, given the increasingly complex network of actors involved, the wider scope of their work, the more data that can be collected from individuals, and the lack of existing legal protection.

-Collection of Development Data

Initially data collection was carried out first through the process of enumeration or population census, which had been carried out since the Dutch East Indies period, in 1920 and 1930. After Indonesia gained independence, the population census was carried out for the first time in 1961, and the latest in 2015.[46] Data collected through the population census process, other than population data, can identify address, gender, age, income, and religion of each resident. In addition to the population census, a number of other censuses were also conducted by the Indonesian government, through the Central Bureau of Statistics, for development purposes, such as economic censuses and agricultural censuses. Poverty data is also collected on an annual basis, through the construction of an integrated database, which is intended to determine the number of poor people, as well as interventions and approaches that need to be carried out in the context of empowerment. The collection of data is of course very important for development, however, because the material is closely related to one's personal data, it must be handled and managed in reference to the principles of personal data protection

- Population Identity Data

Massive recording and collection of data through population identity data are carried out in Indonesia through the National Single Identity policy, by making the Population Registration Number (NIK), which is applied in the Electronic KTP (e-KTP) program. This program was launched in early 2011, and assigns a single identity of each resident, which is valid for life, one card for each resident, in which the NIK is spelled out. Furthermore, recording of population data is carried out by the government in the context of implementing this program. All personal information of citizens is recorded, including their identity and physical characteristics. Specifically for physical features, this is carried out by scanning the fingerprints and retina, which will be used for biometric validation of the ID card holder. According to Ministry of Home Affairs information, the results of recording the data will then be planted on the KTP, first encrypted using a certain cryptographic algorithm.

Some questions need to be asked regarding the practice of recording, due to differences between the regulations and the practice in the field. For example, related to e-KTP security system, according to Presidential Regulation No. 67 of 2011, the biometric validation will only use fingerprint scanning, but in practice, the retina is also recorded. Who will be responsible for the results of the data recording? Even though Law No. 23 of 2006 concerning Population Administration,

46) Wendy Hartanto, The 2010 Indonesia Population Census, Badan Pusat Statistik, dapat diakses di http://jmstat.com/publications/The%202010%20Indonesia%20Population%20Census-final.pdf.

as well as its amendment (Law No. 24 of 2013), have regulated the personal data of citizens, including the government's obligation to protect its confidentiality, Presidential Regulation No. 67 of 2011 does not clearly and in detail regulate the mechanism for protecting the personal data of citizens who have been recorded. This ambiguity appears on the measures undertaken by the Ministry of Home Affairs, which in fact cooperates with at least 218 agencies, both government and private, to the use of population data (e-ID).[47] These institutions also range from law enforcement, intelligence, banking, to telecommunications service providers.[48] The practices carried out without agreement from the owners of the data, of course, are actually a form of violation of privacy protection, because it allows the practice of arbitrary transfer of data.

Especially with regard to personal data, the application of a single identity such as an e-KTP will not only provide convenience in administration, but also create many threats and disasters. Davies (1996) notes that there are at least twelve things that must be considered regarding the use of a single identity. His criticism, among others: the card is not going to stop crime; it actually facilitates discrimination; causes a major disaster if the card is lost; endangers the protection of personal information; will not stop fraud in social security; and will not stop illegal immigration.[49]

- SIM Card Registration

Recently the government, through the Ministry of Communication and Information, announced the obligation for cellular telecommunications customers (prepaid cards) to re-register, by requiring customers to send their Population Identity Number (NIK) and family card number. Referring to Permenkominfo No. 12 of 2016 and Permenkominfo No. 14 of 2017, which was amended again with Permenkominfo No. 21 of 2017, as references for the implementation of this policy, the purpose of re-registration is not clearly stated. It was simply mentioned by the ministry that this is necessary because many SIM cards are misused. Although SIM Card registration obligations are common in many countries, the lack of guaranteed protection of personal data and privacy in general in Indonesia has become a

47) See: "BPK: Dipakai Korporasi, Data Kependudukan Bisa Disalahgunakan", in https://www.cnnindonesia.com/ nasional/20180410111632-20-289638/bpk-dipakai-korporasi-data-kependudukan-bisa-disalahgunakan.

48) See: "Banyak Lembaga Manfaatkan Data, Ditjen Dukcapil Raih Berbagai Penghargaan", in http://www.dukcapil.kemendagri.go.id/detail/banyak-lembaga-manfaatkan-data-ditjen-dukcapil-raih-berbagai-penghargaan. See also: "Kemendagri & sejumlah bank kerja sama akurasi data nasabah dari eKTP", in https://www.merdeka.com/peristiwa/kemendagri-sejumlah-bank-kerja-sama-akurasi-data-nasabah-dari-ektp.html.

49) Simon Davies, Big Brother: Britain's Web of Surveillance and The New Technological Order, (London: Pan Books, 1996).

potential threat to the enjoyment of the right to privacy of citizens. Not to mention, customers are asked to send NIKs and family card numbers at once, to be able to synchronize with Population and Civil Registration Data.

The potential threat to the right to privacy of citizens from the SIM Card registration process, as a result of the lack of guarantee of protection of citizens' personal data, can be reflected in the registration process itself. Although NIK data and family card number collection is carried out centrally by the government, by sending SMS to certain numbers, the data validation process is still carried out by the operators. This means that the first party who will collect and process all personal data of customers is the service provider. In addition, technically, all messages sent by customers (subscribers), will also first enter the Short Message Service Center (SMSC) managed by the operator. Then, what mechanism or rule can ensure that the customer's personal data are protected in the process? This concern is raised in line with the alleged leak of personal data in the implementation of the SIM Card re -registration policy for cellular phone service users.

The experience in various countries shows that the SIM Card registration obligation has created many new problems, as recorded in the following notes:

a. ***The myth of crime busting***, the opinion that the registration of SIM cards is a boon for law enforcers has been consistently proven wrong by experience in several countries. This has been shown many times, criminals generally can avoid this kind of regulation. Worse, in some countries, registration actually causes a surge in the prevalence rates of several crimes (eg, theft of telephones) and even facilitates the emergence of black markets (i.e. for stolen SIMs). The proposal to have a SIM card registration system has been canceled in Canada, after several consultations, and it cannot be proven that crime reduction will occur after the adoption of this system. Mexico revoked the law requiring SIM Card registration after three years of implementation. This revocation was carried out after seeing no improvement in crime prevention, investigation and/or prosecution.

b. ***The specter of surveillance,*** one of the most worrying aspects of the practice of SIM card registration is the potential for its use for surveillance.

The SIM card registration policy has removed the possibility to enjoy anonymity of communication. This can enable location tracking, and make surveillance and interception of communication quite easy. Thus, this practice is certainly a threat to investigative journalists, whistle-blowers, witnesses, marginalized groups, and victims of discrimination. It is important to note, that this mechanism is an even more powerful tool for surveillance, especially after synchronizing with the national ID system.

c. ***Resource nightmare,*** especially budgets, because SIM card registration is actually an expensive and very difficult practice to implement, not only from the side of the cellular service providers, but also for the relevant government agencies. This is especially true for countries that have large populations, most of whom are cellular telephone users. Creating and maintaining a database for many of these subscriptions will require very significant costs. For example in Nigeria, even governments and four telecommunications operators spend around $ 128M on their systems, and telecommunications companies still experience many problems during implementation.

d. ***Creating new discrimination or gaps.*** Another less well-known but very relevant problem is how SIM card registration can also have a negative impact on legitimate cell phone use. Difficulties in resources, finances, and privacy risks associated with the registration process will inevitably affect the purchase and use of SIM cards. This has the potential to cause some sectors or groups to experience disincentives or be discriminated against. For example people with limited mobility (for example, people with disabilities, residents of remote areas, etc.), will be placed in an unfavorable position if

59) Lihat: "Beberkan Rekam Medis Disidik Polisi", dalam https://nasional.tempo.co/read/59094/beberkan-rekam-medis-disidik-polisi.

60) Lihat: "e-health di Indonesia, apa bisa?", dalam https://inet.detik.com/cyberlife/d-2601038/e-health-di-indonesia-apa-bisa.

registration requires a personal appearance/profile. Such a system also would exclude those who do not have official identification (ID number) the practice of which is quite common in developing countries, which typically have poor identification system. For example in South Africa, cellular operators experience a decline in customers after the SIM card registration system was enforced. This is the same with the situation in Indonesia.

e. ***One step behind,*** because in its future development, the use of SIM cards in telecommunications will be considered obsolete. Telco companies themselves have often indicated that we might all be heading towards a future where telephones and other related communication devices no longer need a SIM card to operate. If this is true, we will all invest in complicated and expensive systems that are anchored in something that might not be valid for a longer period of time.

- ***Communication Surveillance***

The practice of communication surveillance in Indonesia as one of the mechanisms of massive data collection is synonymous with the work of intelligence agencies in relation to national security and law enforcement agencies in carrying out their duties. Based on available data and information, several companies providing the tools and technologies of surveillance are known to have relationships with the Indonesian intelligence agencies. This has been acknowledged by the Ministry of Defense, which said it had entered into a cooperation contract with Gamma TSE, a UK-based security company providing surveillance devices.[50] This was also confirmed by the request for warranty and insurance policies from Gamma TSE Ltd to the United Kingdom government in 2013, for the purchase of intelligence equipment by the Indonesian Ministry of Defense.[51] In addition to the purchase by the Ministry of

50). See: "Kemhan: Pengadaan Alat ANTI SADAP Untuk Amankan Informasi Strategis TNI", in http://www.kemhan.go.id/kemhan/?pg=31&id=1203.

51) For complete analisis: https://www.whatdotheyknow.com/request/216377/response/538538/attach/3/ FOI%2014%2028%20Annex%20A%202012%2013.pdf.

Defense, recently a report issued by Citizenlab in Canada also found allegations of the use of the Finfisher software by the State Cypher Agency (Lemsaneg), which has now been transformed into the State Cyber and Cypher Agency (BSSN).[52]

In addition to the collaboration above, recently Wikileaks also provided information on efforts to explore cooperation with Hacking Team of Italy, which offers their surveillance devices to a number of intelligence agencies in Indonesia. According to information from Wikileaks, the joint proposal from Hacking Team has been submitted to the State Intelligence Agency (BIN), Strategic Intelligence Agency (BAIS), Police, and Lemsaneg.[53] Currently there has been no adequate information regarding the use of surveillance tools from Hacking Team by Indonesian intelligence agencies.

In addition to surveillance practices carried out for the purpose of state intelligence, with emphasis on national security, as stipulated in the State Intelligence Law, surveillance practices are also widely carried out in the law enforcement process. This is regulated in a number of laws that regulate the authority and actions that can be carried out by law enforcers, in the process of law enforcement of acts of corruption, terrorism, narcotics, human trafficking, and money laundering. Unfortunately, even though these are aimed at helping the investigation process within the framework of law enforcement, regulatory materials regarding surveillance tend to differ between one law enforcement agency and another. It must be admitted that the practice of surveillance has helped a lot in disclosing criminal cases such as corruption and terrorism. Even the Corruption Eradication Commission (KPK) almost always uses evidence from the results of surveillance (tapping) in the process of verification in court. However, until now unfortunately there is no clarity regarding the audit in the use of surveillance equipment owned by law enforcement officials. Even if

52). See: Citizenlab, Pay No Attention to the Server Behind the Proxy: Mapping FinFisher's Continuing Proliferation, 15 Oktober 2015. Can be accessed on https://citizenlab.org/2015/10/mapping-finfishers-continuing-proliferation/.

53). See: https://wikileaks.org/hackingteam/emails/emailid/447886; https://wikileaks.org/hackingteam/emails/emailid/17660; https://wikileaks.org/hackingteam/emails/emailid/578187; https://wikileaks.org/hackingteam/emails/emailid/17642.

the audit of the surveillance system of law enforcement institutions has been carried out, the public has never received sufficient information regarding the results of the audit.

For example, the cost of the wiretapping equipment owned by the Attorney General's Office, the public only knows the nominal price of the tool, which was obtained from the process of hearing between the Attorney General's Office and the DPR.[54] The wiretapping equipment worth IDR 16 billion has been installed since December 2010, and has been operational since April 2011.[55] Likewise, with the police institution, there has never been adequate reporting regarding communication interception carried out in the context of law enforcement. Information about the surveillance practices carried out by the police only refers to the Chief of Police Regulation regarding Interception Procedures in the Police Monitoring Center. The public is only limited to knowing procedures and the fact that the police have sophisticated equipment to carry out wiretapping, and receive much assistance in the form of wiretapping equipment from the Australian government, especially those aimed at the interests of eradicating terrorism.[56] A similar method was also carried out by the Corruption Eradication Commission (KPK), which has surveillance equipment produced by Gueher GmbH from Germany. The KPK also has the ATIS (Audio Telecommunication International Systems), a new generation of Instant Recall Recorders (IRC) in solid state technology, which can be connected to audio sources in the form of GSM/AMPS/CDMA phones or cellphones and will record or tap all voice communication with an active capacity of more than 680 minutes and 1000 different calls. The compression algorithm in ATIS has increased storage capacity.[57]

54). See: "Harga Alat Sadap Kejaksaan RP 16 Miliar", in http://politik.news.viva.co.id/news/read/71543-harga alat sadap kejaksaan rp 16 miliar.

55) See: "Kejaksaan Sudah Bisa Menyadap", in http://www.hukumonline.com/berita/baca/lt4e2557ce07f7f/kejaksaan-sudah-bisa-menyadap.

56) See: "Polisi Selidiki Alat Sadap Densus 88 Dari Australia", in http://www.majalahict.com/berita-3079-polisi-selidiki-alat-sadap-densus-88-dari-australia.html. See also: "Polri Hentikan Sementara Kerjasama dengan Australia", in http://www.antaranews.com/berita/406332/polri-hentikan-sementara-kerja-sama-dengan-australia.

57) See: "Alat Penyadap Ponsel Milik KPK", in http://www.cyber4rt.id/2013/01/alat-penyadap-ponsel-milik-kpk.html.

- Elections Data

The provisions of Article 201 paragraph (8) of Law No. 7 of 2017 concerning General Elections states that updates of voter data are carried out every six months, with reference to population data submitted by the government to election organizers. The results of this data collection then determine whether a citizen has been registered as a voter and has the right to use his vote or not. These data are in the form of names and addresses, including the Identity Number (NIK) and gender (Article 202 of Election Law), and even the Family Card Numbers (NKK) is also included. Worse, the provisions of this law also obliges organizers of the general election to submit copies of voter data to all political parties participating in the election, including NIK and NKK, and without any rules to hide NIK and NKK numbers that can identify or profile personal data.[58] The NIK itself allows immediate identification of area of residence and date of birth. Unfortunately indeed, so far the registration and announcement of voter data are still considered not to have a correlation with the protection of one's personal data, even though the information contained in the voter data in an election also includes personal data, which should not be freely accessible by other parties, even political parties. In addition, the experience of elections in several countries, such as the United States, also shows the operation of Big Data in the winning scenario, especially those related to data driven political campaigns, as mentioned earlier. This situation is increasingly opening up the potential threat of data exploitation related to electoral affairs, including in Indonesia.

- Smart City Program

A number of cities in Indonesia have declared themselves as "Smart Cities" such as Jakarta, Bandung, Surabaya and Makassar. This step essentially emphasizes the optimization of the use of information and communication technology for the benefit of citizen services, development programs, and crime prevention. Through this program, the city government collects large amounts of data from the public, in the form of direct data collection, complaints

58). There was a public pressure on the KPU to keep NIK and NKK confidential, which was responded by the issuance of the KPU Regulation to close the three numbers on the NIK, and four figures on the NKK.

data, and through the installation of a number of surveillance cameras at important points in the city area. In the implementation of the Smart City program itself, especially in Indonesia, all city governments collaborate with private companies (vendors) as operators, who have the technology and knowledge and resources to operate Smart City services. This means that residents' data or even community activities collected from the monitoring system are not directly managed by the government, but are managed by the private sector. This situation is problematic, given the weak regulation of protecting private citizens' data in Indonesia, as well as the protection of privacy in general. Not to mention, the aspect of privacy is also almost certainly never been discussed or the terms of the Smart City service provisions.

In practice, in addition to the utilization of e-government, transportation data collection, and the establishment of a command center, the development of smart cities in Indonesia is also closely associated with the installation of closed-circuit television (CCTV) on a massive scale. For example, the city of Makassar has installed at least 300 CCTV units out of 3000 units planned throughout the city. This number is smaller than the number of CCTVs installed by the DKI Jakarta Provincial Government, which reportedly has installed at least 6000 CCTV units, spread throughout Jakarta. The reason for the massive CCTV installation is usually closely related to the issue of criminal prevention. Unfortunately, the installation of CCTV often does not fully pay attention to the need to protect the rights to the privacy of the citizens. As part of the operationalization of Smart City, a number of cities have also used social media to monitor the activities of city residents. This monitoring is done to see what topics are often discussed by city residents, and all data recorded through the monitoring system are stored on servers managed jointly with third parties. The situation becomes vulnerable due to the lack of regulations that ensure the limitations in the monitoring and management and use of data from such monitoring, even though it is aimed at crime prevention.

- ***Health and Other Social Security Data***

The practice of collecting health data has been carried out for a long time, especially regarding

data that qualify as medical records. This collection is carried out by health workers and health institutions for the benefit of patients' medical records, as well as other institutions for the purpose of controlling disease. This is regulated by Law No. 29 of 2004 concerning Medical Practice, Law No. 36 of 2009 concerning Health, and Law No. 44 of 2009 concerning Hospitals. Almost all relevant laws and regulations have acknowledged and guaranteed that medical record data are one part of personal data that must be protected. In detail, to regulate this matter, the Minister of Health has issued at least three regulations relating to the protection of patients' personal data, namely: (i) Regulation of the Minister of Health No. 269/Menkes/Per/III/2008 concerning Medical Records; (ii) Regulation of the Minister of Health No. 36 of 2012 concerning the Medical Confidentiality; and (iii) Regulation of the Minister of Health No. 55 of 2013 concerning the Implementation of Medical Recording. The three regulations regulate the scope of medical records (patient health data), storage, opening, supervision, qualifications, and the rights and obligations of health workers conducting medical recordings.

However, even though there are guarantees for the protection of health data, on several occasions there have also been allegations of misuse of health data (medical records), either for reasons of ignorance, or for other purposes, such as the marketing needs of certain drugs.[59] Nowadays, health data collection does not only come from medical records, but also health insurance, both managed by the state through the Health Insurance Agency (BPJS Health), as well as those managed by the private sector. For those managed by the private sector, supervision of the protection of consumer personal data is subject to the rules of the Financial Services Authority (Law No. 12 of 2011). Besides through health insurance, collection of personal data is also carried out through other social security programs, such as BPJS Employment, especially now that the prerequisite for obtaining social security is to first have an e-KTP. Thus almost certainly the organizers of the social security system access

59). See: "Beberkan Rekam Medis Disidik Polisi", in https://nasional.tempo.co/read/59094/beberkan-rekam-medis-disidik-polisi.

the entire personal data of consumers/citizens. The potential for big data collection on health issues is increasingly open, especially with the launch of the national e-health program, namely the development of information technology-based health systems, with the main model of the formation of health Big Data.[60]

-Financial and Tax Data

Bank Indonesia has begun to build a Big Data system by collecting customer personal data, including various activities. This big data collection is carried out both directly and through collaboration with a number of institutions, such as online platform service providers. According to Bank Indonesia, all of the data collected will become a large, diverse and complex unit of data. Bank Indonesia as the central bank in Indonesia wants to optimize the use of Big Data to better understand that the monetary decisions and policies issued can be more efficient. This institution initiated the use of Big Data by working with job portal providers. From the data available on all job provider portals, BI can formulate an employment index. In addition to the employment index, several BI pilot projects in using Big Data have produced a proxy for the property market indicators, mapping of behavior and projections of foreign fund flows on the state bonds market, identification of the structure of the relationship between payment system actors and the proxy of bank financing supply and demand indicators.[61]

This is just one illustration of the use of large data by the banking sector, due to the fact that all banking industries have also developed their respective Big Data systems, to develop their businesses and markets. However, unlike several years ago, the supervision of the protection of personal data of banking customers has been carried out quite tightly, especially after the birth of the Financial Services Authority, which was established through Law No. 12 of 2011. At this time every time during the process of recording the customer's personal data, the bank must first seek approval from the data owner regarding the use of his

60). See: "e-health di Indonesia, apa bisa?", in https://inet.detik.com/cyberlife/d-2601038/e-health-di-indonesia-apa-bisa.

61) See: "Bank Indonesia Incar Big Data yang Berisi Data Pribadi", in https://tirto.id/bank-indonesia-incar-big-data-yang-berisi-data-pribadi-cugr.

personal data, including the transfer of those data.

Specifically with regard to taxation, for reasons of enforcing tax obligations, the Directorate General of Taxes is now authorized to access citizens' financial information for tax purposes. This is as regulated by Law No. 9 of 2017 concerning Ratification of Government Regulation in Lieu of Law No. 1 of 2017 concerning Access to Financial Information for the Interest of Taxation into Law. According to this provision, access to financial information for tax purposes includes access to receive and obtain financial information in the context of implementing the provisions of legislation in the field of taxation and the implementation of international agreements in the field of taxation. Based on this authority, the Director General of Taxes has the right to obtain access to financial information from financial service institutions that carry out activities in the banking sector, capital market, insurance, other financial service institutions, and/or other entities categorized as financial institutions. With this authority, of course, the government through the Directorate General of Taxes has a huge opportunity to build a tax Big Data system, with millions of citizens' personal data in it. Without adequate personal data protection, it brings the risk of misuse of private data on citizens.

- ***Transportation Data***

With the greater penetration of Internet in Indonesia, the process of digital transformation has accelerated, especially related to the use of Internet technology to support daily life. One of these transformations gave birth to various types of online-based transportation platform providers, both from outside, such as Uber and Grab, as well as those developed domestically, such as Go-Jek and conventional transportation service providers, which also began to develop online platforms. By offering cheaper fees than using conventional transportation services, people flock to use this transportation service, including those who register as service provider partners. Even the online transportation platform service providers not only provide transportation services, but also other services needed in everyday life, such as shipping goods, delivering food, etc.

The large number of service users certainly has produced huge amounts of data. Not only the

user's movement data from one place to another, but the overall personal data is recorded when choosing to use this service, such as geolocation data. Additionally, the use of this service for other purposes has added a larger variation of data. Collection of transportation data is truly qualified from volume, velocity, variety, and value as characteristics of Big Data or intensive data systems. Even now, there is no firm instrument that guarantees the protection of consumer personal data collected in this practice, including the use of the data, so the opportunity for abuse is also significant. In addition to collecting transportation data from users of online transportation services, the development of smart city programs in various regions in Indonesia has also become a new mechanism for transportation big data collection, with the intention of developing the city transportation system.

- E-commerce and Fintech

Like online banking and transportation, the large number of Internet users in Indonesia has also encouraged the birth of many new businesses (startups), especially those engaged in the e-commerce sector. In particular, the Indonesian government initiated and encouraged the national movement of 1000 digital start-ups, to optimize the use of the Internet to support the digital economy. The large number of startups, especially those engaged in the e-commerce sector, has given birth to new large-scale data mines. When someone shops online, they must enter their personal data, including name, address, telephone number, and e-mail address. When making a payment, if the buyer chooses a credit card payment method, they will also enter their credit card data. Once once shops online at one site, the buyer's personal data is owned by the shop owner or market place. If they regularly shop at the site, then the behavior data of the buyer associated with the purchased product and shopping intensity are also entered into the database of the online store. This also occurs when a person buys a plane or a train ticket from online travel agencies, in which case the data often involve ID card or passport numbers. The travel data will also be recorded, especially if they constantly use one online travel agent in ordering tickets and hotels. Again, adequate protection regulations for mining personal data have not

been available in Indonesia. The sole reference is Ministry of Communication and Information Regulation No. 20 of 2016 concerning the Protection of Personal Data in the Electronic System, which only provides the threat of administrative sanctions in the event of a violation, and does not provide a recovery mechanism for the harmed data owner.

- Social Networks, Telecommunications and the Internet

All statuses, posts and anything shared on various social media platforms, will also be a record of someone. All information about home and work locations, education, employment, and political views can be traced from social media accounts. The collection of data from social media activity is currently commonly used as a consideration by banking or financial companies, to allow credit for someone, including by insurance companies. Furthermore, when we interact using modern communication infrastructure, the content of our communication does not only contain the information we send. We also send data about communication that allows that communication to successfully reach the desired recipient. This is called communication metadata – data about data. Based on the typical definition of metadata, such information about our communications will include:

- home location, for example the home address of the telephone, subscription information, the nearest cell tower;
- devices that transmit or make the communications, for example telephone identifiers, IMEIs from mobile phones, relatively unique data from computers that send messages;
- the time the message was made and sent;
- recipient of communication along with location, device and time of receipt;
- information relating to the sender and recipient of communications, for example, e-mail address, address book entry information, ISP and IP address, and;
- the length of the interaction in a row or the size of the message, for example, how long is the telephone call? How many bytes in the message?

When using the Internet, we leave similar records – the things we search, the pages we visit and the time we spend and the materials we download. Even our actions on social networks – to whom we are connected, what we "like", what ads we respond to – create behavioral records. All of these actions include communication metadata, including data created by our devices, which includes: the exact location of our phone when it is on; our position when our device checks for new emails; our position when our device checks for new social media updates, application updates or other similar automatic checks. This information is continuously made by our devices and especially our smartphones. A newer and broader definition of metadata is that we send metadata continuously, even when we fall asleep, because our device is always in constant contact with several components of the communication infrastructure. Without adequate protection of metadata, we can be traced and described permanently and continuously.

E. THE IMPORTANCE OF PERSONAL DATA PROTECTION REGULATION IN INDONESIA

The extent of big data collection practice also has an impact on the increasing threat of abuse of collected personal data. The threat is increasingly prominent, especially since the government rolled out the electronic KTP (e-KTP) program. In addition to the government's personal data recording program, recording is also carried out by the private sector, such as banks and telecommunications service providers. In relation to this, some time ago we were shocked by the information about the alleged leaking of data of 25 million cellular telephone subscribers. Later, there were even allegations of personal data leakage that emerged in line with the implementation of the SIM Card re-registration policy for cellular phone service users. The situation has raised awareness regarding the poor mechanism of protecting personal data so far. This occurs at least because of three things: *First,* low public awareness in maintaining or protecting personal data, so that they easily distribute or transfer their personal data to other parties. In fact, in general people have not placed personal data as part of property that is worth protecting; *Second,* there is no adequate set of laws to protect personal data, specifically

related to the obligations of data collectors and managers; *Third,* the massive practice of mass data mining, which is carried out by the government and the private sector, either with the knowledge of the data owner or otherwise.

In the context of legislation, the ELSAM study shows overlapping and lack of synchronization of rules regarding the protection of personal data in Indonesia. ELSAM found at least 32 laws with content related to the regulation of private data of citizens. The majority of them give authority to both public (government) and private authorities to collect and manage citizens' personal data, including the authority to carry out intrusions with a few exceptions. The sectors are also varied, ranging from telecommunications, finance and banking, taxation, population, filing, law enforcement, security, to health.

The overlapping of the 32 laws is seen, among other things, in the following aspects: (a) the purpose of processing personal data; (b) notification or approval from the owner of personal data; (c) retention time of personal data; (d) destruction, deletion or alteration of personal data; (e) the purpose of opening personal data to third parties; (f) the licensor to open personal data to third parties; (g) the period of which personal data can be opened to third parties; (h) sanctions for violators of the protection of personal data; and (i) recovery mechanisms for victims whose rights have been violated.

Another weakness of the current laws also appears from the lack of guarantees of recovery for victims whose rights are violated. Only the National Intelligence Law and the Health Law acknowledged the existence of a guarantee for recovery for victims whose personal data was exposed by third parties. In fact, one law should contain provisions regarding privacy of data, but does not have any arrangements, i.e. the Consumer Protection Law. This actually reinforces the fact that the provisions of legislation in Indonesia today are still limited and has not been able to protect the enjoyment of personal data as a whole.

In the case of overlapping arrangements, weaknesses related to the protection of personal data in Indonesia occur due to a lack of understanding of the definition of personal data, as a consequence of the diversity of regulation.

Until now, Indonesia does not have a solid definition of personal data, each law has its own understanding of personal data in accordance with the field of regulation. For example, in the Population Administration Law and Government Regulation concerning Electronic Systems and Transactions Administration (PSTE), personal data is defined in general, as certain data about persons that is stored, maintained and maintained by the truth and protected by confidentiality. However, in other laws, such as the Banking Law, the Sharia Banking Law and the Telecommunications Law, the definition of personal data is specifically intended for the information of depositors and their deposits as well as investors and investments, and records of telecommunication service usage. This conceptual difference regarding personal data will certainly affect the level of security of the data itself. The confusion in personal data protection rules is highly detrimental to the data owner, especially with the era of Big Data or intensive data systems today, when all parties are competing to collect as much personal data as possible, along with the data revolution process.

F. CONCLUSION AND RECOMMENDATIONS

Protection of personal data is a form of respect for the right to privacy, as part of human rights. Although Indonesia does not yet have a *lex specialis* regulation that regulates the protection of personal data, the guarantee of the right to privacy is contained in the Indonesian constitution, as stipulated in the provisions of Article 28G paragraph (1) of the 1945 Constitution, which states: "*Everyone has the right to protection of the person, family, honor, dignity, and property under their control, and has the right to security and protection from the threat of fear of doing or not doing something that is a human right*". A similar assertion also appears in the provisions of Article 29 of Law No. 39 of 1999 concerning Human Rights. Unfortunately, these constitutional guarantees have not been properly implemented at the level of legislation. Indeed, the majority of the public in Indonesia has not made personal data a part of property and human rights that must be protected, so we often find people who

unconsciously but carelessly indulge their privacy, including personal data. Responding to large-scale data collection practices, as a logical consequence of the new era of intensive data systems, which seek to collect various types of data for various interests, it is urgent for the state to prepare a strong data protection regulation for its citizens. This effort can be done through the following steps:

1. The need for a roadmap that describes and integrates various mandates and obligations for the protection of personal data, in every direction for the development of government programs, and various other sectors, so that there are complete guidelines in the process of institutionalizing future policies.

2. The need for an immediate initiative to discuss a law that specifically contains provisions regarding the protection of personal data, in accordance with international human rights law standards.

3. Due to the wide scope of protection of personal data, including entities that perform the collection and storage of data (public and private), it is important to encourage the establishment of an independent body that specifically has the authority to supervise the implementation of these laws effectively.

4. The importance of involving related actors (government, society, business sector), in every policy making process, in an emancipatory space, so that the resulting policies have minimal risk and provide benefits for all actors.

5. The importance of the integration of business and human rights principles in every regulation that regulates the private sector, also encouraging the integration of these principles in corporate operations.

ORGANISATIONAL PROFILES

The Institute for Policy Research and Advocacy (Lembaga Studi dan Advokasi Masyarakat, ELSAM) is a policy advocacy organisation founded in August 1993 in Jakarta. Its purpose is to participate in the attempts to grow, advance and protect civil and political rights, and human rights in general – as mandated by the 1945 Constitution and the 1948 Universal Declaration of Human Rights. Since its inception, the spirit of ELSAM is to develop a democratic political order in Indonesia through empowerment of the civil society through advocacy and promotion of human rights. ELSAM conducts a number of activities in order to achieve its objectives, which consist of:

1. **Mainstreaming Human Rights in Policymaking**

Encouraging the integration of human rights principles in every policy-making process, such as: Providing inputs and recommendations for legislative and government institutions; Preparing critical notes through various policy briefs on draft policies; Monitoring the implementation of legislative functions; Providing technical assistance to government institutions; Conducting various strategic partnerships with various institutions.

2. **Studying and Producing Human Rights Knowledge to Support Policy Advocacy**

ELSAM conducts various research and knowledge production to encourage the formation of evidence-based policies. ELSAM's study focuses include: Business and human rights: Assessing the impact of corporate operations on human rights, including encouraging the creation of the national business and human rights action plan; Internet and human rights: Examining the implications of technological progress on human rights protection, including issues of internet

content governance, personal data protection, communications surveillance, and cybersecurity; Development of human rights friendly regencies/cities: ELSAM is actively involved in the development of Human Rights Friendly Regencies/Cities, by assisting local governments in carrying out the mandate of the fulfillment of human rights.

3. Encouraging the Fair and Dignified Settlement of Past Human Rights Violations

ELSAM has intensively promoted the adoption of a transitional justice approach to resolve past cases of gross human rights violations. The goal is to ensure justice and recovery for the victims, and to prevent repetition, through: Promoting disclosure and truth-seeking policies; Supporting institutional reform, in particular of law enforcement and security institutions; Assisting victim organizations to strengthen advocacy capacity; Representing victims in the litigation process of past human rights violations cases.

4. Human Rights Education

As part of promoting and strengthening the capacity of actors, ELSAM continues to provide human rights education, such as: Human Rights Course for Lawyers, held annually; Organizing trainings with specific topics, such as the handling of serious human rights cases, human rights in conflict areas, freedom of religion/belief, and capacity building for law enforcement officers; Training to promote the use of a rights-based approach for policy makers, including local governments, and business sector.



Privacy International is committed to fighting for the right to privacy across the world.

We investigate the secret world of government surveillance and expose the companies enabling it. We litigate to ensure that surveillance is consistent with the rule of law. We advocate for strong national, regional, and international laws that protect privacy. We conduct research to catalyse policy change. We raise awareness about technologies and laws that place privacy at risk, to ensure that the public is informed and engaged.

To ensure that this right is universally respected, we strengthen the capacity of our partners in developing countries and work with international organisations to protect the most vulnerable. Privacy International envisions a world in which the right to privacy is protected, respected, and fulfilled. Privacy is essential to the protection of autonomy and human dignity, serving as the foundation upon which other human rights are built. In order for individuals to fully participate in the modern world, developments in law and technologies must strengthen and not undermine the ability to freely enjoy this right.

Privacy International, a registered UK charity (No. 1147471), was founded in 1990 and was the first organisation to campaign at an international level on privacy issues.

Privacy International is located at:

62 Britton Street, London
EC1M 5UY United Kingdom



Lembaga Studi dan Advokasi Masyarakat (ELSAM)

Jl. Siaga II No. 31, Pasar Minggu, Jakarta 12510 INDONESIA
Tel.: (+62 21) 797 2662, 7919 2564 Telefax.: (+62 21) 7919 2519
Surel: office@elsam.or.id | Laman: www.elsam.or.id
Twitter: @elsamnews, @elsamlibrary
Instagram: @elsamnews
Fanspage: @elsamjkt