September 2017 SDN/17/06

IMF STAFF DISCUSSION NOTE

Big Data: Potential, Challenges, and Statistical Implications

Cornelia L. Hammer, Diane C. Kostroch,

Gabriel Quirós, and STA Internal Group

September 2017 SDN/17/06

IMF STAFF DISCUSSION NOTE

Big Data: Potential, Challenges, and Statistical Implications

Cornelia L. Hammer, Diane C. Kostroch,

Gabriel Quirós, and STA Internal Group

Statistics Department

Big Data: Potential, Challenges, and Statistical Implications

Prepared by Cornelia L. Hammer, Diane C. Kostroch, Gabriel Quirós, and STA Internal Group[1,2]

Authorized for distribution by Louis Marc Ducharme

JEL Classification Numbers:	E0, Y2, Z0
Keywords:	Big Data, Macroeconomic and Financial Statistics, Official Statistics, Data Quality, Surveillance
Authors' E-mail Addresses:	CHammer@imf.org, DKostroch@imf.org, GQuiros@imf.org

[1]The Internal Group on Big Data of the Statistics Department (STA) was established in August 2016 to investigate opportunities and challenges of big data for macroeconomic and financial statistics. The Group is led by Gabriel Quirós, STA Deputy Director, and composed of Serkan Arslanalp, Jose Maria Cartas, Aimee Cheung Kai Suet, Daniela Comini, Saurabh Gupta, Andreas Hake, Cornelia L. Hammer, Gary Steven Jones, Venkateswarlu Josyula, Diane C. Kostroch, Stephanie Medina Cas, Edgardo Ruggiero, and Patrizia Tumbarello.

[2] The authors would like to thank, in particular, for valuable comments received on previous drafts Louis Marc Ducharme, El Bachir Boukherouaa, and ITD colleagues for their excellent IT-related contribution; Claudia Dziobek and Mark Van Wersch for review and editing; and Manuela Goretti for her valuable comments. We would like to thank IMF Multimedia Services for graphics and our COM colleagues Jim Beardow, Linda Long, and Lucy Scott Morales for their editorial support. The paper has also greatly benefited from the review and comments received from IMF area departments (AFR, APD, EUR, MCD, WHD) and functional departments (COM, FAD, ICD, ITD, LEG, MCM, RES, SPR), and from the analytical input and administrative support provided by James Chan, Liliya Nigmatullina, and Zula Oimandakh (all STA).

CONTENTS

EXECUTIVE SUMMARY

This Staff Discussion Note reflects on the potential, challenges, and implications of big data for macroeconomic and financial statistics. It addresses the wide range of stakeholders of "official" data and statistics and covers interested users and producers. Good data and statistics, strategic elements for any society and economy, are essential for sound policy decision making in both the private and public sector. By now, many private companies as well as national and international organizations see that "big data" is no mere buzzword, but a medium-term concept that requires a long-term vision.

Big data is evolutionary and can provide innovative, real-time, and more granular insight for economic and financial analysis. Yet **big data opportunities for individual countries will be asymmetric and will depend on the country's characteristics and the availability of the systems and networks generating big data.** Big data offers opportunities, challenges, and implications for official statistics that compilers and users of statistics need to be aware of when they start to incorporate big data into their work plan to the extent relevant.

Numerous individual applications of big data are already being carried out, either by users or compilers of data and statistics. However, a systematic and structured discussion is lacking: this SDN attempts to offer such, emphasizing implications for macroeconomic and financial statistics. Further research and detailed analyses are essential to understanding if and how big data can directly and indirectly support IMF surveillance work.

What does big data mean? Although there is no agreed-on definition, the term is often characterized by the 3Vs—high-volume, high-velocity, and high-variety. More Vs have been added over time, such as veracity and volatility. Unlike statistical data compiled for specific purposes, big data is made up of byproducts found in business and administrative systems, social networks, and the internet of things. To structure the discussion, the paper presents a big data classification that is relevant for macroeconomic and financial statistics.

What is the potential of big data? Big data can benefit macroeconomic and financial statistics and ultimately policymaking through at least three features:

1. By answering new questions and producing new indicators
2. By bridging time lags in the availability of official statistics and supporting the timelier forecasting of existing indicators
3. As an innovative data source in the production of official statistics

What challenges come with big data? Data quality concerns, difficulties with access, and new required skills and technologies are the main challenges of big data. And while big data mainly measure insights, correlations, trends, and sentiments, detailed country-by-country time series in

accordance with internationally agreed standards remain crucial for measuring and monitoring countries' economic performance and policies over time.

Statistical Implications: Moving forward, international statistical cooperation is key to overcoming big data challenges and to building lasting partnerships among national and international statistical agencies, users, and data owners. The implications include the need for interested parties to build up the required skills and technologies in their organizations. Opportunities, challenges, and potential implications are particularly high for national statistical agencies: **the incorporation of big data as new data sources, either supplementing or substituting for traditional data sources, will not be exempt from methodological, organizational, and budgetary challenges.**

The success of big data projects lies not in implementing a particular piece of technology, but rather in establishing an environment of people and processes that take big data innovations forward and put them to work. **Given the diverse skills needed to deal with big data, it also provides an opportunity for organizations to break their internal silos, including between users and producers of data and statistics.**

From individual applications of big data to its incorporation into the systematic, regular, and large-scale production of statistics—before engaging in costly and time-consuming investments, organizations should begin with a proof of concept and should operationalize the project only after findings have proved valuable and feasible from an organizational point of view. Statistical agencies should decide on a case-by-case basis and select the most promising big data projects to complement existing statistics. Moreover, to keep abreast of developments, agencies should proactively search for big data sources to address the most urgent research needs. Selected big data projects may also be incorporated into capacity development activities to support the membership in building their capacity to benefit from available big data sources. Going forward, research in and compilation of best practices—for statistical techniques and methodologies that address veracity and volatility, specifically—need to be at the top of the statistical community's agenda.

Given that big data is not static but dynamic, the systems and networks generating big data continue to evolve, and with them the possibilities, challenges, and limitations of big data for statistics. Consequently, **the overall assessment made in this paper will need to be revisited as the worlds of big data and official statistics evolve.**