

International Seminar on Big Data “Building Pathways for Policy-Making with Big Data” Bali, 26 July 2018

Opening remarks by Erwin Rijanto, Deputy Governor, Bank Indonesia

Honorables:

- Bapak Fadhil Hasan, Ketua Badan Supervisi Bank Indonesia,
- Mr. Bruno Tissot, Head of IFC Secretariat, Bank for International Settlements,
- Mr. Gabriel Quiros, Statistics Department, International Monetary Fund,
- Our esteemed chairs and speakers,
- Distinguished guests, ladies, and gentlemen,

Good morning to all of you,

It is my pleasure to welcome you to the International Seminar on Big Data, “Building Pathways for Policy-Making with Big Data”, co-organized with our wonderful colleagues at the Bank for International Settlements – Irving Fisher Committee for Central Bank Statistics.

I would like to express my sincere gratitude to our prominent speakers who have travelled from around the world to gather here in Bali and share their valuable knowledge and experience to us.

Let me begin by remarking that it would be somewhat understated to say that the Big Data revolution has just begun. It has been a decade since the Google Flu Trend was publicized, from which we learned that Internet search data could predict disease outbreak faster than careful analysis by experts. It has also been a decade since two brilliant fellows at MIT first launched The Billion Prices Project, a robust alternative to the long-established methodology for measuring consumer prices inflation using data scraped from online retailers’ website. We are already quite far into the era of Big Data, and it is up to us to make the best use of it.

Although it was first developed and adopted by IT and digital companies, **Big Data tools and methodologies have made their way into public institutions as well.** As of July 2018, there are almost 200 projects listed in the Big Data Project Inventory compiled by the World Bank and the United Nations¹. Very comprehensive though it is, I am sure the list is not complete and there are many other public sector Big Data projects that are being developed in countries all over the world.

¹ <https://unstats.un.org/bigdata/inventory/>

Central banks are not missing out from this surge in Big Data adoption as well. I am delighted to realize that the share of central banks who have incorporated Big Data analytics into their policy-making and supervisory processes has gone up significantly, from 30% in 2015² to almost 60% in 2017³. Even though Big Data implementation presents various challenges, as I will discuss later on, this wider adoption shows that Big Data has proven its merits for us policy-makers.

Distinguished Guests, Ladies, and Gentlemen,

Let us now explore **why institutions are racing to adopt Big Data** in the first place, mentioning relevant examples along the way.

We observe **that there are at least three key factors that drive the widespread adoption of Big Data. The first and perhaps most significant one is the ubiquitous recording of our activities in digital format.** This is closely related to the increased mobile phone and Internet penetration and the recent trend towards digital commerce and interactions. With the cost of data storage and computing power continuously declining over the past decade, thanks in part to the development of cloud services, it has now become very in-expensive for companies and institutions to log and store the data of all their transactions and activities. These massive and granular sets of records provide a gold-mine equivalent for decision-makers from which they can draw insights and base decisions. E-commerce, fintech, social media, and all kinds of companies and institutions play an active part in producing, distributing, and consuming this explosion of digital data.

The next important factor that drives Big Data adoption is the various shifts in data analytical paradigms. Conventional data and econometric analysis have usually been applied on aggregated datasets and time series. Now, we are seeing some major change **from aggregated analysis towards analytical methods that depend on granular/large datasets**, often individual, transaction-by-transaction, or tick-by-tick data. Examples would be the Billion Prices Project that I mentioned earlier, and the AnaCredit⁴ initiative by the European Central Bank, which allows individualized analysis of lending trends and behavior.

The high level of granularity in the data allows us to reveal interesting patterns and behaviors of economic agents. One of the source that readily accessible to central banks is the payment system settlement data. By applying network analysis and entity-matching algorithms, we are able to discover core-periphery structure in interbank payments from our own Real-Time Gross Settlement system. In addition, we can identify flows of funds both within and between groups of corporations. The overall analysis opens up new opportunities for systemic risk assessment.

Besides aggregate-to-granular analysis, we are also seeing **more applications of predictive analytics in addition to descriptive statistics.** Big Data offers added-value that allow us to base our predictions on richer, more granular and more varied types of data, including unstructured data such as texts and images. Big Data is also

² IFC Report on "Central banks' use of and interest in big data", 2015

³ Central Banking and BearingPoint's Joint Survey on "Big Data in Central Banks", 2017

⁴ https://www.ecb.europa.eu/stats/money_credit_banking/anacredit/html/index.en.html

well-suited for **nowcasting**, “predicting” data for the current time period using alternative datasets. Big Data⁵ can help alleviate the problem of **data lag** through higher-frequency collection of granular, publicly accessible datasets and subsequent processing on top of streaming and other real-time Big Data technologies. **Bank Indonesia is also developing and has benefited from this approach for measuring the trends in job vacancy, secondary property, and used-car markets**, for which official statistics are published with long time lags or are not available with the desired level of detail.

Finally, it feels incomplete to talk about reasons for Big Data adoption without alluding to **Artificial Intelligence and machine learning**. The recent development and subsequent boom of specialized machine learning algorithms called deep learning⁶ have allowed computers to see and discern images and videos, understand and generate human-language texts and speeches, drive cars and control robots, and perform a plethora of other tasks nearly as well as human do. With its potential of automating manual human labors, machine learning will certainly impact employment and the general economy in the years to come.

Ladies and Gentlemen,

How could central banks join the broader industry in this issue? One example would be the application of **text mining**: the automated analysis of text data.

Perhaps the most widely cited and widely replicated application of text mining for economic analysis is the **Economic Policy Uncertainty Index**⁷. It has long been believed that high policy uncertainty undermines macroeconomic performance. The index developed by Baker, Bloom and Davis (2016) provides an important advancement in the area of measuring policy uncertainty. Through machine-reading of newspaper articles, an index that measures policy-related economic uncertainty can be constructed. This new approach has been followed by many central banks, including Bank Indonesia, and several robustness checks has shown that the index well correlated with important events that might affect this uncertainty, despite some shortcomings.

In recent years there were also some number of text mining researches that went into **understanding central banks’ statements**⁸. As we know, these statements are an integral part of our policies and being able to extract information in a quantitative way may provide decision-makers and central banks themselves with useful insights on past and future policy decisions and their impacts on the economy.

Distinguished Guests, Ladies, and Gentlemen,

Let us now turn to building pathways for Big Data in our policy-making. There are numerous obstacles that hinder Big Data implementation in central banks and government institutions, as have been brought into attention in previous conferences and surveys. The most common challenges include lack of adequate support of Big

⁵ https://www.ecb.europa.eu/pub/conferences/html/20140407_workshop_on_using_big_data.en.html

⁶ <https://www.mckinsey.com/featured-insights/artificial-intelligence/notes-from-the-ai-frontier-applications-and-value-of-deep-learning>

⁷ <https://academic.oup.com/qje/article-abstract/131/4/1593/2468873>

⁸ http://sekhansen.github.io/pdf_files/fomc_transparency.pdf
<https://www.ecb.europa.eu/pub/pdf/scpwps/ecb.wp2085.en.pdf>

Data technology infrastructures and software, lack of capable human resources (data scientists) and dedicated Big Data unit, lack of procedures in place for ensuring data privacy, not to mention questionable data quality and the complex pre-processing required to clean such data. I will focus on one specific challenge: **how policy-makers should obtain access to Big Datasets.**

It is quite understandable that **institutions and companies are concerned and wish to keep their data private to themselves.** Data has been termed as the new oil in the 21st century, considering its ability to generate and maintain competitive advantage for its owners⁹. Absent overarching regulations that require other institutions and companies to report their proprietary data, government institutions are ill-positioned to obtain most Big Datasets relevant for analysis and policy-making, since they are not themselves producers of such datasets.

As policy-makers, it becomes our task to explain to the greater public that access to Big Datasets are aimed for and only for crafting better policies, and to make good on this promise. A clear regulation in conjunction with clear objectives should help convince the public of the necessity for government access to Big Datasets and elicit more cooperation and compliance.

Central banks, statistics offices, and government institutions should also establish a close coordination for data sharing¹⁰. It is very burdensome to the public if they are required to report the same datasets multiple times, especially if different government institutions each require different specifications for the reports. This is even more relevant in the case of Big Data, since transferring voluminous amount of data in near real-time to multiple institutions will be impractical, if not impossible. Depending on the cross-cutting of jurisdictions in the country, relevant datasets are also often captured in silos in different government institutions. **Thus, besides reducing reporting load, data sharing allows a more complete picture for analysis and policy-making.**

Of course, all data access and data sharing initiatives should be established with stringent privacy and confidentiality measures that protect both the data producers and the respective individuals. For example, reporting and sharing of identifiable information that are not needed for analysis, such as people's names and exact addresses, should be kept at a minimum and kept private. Sufficient IT standards that cover all relevant aspects of data privacy, including network and database security, should be implemented and duly observed. In addition, internal access to confidential Big Datasets should be assessed and authorized on a need-by-need basis.

Distinguished Guests, Ladies, and Gentlemen,

Allow me to conclude my remarks.

I have discussed what we consider to be the main factors driving Big Data development and its widespread adoption. Big Data analytics is made possible through massive and ubiquitous digital recording of our activities, complemented by the growth of computing power. Big Data can inform better decisions through

⁹ <https://www.economist.com/leaders/2017/05/06/the-worlds-most-valuable-resource-is-no-longer-oil-but-data>

¹⁰ <https://www.bis.org/ifc/events/7ifc-tf-report-datasharing.pdf>

analysis of richer, more granular data, in more timely manner. Recent progress in machine learning algorithms also allows more accurate analysis of unstructured datasets, such as texts and images.

I have also discussed the need for data access and data sharing for public institutions, along with several considerations that need to be taken into account in the implementation.

I am fully aware that it will not suffice to discuss all aspects of and trends in Big Data in a single speech nor a single day of seminar. I sincerely hope that we all have much to learn from this Seminar, and we can go back to our institutions with concrete strategies and improvements to be put in practice.

Finally, I would also like to mention that this event is held as part of the Voyage to Indonesia (VTI) series of activities that we prepare as the groundwork for the 2018 IMF-WBG Annual Meetings here in Bali. The theme Voyage to Indonesia reflect a journey that will bring the world to the renewed Indonesia; a reformed, resilient, and progressive economy. We hope that you will continue to be engaged in various VTI programs.

With that, allow me to declare the opening of this Seminar.

Thank you very much.