Injuruty: Interdiciplinary Journal and Humanity
Volume 2, Number 5, Mei 2023
e-ISSN: 2963-4113 and p-ISSN: 2963-3397

REFORMULATION OF STATISTICAL DATA SOURCES: BIG DATA NEW DATA SOURCES SUPPORTING FUTURE OFFICIAL STATISTICS ?

Ari Ardiansyah, Amir Ilyas, Haeranah
Universitas Hasanuddin Makassar, Indonesia
Email: ariabps4@gmail.com

Abstract

Big Data is the new oil. The real issue of the development of information technology is the existence of Big Data. Big Data is very important, talking about Big Data not only touches the issue of how big data you have, but what can be done with the data. The high cost of Big Data access is inversely proportional to the collection of data through surveys or censuses for free. analysis of the application of efficiency theory in an economic approach in imposing access costs to statistical activities, knowing the impact of imposing Big Data access fee tariffs, knowing the potential of Big Data as a supporter of official statistics in the future, what are the obstacles and solutions in implementing it. The preparation of this paper is juridical-empirical research with the nature of descriptive research with an economic law approach. Big Data is managed for different purposes using different systems and methods and not necessarily using statistical rules. The implementation of Big Data as a new data source is carried out through a combination of data sources. The imposition of Big Data access fees as a source of data supporting official statistics causes low utilization of Big Data as a new data source. The information technology revolution makes Big Data has the potential to complement, replace, improve, add, and improve the composition of existing statistical data sources, and produce more timely outputs. Difficult access and high costs for data collection are major obstacles. Therefore, it needs to be supported by legal instruments that facilitate its implementation in Indonesia. One is the reformulation of existing regulations that make it easier for basic statistical organizations to obtain such data sources for free and use only in the national interest.

Keywords: reformulation; statistics; big data; official statistics

INTRODUCTION

Today, millions or even billions of bytes of data are generated at unprecedented speeds from heterogeneous sources every day. This is due to the trend of technological development, including the Internet of Things and cloud computing (Botta et al., 2016). This system forms a distributed and resilient system that supports several interconnected systems such as the smart grip system (M. Chen et al., 2014). Health care systems Kankanhalli et al., (2016), retail systems such as Walmart Schmarzo, (2013), and Government systems Sirait, (2016), such as law enforcement security infrastructure (Stoianov et al., 2015).

Technological advances have created a digital revolution in the form of new innovations in the collection, storage, processing, and transmission of large, complex amounts of data in real time (Djafar, 2019). Therefore, the digital revolution is often considered synonymous with the data revolution (Djafar, 2019). These developments have encouraged the collection of a wide variety of data, no longer relying on consideration of what data might be useful in the

future (Djafar, 2019). Data is treated like a tangible asset. This new era of data management is commonly referred to as Big Data (Malik, 2013).

Big Data, a term typically associated with very large and rapidly growing, complex and complex data sets, as well as the diversity of data types collected, processed, and analyzed by organizations (Malik, 2013). Before the Big Data revolution, conventional data collection through surveys or censuses to produce Official Statistics was considered less efficient in terms of time. Conventional data collection takes a long time starting from data collection, data processing, data dissemination to publication. However, with the advent of Big Data technology, there has been a paradigm shift in terms of fast and accurate data availability. Big Data is one of the new data sources that allows quick and accurate decision making.

Big Data emerged from the need for large companies like Yahoo, Google, and Facebook to analyze large amounts of data (Garlasu et al., 2013). Doug Laney describes Big Data in terms of volume, velocity and variety or often referred to as 3V (Kitchin & McArdle, 2016). The development of Big Data technology in Indonesia is increasingly attracting interest not only from the corporate sector but also from government agencies.

The trend of Big Data utilization in government agencies is increasing along with the increasing need for real-time updated data for timely and accurate public policies. Big Data technology is of course used by many organizations, including large, small and medium-sized companies, governments, and statistical organizations (Agusta, 2021).

The Big Data revolution has succeeded in changing the paradigm of data. Big Data as new oil, is now the most valuable source of wealth besides oil. The Economist notes that "A century ago the resource in question was oil. Today, in the digital age, similar concerns are raised by giants dealing with data, oil in the digital age" (Www.economist.com, 2017).

One of the keys to winning the competition between countries is mastering data. There is a strong correlation between a country's development capacity and the availability of quality data. Quality data definitely meets the Quality Assurance Framework (QAF) (Nurseto Wisnumurti et al., 2022).

The real issue of the development of information technology is the existence of Big Data. This factual condition is a challenge in the implementation of statistics. Big Data is very important, because talking about Big Data not only touches the issue of how much data you have, but what can be done with the data. Big data is a reality and has the potential to be used as a source of data in the implementation of statistics in Indonesia. Kitchin, (2015)identified that Big Data can be used to: replace all existing data sources, replace some existing data sources, generate complementary data with different perspectives to complement existing data, increase estimates from other data sources, generate new data (Kitchin, 2015).

Table.1 Mobile network operator (Big Data) data access costs for 2021-2022

Big Data	Cost	Information	Year
Mobile Network Operator Data	30.650.000.000	Nusantara Tourism and Mobility Statistics	2021-2022

Source: DIPA BPS for 2021-2022

BPS asa provider of basic statistics can obtain data in other ways in accordance with the development of science and technology. Setting the phrase is still general in nature, causing BPS limitations to access data sourced from Big Data and the high cost of access to these data sources. During 2021-2022, Rp. 30,650,000,000 is allocated to get mobile network operator data access for statistical activities, the greater the data needed, the higher the costs that must be incurred by BPS. This is inversely proportional to data collection through surveys or censuses carried out free of charge. Where the two data produced are used as Official Statistics for national interests (Sabtiana et al., 2018).

The ideal theory of harmonizing economic and legal aspects is the theory of economic efficiency. The use of efficiency theory is in line with the goal of national economic efficiency as an effort to improve people's welfare through statistical activities. The author is interested in discussing the application of efficiency theory in legal and economic approaches in the imposition of access costs to statistical activities, how is the impact of the imposition of access costs and the difficulty of access to the use of Big Data, what is the potential of Big Data as a supporter of official statistics in the future, what are the obstacles and solutions in implementing Big Data as a source of data supporting official statistics and what is the imposition The cost of access to statistical activities is appropriate in terms of the Economic Analysis of Law approach.

METHOD RESEARCH

The type of research used in the preparation of this paper is juridical-empirical research with the nature of descriptive research. Juridical-empirical research with a descriptive nature is a literature research conducted by examining secondary data. The research was conducted by examining the provisions in laws and regulations and literature related to law formation and examining the socio-legal symptoms with an economic approach using the theory of economic analysis of law (Irwansyah, 2017).

RESULT AND DISCUSSION

Reformulation of Statistical Data Sources: Big Data New Data Sources supporting Future Official Statistics.

The concept of big data focuses more on the characteristics of big data. Doug Laney with his concept of 3V whichis Volume, Velocity, Variety, and there are also those who add other V characteristics, namely exhaustive. Volume relates to the size of very large data storage media; Velocity relates to the Kaur et al., (2018) speed at which data is created, reprocessed, and generalized; variety relates to the type or types of data that can be processed ranging from structured or unstructured data; and exhaustive covers all populations.

Big data is different from conventional data collection (survey / census) or often called Small Data. Based on a review of the definition of Big Data Kitchin, (2015) argues that Big Data differs qualitatively from Small Data in seven characteristics, Lovelace, (2016)see in table 2. While MacFeely, in contrast to Kitchin, claims that "there are 6 characteristics of Big Data or often referred to as 6V, see tabel 3.

Basically, Kitchin and Macfeely have a lot in common, but in current conditions, Macfeely's 6V concept is more relevant to Official Statistics' data business processes. The characteristics of Big Data Value are termed new oil. In addition, the characteristics of Volatility allow data to be collected faster than conventional data sources (census/survey) which allows quick and precise decision making.

Table 2. Conventional Characteristics Vs Big Data

Characteristics	Small Data	Big Data
Volume	Limited to large	Very large
Velocity	freeze-framed/bundled	Fast, continuous
Variety	Limited to wide	Wide
Exhaustivity	Sample	Entire populations
Resolution andindexicality	Course and weakto tight and strong	Tight and strong
Relationality	Weak to strong	Strong
Extensionality and scalability	Low to middling	High

Source: (Kitchin, 2015)

Table 3.Conventional VS Big Data Characteristics

Characteristics	Conventional (survey/Census)	Big Data
Volume	Limited to very large	Very large
Velocity	freeze-framed/bundled	Fast, continuous
Variety	Limited to wide	Wide
Veracity	known to be true	Unknown
Volatility	weakbecome fast	Very fast
Value	very valuable	very valuable

Source: adjustment of the Mac Feely 6v (2018).

Big Data is often associated with data science, data mining, or data processing. However, Big Data involves a larger infrastructure than ever before. There are 4 important elements in implementing Big Data technology in Rumata, (2015) organizations related to statistics , including:

a. Data

The basic description of data refers to objects, events, activities, and transactions that are documented, classified, and stored to give it a specific meaning. Data that has been processed so that it can provide meaning is called information. Data availability is the initial key of Big Data technology. There are some organizations that have a lot of data from their business processes, both structured and unstructured data.

b. Technology

It deals with the tools and infrastructure used to run Big Data, such as computational specialists, mathematicians, and statisticians. BPS will usually not face major technology limitations because BPS can buy technology or work with third parties to obtain the technology.

c. Process

The process of adoption of Big Data technology requires a change in organizational culture. For example, before Big Data, BPS only collected data from Surveys and Censuses as Official Statistics in policy making. But after the existence of Big Data technology, there has been a paradigm shift in the importance of the availability of fast and reliable data to take fast and appropriate policies. One of the new data sources that enable such policy making is through Big Data.

d. Human Resources

The implementation of Big Data technology requires human resources who have analytical and creative abilities, namely the ability to determine new methods that can be used to collect, interpret and analyze data, and computational capabilities (computer programs).

Definition of Official Statistics

Official statistics are statistics produced in the national statistical system. The data are collected within a legal framework, and conform to statistical principles such as independence and objectivity. Statistical concepts are standardized and output requirements are internationally harmonized, and often governed by binding regulations (Braaksma et al., 2018.)

The use of Official statistics is intended to support national development. Meanwhile, according to MacFeely, (2020), the purpose of official statistics is to provide statistical data needed by the government, private sector and society related to economic, demographic, social and environmental conditions (MacFeely, 2020)

Official statistics are based on strong principles. The main foundation of this principle is the protection of data confidentiality. Data collected for statistical purposes (in particular in relation to personally identifiable or corporate data) may not be disclosed and disseminatedin aggregate data form and may only be used for statistical purposes. The principle of maintaining the confidentiality of data provided by respondents is very important in the implementation of statistics. The existence of confidentiality guarantees will encourage respondents' trust in BPS, which will affect the way basic statistics are administered.

Economic Analysis Of Law

Legalquestions can be analyzed using rational choices byapplying economic principles or often called Economic analysis of law (LAKO, 2016). The Economic Approach to Law is a branch that is beginning to grow and is increasingly in demand among scholars. One of these can be found in the legal literature entitled Economic Analysis of Law by Richard A. Posner.

Like economics, the legal system is concerned with rational behavior. Standard analysis begins with the assumption that in deciding to carry out statistical activities, BPS has carried out a rational assessment by calculating the benefits and costs of these statistical activities to maximize the provision of statistical data to the wider community. Thus, when BPS assumes

that the benefits of Big Data sources outweigh the costs of Big Data access, BPS will be more likely to utilize these Big Data sources, but if the costs incurred are greater than the benefits, then BPS will tend not to do so.

In addition, how these microeconomic concepts are applied to legal issues. The concepts of biaya, harga, nilai, utility are very important, including in evaluating a rule.

The impact of the imposition of Big Data access fees and the difficulty of Big Data access on the use of Big Data as a source of data supporting Official Statistics

The presence of Big Data is expected to have a major impact on institutions that produce, process and analyze data and information (De Broe et al., 2021). Band the Statistics Center or often abbreviated as BPS is one of these organizations. The Central Bureau of Statistics in charge of official statistics uses it for various purposes, both for state and private and public interests. Arguably, the way BPS takes data from Big Data will ultimately have an impact on the entire community.

Data plays an important role in strategic decision making, especially in the era of B ig Bata where official statistics are not only derived from surveys and censuses but are required to handle a number of Big Data in various sectors. In fact, it not only provides valuable insights, but also a competitive advantage when supported by technology and organizational resources (Mahrani et al., 2021)

Big Data is generally defined as 3V, i.e. Volume, Velocity and Variety. The volume associated with data size increases as data storage capacity increases. Velocity refers (C. L. P. Chen & Zhang, (2014) to how quickly data can be transmitted in real-time, streamlined, or in batches Assunção et al., (2015) and Variety refers to data mining that can be obtained from various data sources such as social media, the Internet of Things (IoT), or sensors in a structured, unstructured data format (Kitchin & McArdle, 2016).

Along with the development of science and technology, these three characteristics are transformed into: Volume (amount of data), Velocity (data processing speed), Variety (data diversity), Variability (data variability), Veracity (data quality), Validity (data validity), Viscosity (data complexity), Volatility (data volatility), Visualisation (data visualization), and Value (data / utility) or often called 10V (Khan et al., 2018).

Meanwhile, in relation to official statistics, Mac Feely stated that the 6V characteristics of Big Data are very relevant to the business process of official statistical data (MacFeely, 2020)v. In addition to the 3V outlined above, Mac Feely adds three other characteristics: Veracity, Volatility, and Value. The characteristic "Veracity" refers to errors and biases in Big Data. The characteristic of "Volatility" MacFeely, (2020) refers to the rapid changes in technology and business processes that generate Big Data. The "Value" characteristic contains a very valuable value, see chart 2.

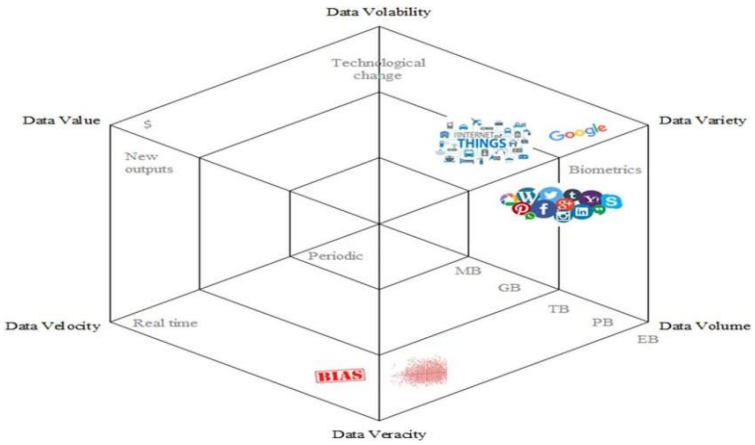


Chart 2

Figure 1 : 6V Characteristics of Big Data Relevant to Official Statistical Data

The existence of big data can be both a challenge and an opportunity for national statistical agencies to produce official statistics. Official statistics abbreviated as NSO as a source of government decision-making and key public information have undergone changes (De Broe et al., 2021). In facing these challenges, NSO must be able to utilize big data and manage it according to procedures. Big Data is expected to be a source of data other than censuses and surveys. if managed properly, Big Data has the potential toreplace official statistics or official statistics that are completely new or as a complement to official statistics Florescu et al., 2014; Piwowar-Sulej, (2021)

Big Data technology has only begun to be known in statistical activities since 2016. BPS utilizes Big Data through web crawling, Google and Facebook mobility, satellite imagery, and mobile phone data in data collection. So far, one of the difficulties in using Big Data is one of the characteristics of Big Data, namely Variety refers to structured and unstructured data formats. Big Data is managed for different purposes, different systems and methods and does not always conform to statistical principles. No Big Data has complete data, so to make it an official statistic requires a combination of various data sources such as censuses or surveys. Currently, the role of Big Data in the Implementation of Statistics as a source of data supporting official Statistics.

Implementation of the 2016 Economic census, a combination method between the use of Statistical Business Register (SBR) as initial data with census data sources. In addition, the implementation of the 2020 census will be combined with census data sources that use the Big Data Population Administration dataset as initial data combined with census data sources. Implementation of Area Sample Framework (KSA) to calculate rice production by combination method using satellite imagery as rice field area data and surveys. The combination method uses Big Data from Mobile Positioning Data (MPD) to collect the number of domestic tourists and length of stay with survey data sources. Data in the form of Google and Facebook Mobility Index to compare people's mobility in various places, such as offices, homes, and grocery stores, to meet daily needs before and after the pandemic to make mobility statistics.

Arrangements related to basic statistical operations can obtain data in other ways in accordance with the development of science and technology. Setting phrases in other ways in accordance with the development of science and technology is still general in nature, causing low utilization of these data sources.

Interms of the aspect of Big Data Utilization Access Costs in statistics is still very limited due to the high cost of Big Data access. In 2021-2022, a budget of Rp. 30,650,000,000 is needed to accessmobile positioning data (MPD), see table 1. In addition, the high cost of accessing these data sources causes the use of paid Big Data in basic statistical activities is still very limited, there is only 1 routine data collection activity each year see in table 4.

Judging from the aspect of the difficulty of accessing Big Data sources, there are only 2 new data sources that can be utilized every year. Only 2 new data sources can be used each year. These data sources are utilized through inter-ministerial cooperation or through free and open access. It is difficult to find open sources of Big Data (0pen access) and of course free for statistical purposes today, one of which is due to the high sectoral egos among data owners.

Table 4. Official Statistics that use Big Data as a Source of Supporting Data

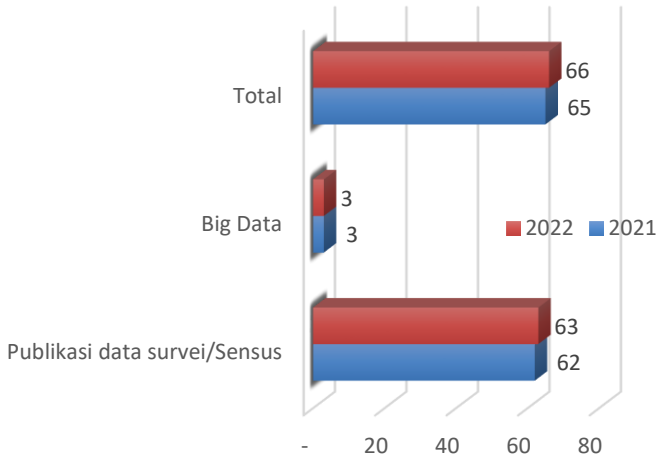
Official Statistics	Data Sources	Year
Economic Census 2016 (SE2016)	Combination Method (use of Statistical Business Register (SBR) and field census)	2016
Tourism statistics to calculate foreign tourists*	Combination Method (Big Data (use of mobile positioning data (MPD) and surveys)	Every Year
Mobility statistics	Combination Method (Big Data (google and facebook mobility index and survey)	Every Year
Food Survey Area Sample Framework	Combination Method (survey and utilization of Big Data from satellite imagery (GIS)	Every Year
Population Census 2020	combination method (Big Data (population administration records) and Census)	2020

When viewed from the utilization of Big Data on an annual basis, the utilization of new data sources is still very low. Thehigh cost of Big Data access and the difficulty of accessing these data sources are the main reasons. In 2021, out of 65 survey/census activities produced only 3 data sources using Big Data, then in 2022 out of 66 publications produced only 3 data came from Big Data, see Table 5. P in 2021-2022 there are only 6 publications or 4.6% sourced from Big Data, the remaining 95.4% comes from censuses/surveys, see in Graph.1. Effective utilization of data is important because it is considered the foundation for an organization (Kaur et al., 2018).

Table.5 Comparison of official statistics derived from Big Data vs Conventional

Publications	2021	2022	amount	Average
Publication of survey/census data	62	63	125	95,4
Publications from Big Data sources	3	3	6	4,6
number of publications	65	66	131	100

Source: the central statistics agency is taken from the output achievements of each surve



Graphs. 1. Conventional VS Combination of Conventional Statistics with Big Data
Source: *BPS Pusat Statistik, BPS census/survey calendar for 2021-2022*

Statistical data provide an objective basis for the true picture of an event. By knowing the situation correctly and as it is, individuals, society and society, as well as state administration can easily determine appropriate, effective and efficient measures to solve various problems. A clear and well-structured picture of reality will provide benefits for the enhancement of intelligent knowledge (Hasbullah, 2023).

Statistical data also helps the government in monitoring and evaluating national development performance. By monitoring statistical data, the government can find out the extent to which development programs have succeeded in achieving their goals and improve programs that have not succeeded in achieving targets.

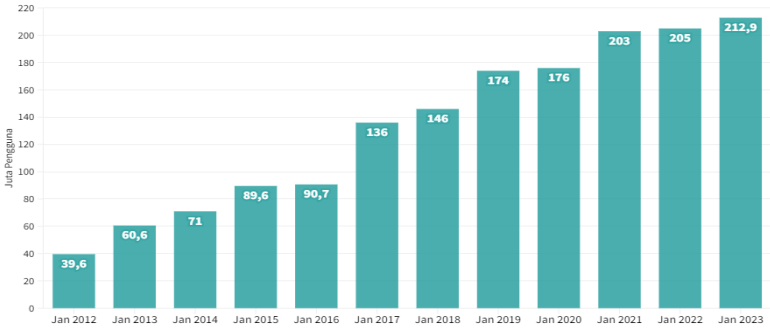
In addition, statistical data is also useful for the private sector to develop relevant products and services and to support business planning. Companies can use statistical data related to public consumption to understand consumer behavior (Hasbullah, 2023). What commodities are favored by the people. What goods are not yet known and are now consumed on a wide scale. Companies can develop business strategies to develop market potential and prepare strategies to win market competition.

Big Data Opportunities as a Source of New Bricks support the Official Statistics of the future.

The information technology revolution causes the need for fast and reliable data is a necessity. One of the new data sources for policy making is Big Data. Big Data as *new oil*. Data is a new kind of wealth, now more valuable than oil.

The era of digital society in Indonesia is marked by the rapid development of information technology (ICT), especially the increasing use of the internet through cellular phones. In 2021, citing data from Indonesia.Id, it was reported that internet users in Indonesia had reached 203 million and became 212.9 million in 2023.Interestingly, the average internet user spends

8 hours 36 minutes a day surfing in cyberspace. Development of internet users in Indonesia January 2012 to January 2023), see graph.2.



Source : DataIndonesia.Id.

Graphs. 2 Growth of internet usage in Indonesia (January 2012-January 2023).

Every day without realizing it, our activities generate a lot of data. We use various online applications toreveal our work. Financial apps, vehicle booking apps, food and beverage booking apps, online meeting apps, online shopping and a variety of other apps all require internet and access to personal data. These activities are then stored as data, and will continue to grow to form Big Data. Big Data can be turned into information that can help users to get a lot of input in determining strategic policies in a government agency or company.

The fundamental difference between the application of statistics in Big Data and conventional data systems (surveys and censuses) is that in conventional systems, data will be published through planning from sampling, data collection, analysis, to dissemination. Whereas in Big Data, data is available, created, and continues to grow all the time so that what is needed is the ability for what the data is used for. Differences in the analytical capabilities of conventional data collection (survey/census) with Big Data, see table 6.

Table 6. Differencesbetween conventional data analysis and Big Data analysis data

Conventional Statistical Analysis	Data Analytics
Confirmative	Explorative (predictive)
Small data set	Large data set
Small number of variables	Large number of variables
Deductive (no predictions)	Inductive
Numeric data	Numeric and non-numeric data
Clean data	Data cleaning

Source: Data Mining and Statistics: What are the Connections? (Jerome Friedman, 1997)

Due to the rapid development in the era of Big Data, the need for fast data and real-time statistical information is increasingly needed, especially as a basis for public policy, or in other words as a new source of supporting official statistics. Statistics Center often abbreviated as BPS as one of the official statistical data providers has implemented big data in a number of activities, including the use of satellite imagery for Area Sample Framework, commuter mobility, tourism statistics, and other statistical activities. Utilization of Big Data through the method of combining Big Data data sources with census or survey data sources to support official statistics. Currently, the use of Big Data as a new data source is still limited and will increase in the future. This is in line with the mission of the UN statistical committee Global

Working Group (GWG) related to global programs including the use of Big Data for official statistics.

Furthermore, the existence of statistical activities is influenced by BPS's ability to utilize Big Data in the future. The impact of rapid technological progress will cause new problems, therefore fast and appropriate action is needed. Policy making should be based on the timely and accurate availability of data, one of the data sources is Big Data. Big D or a technology that enables the processing, storage, and analysis of data in various forms or formats. In addition, Big Data can be used to increase large volumes of data and transfer data quickly in a relatively short time. Big Data has characteristics, namely: 1). A much larger number (volume), 2). Data is transmitted at a very high velocity, 3). Data formats are variegated.

The rapid development of information technology makes big data a provider of dynamic new data sources that can complement, replace, expand and complement and improve the composition of existing statistics, as well as produce more timely output data (Kitchin, 2015). Florescu et al. are cited by Kitcin in their journal, which clarifies that big data sources can be used in five ways in current statistical systems:

a. To completely replace existing statistical sources (existing statistical output);
b. To partially replace existing statistical sources (existing statistical outputs);
c. To provide complementary statistical information in the same statistical activity but from different perspectives(additional statistical output);
d. To improve estimates from statistical sources (including surveys) (better statistical output);
e. To provide completely new statistical information in a particular statistical domain (new alternative statistical output) (Kitchin, 2015).

The potential for the use of Big Data as a source of new data supporting Official Statistics will be even greater in the future due to, among others:

1. The utilization of Big Data in the implementation of new statistics is around 4.6%, indicating that the potential for the use of Big Data as a new data source is still large and wide open. Of course, this must be supported by all parties, especially from the legal aspect which makes it easy for statistics organizers to get access and of course for free to Big Data sources in the government and private and used only for official statistics for national interests.
2. The United Nations Economic Commission for Europe (UNECE) classifies three main sources of Big Data: social networks, traditional business systems, and the Internet of Things. Government business processes and private institutions can also generate data. This type of Big Data group can be utilized optimally in the future as a new source of data.
3. The mission of the UN statistical committee is the Global Working Group (GWG), which provides direction, analysis, and strategic coordination for global programs including the use of Big Data for official statistics.
4. The single data system in Indonesia, which is gaining popularity in government data governance, is predicted to improve government data governance. One Data aims to overcome the occurrence of data differences between data guardians. The One Data concept

is a Government initiative to accelerate the process of data-based policy making. Data obtained by a government or organization must be accurate, complete, and interoperable in order to be easily distributed to data users. With this system, it is expected that data collection will be easier, especially data related to Big Data.

5. The rise of data leakage cases that occur, some members of the public are worried about the security and confidentiality of data collected by BPS, especially those related to respondents' personal data. If this concern continues in the long run, respondents will feel hesitant to provide actual data. Thedevelopment of technology has caused changes in people's behavior, especially urban communities, towards privacy. The protection of privacy causes people to be closed to social interactions, including sharing personal data directly with data collection officers. Thecommunity will be very active in cyberspace. Big Data is one of the new alternative data sources related to population data because applications that use the internet will access personal data.
6. The implementation of the Electronic-Based Government System (SPBE) or better known as e-government, has an impact on system integration. These activities are then stored in the form of data, and will continue to grow to form Big Data. Big Data can be converted into information that can help users to get a lot of input in determining strategic policies in a government agency or company
7. Kennedy et al., (2007) Technology disruption changes BPS business processes in termsof data collection mode. Data collection through PAPI is time-consuming, and expensive. The eraof information technology requires fast, precise, cheap and technology-based data such as CAPI and CAWI. The implementation of both modes requires a combination of methods between data sources, one of which is through Big Data. For example, the implementation of the 2020 population census using CAPI and CAWI modes by utilizing Big Data can shorten the population census data collection time by 10-15 minutescompared to using conventional PAPI modes and at a lower cost, no printing costs, and shipping uesionary as well as data entry fees.

Currently, the use of Big Data in statistical activities is only at the stage of additional statistical output and as a source of new data supporting official statistics has not yet reached the stage of replacing statistical data sources. This is because its implementation has encountered obstacles. Some obstacles in the adoption of Big Data technology, especially in the implementation of statistics, include:

1. Data availability

   One of the keys to making Big Data a new data source is data availability. Currently, Big Data owners and managers are spread across various sources ranging from the government to private parties. Access to data sources requires a lot of effort and cost as permits and licenses are required to legally access non-public data. This is especially true because there are still sectoral egos between agencies as data owners. The data collection phase is the main bottleneck phase.
2. Data standardization

Big Data is managed for different purposes using different systems and methods that are not necessarily in accordance with statistical standards. No Big Data has complete data, to make it an official statistic requires a combination of data from various data sources. currently the use of new Big Data as supporting data for Official Statistics.

3. Infrastructure
   a. A very large volume of Big Data requires a very large data storage infrastructure.
   b. The speed of Big Data is very fast requires the speed of data collection, storage, and processing so that the resulting analysis results do not lose momentum.
   c. Diverse Big Data requires specialized expertise to benefit from a wide variety of highly diverse content.
4. Study on the Utilization of Big Data as a new data source

   There is no comprehensive and continuous study of ideal statistical methods related to Big Data. This was seen during the Covid-19 outbreak. When there are social restrictions that force data collection without face-to-face. To ensure that official statistics can continue to be produced on schedule, there has been no review of the best statistical methods that can be used when comparing data collection with face-to-face.
5. Data security and confidentiality levels

   According to the Microsoft Digital Defense Report 2021, government data is one that is at risk of leakage and misuse. The case of leakage of 1.3 billion card registration data. Then, in 2021, the case of leaking 279 million BPJS Health user data in May 2021. Leak of 1.3 million e-HAC application user data in August 2021. BPS as a government agency that provides statistical data in Indonesia, is very risky to be the target of hacking because it acts as a data center in charge of collecting and processing national data.

Big Data Solutions **as Official Statistics**

1. Reformulation of regulations that make it easier for official statistical institutions to manage and access new data sources, namely Big Data free of charge (free). The Personal Data Protection Law that has been passed already regulates the transfer of personal data but is still general, regulations are needed that specifically regulate the transfer of Big Data for the benefit of government policies, especially statistical institutions.
2. Optimizing the role of BPS as fostering sectoral and special statistics organizers, especially Big Data managers to improve the quality of the data produced. Development starts from government agencies that manage Big Data such as the Director General of Population and Civil Registration, BPJS Health, BPJS Employment, Directorate General of Taxes, Education Office, Government Hospitals.
3. Infrastructure development that supports Big Data Implementation, provision of large storage capacity, data processing tools and building human resource capacity that has data science capabilities.
4. Internally, BPS needs to be encouraged to review statistical methods that utilize Big Data for official statistics on an ongoing basis.
5. Increased Security and confidentiality of data.

Strengthening the protection of confidentiality and security of respondents' personal data starting from the stages of data collection, data processing and data dissemination. Formation of a work team tasked with auditing data security systems and collaborating between institutions, especially with the National Cyber Code.

The imposition of access fees to statistical activities is viewed from the approach of the Economic Analysis Of Law Theory

Nikolas Anova in Statistical Thinking in the Era of Big Data states that the new era of information technology changes everything for statistics. An era where the flow of various types of data (velocity) is very high and fast, resulting in very large data (volume) with high variety or what is called the era of Big Data.

The United Nations Economic Commissions for Europe (UNECE) classifies three main sources of Big Data: social networks, business systems, and the internet of things. Government business processes and private institutions can also generate data. This type of Big Data group can be utilized optimally in the future as a new source of data.

First, a social network is a social structure formed of nodes intertwined with one or more specific types of relationships such as values, visions, ideas, friends, offspring, and so on. the number of active social media users in Indonesia reached 191 million people in January 2022, becoming a potential source of data for statistical activities.

Second, a study conducted by Frost & Sullivan states that online businesses in Indonesia are growing by 17% per year. This growth is due to: No need for large capital, flexible sales time, wider market reach, and easier service. The data can be used to support economic statistics.

Third, the Internet of Things was first proposed by Kevin Ashton in 1999 and first recognized through the MIT Auto-ID Center. Currently being hotly discussed by several parties. The Internet of Things is also starting to be known in Indonesia and began to be used to support various daily activities to connect many objects both physically and virtually through the internet. The Internet of things can also be used by governments through e-Government to manage government and public service delivery more effectively.

Rapid technological development requires the availability of up-to-date, accurate and fast data. One such source comes from Big Data. Here's a comparison between survey, census and Big Data data, see table 7.

Tabel 7 Comparison of Surveys, Census Data and Big Data

Indicators	Survey Data	Census Data	Big Data
Specifications	Statistical Products specified Ex-aunt	Statistical Products specified Ex-aunt	Specified statistical products ex-post
Purpose	Designed for statistical purposes	Designed for statistical purposes	Organic or designed for other purposes
By-products	lower potential by-products	lower potential by-products	higher potential by-products
Method	Statistical methods	Statistical methods	Haven't used the static method yet
Data structure	Structured	Structured	structured and unstructured

Representation	representativeness and coverage according to design	All representation and scope	Representation and coverage are difficult to assess
Data bias	Possible bias (sampling and non-sampling bias)	possible bias, non sampling bias	Unknown
Data errors	Common errors, sampling and non-sampling	Non sampling errors	Non sample error (error reporting)
Timeliness	medium depending on the number of samples	slow	Potentially faster
Cost	medium depending on the number of samples	Expensive	can be more expensive or cheaper even free
Demogfrafi	Featured hotels	all	Service users
Intellectual property	Government/BPS	Government/BPS	Government/private/copyright users

Source : Comparison: survey, census and Big Data.

The data revolution can be an input for strategic policy decisions in looking at a problem. The size and progress towards the sustainable development agenda, in today's digitally connected world, will depend on our ability to look at new sources of data in terms of time and innovative technologies to inform policy formulation. The data revolution leverages existing and new data sources to fully integrate statistics into decision-making, encourage open access to data use, and ensure increased support for statistical systems.

The Center for Statistics is a non-departmental institution that is obliged to provide basic statistical data used for its use intended for broad interests, both for the government, private sector, and society, characterized by cross-sectoral, national-scale, and macro characteristics. As a government agency that provides official statistical data in Indonesia. BPS Pusat Statistik or often abbreviated as BPS conducts data collection activities, both in the form of censuses and surveys. Furthermore, technological developments have an impact on the development of new data sources which are often called Big Data.

Big Data is termed as a kind of "oil mine" for the future that will affect many people. Big Data analytics does not always rely on internal data. Big Data adopts different infrastructure methods and systems. Philosophically, the use of Big Data reduces costs to be cheaper, but in its development, the use of Big Data is actually expensive because most of these new data sources are paid.

The Law on Statistics regulates the implementation of basic statistics, in this case BPS can obtain data in other ways in accordance with the development of science and technology. Setting the phrase is still general in nature, causing BPS limitations to access data sourced from Big Data and the high cost of access to these data sources. The use of Big Data as a new source of data in statistics is limited.

The purpose of law is also expediency in addition to legal certainty and justice. Law has many aspects, one of which is the Economic aspect. Legal weaknesses in the development process only look at the present with the approach of the past and do not see the future. Instead, economics sees the future of legal policy.

Economic law analysis is important because law is a dynamic process or is Law in the making (Satjipto Raharjo). Law is not just about maintaining the existing conditions of society

but continuing to seek and discover about the efficiency and effectiveness of laws that work in society.

Legal issues continue to be objects constellated (organized, created, linked) with fundamental economic ideas. The idea is to be able to place legal issues so that the flexibility of legal analysis (rather than economic analysis) can be developed more thoroughly. An area of study known as "legal economic analysis" focuses on how economic principles can be used to address real-world legal problems.

Judging from the cost of respondents percomparison of conventional data collection (census or survey) with Big Data, the use of mobile positioning data (MPD). The cost of Big Data access to the use of mobile positioning data (MPD) is very expensive where the collection of survey / census data for official statistical purposes by BPS is carried out free of charge. See table 8

Table 8.
Comparison of Costs and Books Publication of Big Data &; Survey/Census sources

Years	Big Data Respondents		Respondents from Surveys/census	
	Cost (Rp)	Publication Price (Rp)	Cost (Rp)	Publication Price(Rp)
202	15.400.000.000	30.000	0	30,000 to 300,000
2022	15.250.000.000	30.000	0	30,000 to 300,000

Source: DIPA BPS for 2021-2022 & PerBPS Head Regulation No.7 of 2015.

Judging from the price of statistical publication books sourced from paid and free Big Data. Theamount of both publications is only Rp. 30,000.-. and both are grouped as non-tax state income (PNPB).

Furthermore, in terms of state revenue, in 2021, the total state revenue for publication sales of Rp. 6,568,494,459 is not comparable to the cost of accessing Big Data sources, the use of mobile positioning data (MPD) required of Rp. 15,400,000,000 in 2021.

Judging from the utility (usefulness) of data collection sourced from censuses / surveys and Big Data both the use of data is used for statistical activities. The use of this data is for broad purposes, both for the government, the community, and the private sector in supporting national development. See tables 9 & 10.

Tabel 9.Comparison of Big Data Access Costs Vs Publication Prices for 2021-2022

Year	Publication of Foreign Tourism Statistics			
	Sum	Big Data access fees*	Softcopy (Rp)	Hardcopy (IDR)
2021	1	15.400.000.000	30.000	30.000
2022	1	15.250.000.000	30.000	30.000

Source: Band Statistics Center, DIPA BPS &; Perka No.7 Year 2015 Based on DIPA Ceiling

Table 10. Cost and price comparison of Paid and Free Big Data publications 2021-2022

Years	Big Data paid	Big Data for Free

	Cost (Rp)	Publication Price (Rp)	Cost (Rp)	Publication Price(Rp)
2021	15.400.000.000	30.000	0	30.000
2022	15.250.000.000	30.000	0	30.000

Source: Badan Pusat Statistik&; Perka BPS No.7 Year 2015

In economics, the level of demand can be influenced by prices, where if prices are high then demand will decrease, and vice versa. When economic principles are used to analyze law, demand is used and price is access to data. So if the price of a data source is high, the utilization rate of the data will decrease. The imposition of data access fees on official statistics is inefficient and reduces the usefulness of the statistical data. BPS as the sole authority in the administration of statistics must be given the convenience to access new data sources free of charge for official statistics purposes only for national purposes. In 2021-2022, data utilization from Big Data averaged 4.6% or only 6 official statistics publications whose data was sourced from Big Data, see table 5.

Rationally, the imposition of fees for official statistics is not appropriate because basic statistics as official statistics concern the public interest and national interest.

When viewed from an economic approach in terms of efficiency, the imposition of costs to utilize Big Data data sources causes inefficiencies where in the implementation of statistics so far free of charge or free resulting in limited use of these new data sources in the implementation ofbasic statistics.

In the era of the industrial revolution 4.0, the provision of statistical information based on economic, social, political, cultural, and natural resources is increasingly needed, because development planning is increasingly widespread, diverse, and becomes the basis for determining the direction of national development policies. Statistical data is essential for decision-making and policy-making. Decision making and policy making can be done more simply, quickly, and accurately with statistical data.

The current data revolution has fundamentally affected the stages of the public policy decision-making process. Data related to statistics play a role in almost all stages of the public policy decision-making process. In the early stages of initiation and formulation of public policy, the role of statistics is very high as inputs which can be in the form of statistical information about socio-economic conditions or through research. The role of statistics is also very high at the stage of monitoring and evaluating the implementation of public policies that have been established at the policy selection stage. Even at the policy selection stage, not only statistical information is required, but statistics can also be utilized

It is necessary to reformulate regulations on statistics to accommodate the use of Big Data into statistical activities. Reformulation is needed for more specific (special) arrangements for the explanation of phrases in other ways in accordance with the development of science and technology in the Law on Statistics which regulates the use of Big Data as a new source of data in the implementation of statistics. The principle of legality is needed to provide legal certainty in statistical providers that make it easier for BPS to be able to access all new data sources in the form of Big Data in the government and private sector for free, so

that the implementation of statistics runs more efficiently and maximum utilization of the data for offical statistics purposes.

CONCLUSION

MacFeely's 6V characteristic concept is relevant to the business process of official statistical data. Big Data is managed for different purposes using different systems and methods that do not necessarily use statistical rules. The absence of Big Data that has complete data to make it an official statistic requires a combination of data from various data sources. Implementation of the 2016 Economic census, a combination of data between the use of the Statistical Business Register (SBR) and the field census. The use of Big Data in statistics is still very limited due to the high cost of access to obtain the data.

The information technology revolution causes the need for updated, fast and reliable data is a necessity. One of the new data sources that enable such policy making is through Big Data. Big Data as a provider of new dynamic data sources, has the potential to complement, replace, improve, and augment and improve the composition of existing statistics, as well as produce more timely outputs.

Access to data sources requires considerable effort and cost due to the need for permits and licenses to legally access non-public data. This is especially true because there are still sectoral egos between agencies as data owners. The data collection phase is the main bottleneck phase. Reformulation of regulations related to new data sources in statistical providers urgently needs to be carried out immediately to support new data sources in statistical operators. The technological revolution has changed the paradigm about data. The availability of fast, accurate data is needed for policy making, one of which is Big Data. Regulatory support needs to be encouraged for the implementation of statistics to obtain these new data sources with access to Big Data in the government and private sectors with young and free costs and only for official statistics to be used for national interests.

REFERENCES

Agusta, H. (2021). Keamanan dan Akses Data Pribadi Penerima Pinjaman Dalam Peer to Peer Lending di Indonesia. *Krtha Bhayangkara*, *15*(1).

Assunção, M. D., Calheiros, R. N., Bianchi, S., Netto, M. A. S., & Buyya, R. (2015). Big Data computing and clouds: Trends and future directions. *Journal of Parallel and Distributed Computing*, *79*, 3–15.

Botta, A., De Donato, W., Persico, V., & Pescapé, A. (2016). Integration of cloud computing and internet of things: a survey. *Future Generation Computer Systems*, *56*, 684–700.

Braaksma, B., Daas, P., Offermans, M., Puts, M., & Tennekes, M. (n.d.). *Big Data and official statistics: local experiences and international initiatives*.

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

Chen, M., Mao, S., & Liu, Y. (2014). Big data: A survey. *Mobile Networks and*

Applications, *19*, 171–209.

De Broe, S., Struijs, P., Daas, P., van Delden, A., Burger, J., van den Brakel, J., ten Bosch, O., Zeelenberg, K., & Ypma, W. (2021). Updating the paradigm of official statistics: New quality criteria for integrating new data and methods in official statistics. *Statistical Journal of the IAOS*, *37*(1), 343–360.

Djafar, W. (2019). Hukum perlindungan data pribadi di indonesia: lanskap, urgensi dan kebutuhan pembaruan. *Seminar Hukum Dalam Era Analisis Big Data, Program Pasca Sarjana Fakultas Hukum UGM*, *26*.

Florescu, D., Karlberg, M., Reis, F., Del Castillo, P. R., Skaliotis, M., & Wirthmann, A. (2014). Will 'big data'transform official statistics. *European Conference on the QualityofOfficial Statistics. Vienna, Austria*, 2–5.

Garlasu, D., Sandulescu, V., Halcu, I., Neculoiu, G., Grigoriu, O., Marinescu, M., & Marinescu, V. (2013). A big data implementation based on Grid computing. *2013 11th RoEduNet International Conference*, 1–4.

Hasbullah, J. (2023). *Tangguh dengan statistik: akurat dalam membaca realita dunia*. Nuansa Cendekia.

Irwansyah, I. (2017). based Environmental Law: The Debate Between Ecology Versus Development. *Sriwijaya Law Review*, *1*(1), 44–66.

Kankanhalli, A., Hahn, J., Tan, S., & Gao, G. (2016). Big data and analytics in healthcare: Introduction to the special section. *Information Systems Frontiers*, *18*, 233–235.

Kaur, P., Sharma, M., & Mittal, M. (2018). Big data and machine learning based secure healthcare framework. *Procedia Computer Science*, *132*, 1049–1059.

Kennedy, R. E., Cohen, W. B., & Schroeder, T. A. (2007). Trajectory-based change detection for automated characterization of forest disturbance dynamics. *Remote Sensing of Environment*, *110*(3), 370–386.

Khan, N., Alsaqer, M., Shah, H., Badsha, G., Abbasi, A. A., & Salehian, S. (2018). The 10 Vs, issues and challenges of big data. *Proceedings of the 2018 International Conference on Big Data and Education*, 52–56.

Kitchin, R. (2015). The opportunities, challenges and risks of big data for official statistics. *Statistical Journal of the IAOS*, *31*(3), 471–481.

Kitchin, R., & McArdle, G. (2016). What makes Big Data, Big Data? Exploring the ontological characteristics of 26 datasets. *Big Data & Society*, *3*(1), 2053951716631130.

LAKO, A. (2016). Revisi RUU Pengampunan Pajak. *Revisi RUU Pengampunan Pajak*.

Lovelace, R. (2016). Book Review: The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences, by Rob Kitchin. *Journal of Regional Science*, *56*(4), 722–723.

MacFeely, S. (2020). Measuring the sustainable development goal indicators: An unprecedented statistical challenge. *Journal of Official Statistics*, *36*(2), 361–378.

Mahrani, S., Pasi, I. D., Mutmainnah, A. K., Samosir, S. W. P., & Gunawan, I. (2021). Proses Pembangunan Smart City Di Indonesia Menggunakan Metode Big Data Analytis Dalam Penerapan E-Commerce. *Media Jurnal Informatika*, *13*(2), 57–63.

Malik, P. (2013). Governing big data: principles and practices. *IBM Journal of Research and Development*, *57*(3/4), 1.

Nurseto Wisnumurti, M., Umum, P. J. A., Bambang Nurcahyo, S. E., Penyelenggara, M. M. K., Sulistiadi, Y. A., Nasrudin, S. S., Muchlisoh, M. E. D. S., Agustina, N., Hardiyanti, R., & Wahyuni, S. S. T. A. (2022). *Pengarah: Dr. Erni Tri Astuti, M. Math. Penanggung Jawab Akademik: Dr. Hardius Usman, M. Si. Penanggung Jawab Keuangan: Ir. Titik Harsanti, M. Si. Penanggung Jawab Kemahasiswaan: Ir. Agus Purwoto, M. Si.*

Piwowar-Sulej, K. (2021). Human resources development as an element of sustainable HRM – with the focus on production engineers. *Journal of Cleaner Production*, *278*, 124008. https://doi.org/https://doi.org/10.1016/j.jclepro.2020.124008

Rumata, V. M. (2015). *Media and Telecommunication Policy and Governance in Indonesia towards Convergence*. The University of Melbourne.

Sabtiana, R., Yudhoatmojo, S. B., & Hidayanto, A. N. (2018). Data Quality Management Maturity Model: A Case Study in BPS-Statistics of Kaur Regency, Bengkulu Province, 2017. *2018 6th International Conference on Cyber and IT Service Management (CITSM)*, 1–4.

Schmarzo, B. (2013). *Big Data: Understanding how data powers big business*. John Wiley & Sons.

Sirait, E. R. E. (2016). Implementasi Teknologi Big Data Di Lembaga Pemerintahan Indonesia. *Jurnal Penelitian Pos Dan Informatika*, *6*(2), 113–136.

Stoianov, N., Urueña, M., Niemiec, M., Machnik, P., & Maestro, G. (2015). Integrated security infrastructures for law enforcement agencies. *Multimedia Tools and Applications*, *74*, 4453–4468.

Www.economist.com. (2017). *The world's most valuable resource is no longer oil, but data*. https://www.economist.com/leaders/2017/05/06/the-worlds-most-valuable-resource-is-no-longer-oil-but-data

Copyright holders:
Ari Ardiansyah, Amir Ilyas, Haeranah (2023)
First publication right:
Injurity - Interdiciplinary Journal and Humanity

This article is licensed under a Creative Commons Attribution-ShareAlike 4.0 International