BIG DATA SOLUTIONS

Harnessing the Power of **Big Data** for **Trade and Competitiveness Policy**

WORLD BANK GROUP

CONTENTS

FOREWORD

Big data - data that is vast at an unprecedented scale, comes from a range of new and old sources, is extremely high in frequency, and which can now be analyzed and combined in very sophisticated ways - has the potential to augment the efforts of institutions like the World Bank Group working to deliver high-impact development solutions to countries around the world.

Within the Bank Group's Trade & Competitiveness Global Practice, we believe that investigating the potential of big data for development is a worthwhile pursuit as we support countries to boost the volume and value of trade, enhance their investment climates, improve competitiveness in sectors, and foster innovation and entrepreneurship.

Big data solutions have the potential to accelerate the work of our teams by deriving timely, accurate, and actionable insights from alternative data sources in order to close data gaps and inform policymaking. For example, pilot projects underway in the Trade and Competitiveness practice are exploring the use of data science techniques to harness publicly available government and commercial data to aid competition authorities in detecting cartels and other anti-competitive practices. We are also mining Internet data to measure innovative economic activity in cities, so agencies can make better informed policy decisions and, we are collecting regulatory data to classify and assess the impacts of non-tariff measures on economies and their competitiveness.

Of course, there are limitations to operationalizing big data to inform our work—navigating privacy policies and other challenges to access, for example—but we continue to pursue big data's potential because we know it can support the Bank Group's goals of ending extreme poverty and promoting shared prosperity.

This paper, prepared in collaboration with Deloitte and with other global practices within the Bank Group, highlights data-driven pilot projects underway in the Trade & Competitiveness Global Practice and shares compelling cases of how big data is changing the way we look at the challenges countries are facing and how we can best support them.

These are exciting times for big data for development. We hope that these cases will prove useful for policymakers, academics, students, trade and competitiveness practitioners, data enthusiasts and many others as they pursue and develop breakthroughs in big data-driven development solutions.

Anabel Gonzalez
Senior Director
Trade & Competitiveness Global Practice

ACKNOWLEDGMENTS

This publication has been prepared jointly by the staff of the World Bank Group (WBG) and Deloitte Consulting LLP (Deloitte). Prasanna Lal Das (WBG) and Trevor Monroe (WBG) coordinated the publication. Alyona Polomoshnova (Deloitte), Dale Kim (Deloitte), and Jack Sullivan (Deloitte) are the main authors. Several WBG staff provided valuable contributions to make this publication possible, including Andrew Whitby, Ankur Huria, Anwar Aridi, Georgiana Pop, Juni Zhu, Martin Molinuevo, Megha Mukim, Michael Ferrantino, Peter Kusek, Siddhesh Kaushik, Victor Mulas, and Yehia Eldozdar. We are grateful for the advice and suggestions of peer reviewers Daniel Reyes, Hooman Dabidian, Ian Gillson, Jean-Francois Arvis, Maja Andjelkovic, Mariem Malouche, and Michele Ruta. The preparation of the publication was carried out under the guidance of Anabel Gonzalez, Cecile Fruman, Klaus Tilmes, Dahlia Khalifa, and Mary Hallward-Driemeier at the WBG.

Photo Credits:
Cover page: Deloitte; Page 6: Nahuel Berger / World Bank; Page 10: Stephan Gladieu / World Bank; Page 19: Dominic Sansoni / World Bank; Page 28: Chlor Sokunthea / World Bank; Page 31: Dominic Chavez / World Bank

ACRONYMS

CDR	Call Detail Records
FDI	Foreign Direct Investment
GDP	Gross Domestic Product
GIS	Geographic Information Systems
GP	Global Practice (World Bank Group)
GPS	Global Positioning System
ICT	Information and communications technology
IoT	Internet of Things
LDC	Least Developed Countries
NTM	Non-Tariff Measures
NYC	New York City
OECD	Organisation for Economic Co-operation and Development
PII	Personally Identifiable Information
RFID	Radio Frequency Identification
SIM	Subscriber Identity Module
SMS	Short Message Service
UN	United Nations
US	United States
WBG	World Bank Group
WEF	World Economic Forum

EXECUTIVE SUMMARY

Integrating developing economies into global trade and investment markets is essential to the World Bank Group's twin goals of ending extreme poverty and promoting shared prosperity. The Trade & Competitiveness Global Practice helps low- and middle-income countries spur economic growth by developing strategies that boost trade integration, enhance investment climates, improve sector competitiveness, and foster innovation and entrepreneurship.

Effective trade and competitiveness interventions require high-quality data and analysis to help policymakers identify economic growth opportunity areas, formulate trade and investment policy, and design development interventions. Data is equally indispensable to determining the success of economic growth initiatives, enabling the benchmarking of initial economic conditions, and measuring the performance of new trade and competitiveness strategies. Consequently, a data-driven intervention agenda that focuses on gathering high-quality economic and market outcome data, and that connects the data to relevant policy areas, such as labor markets or agriculture, provides a strong foundation for designing trade and competitiveness policies that maximize benefits for the global poor.

However, low- and middle-income countries, which are most in need of policy interventions, often lack the kind of high-quality data about their economies that can help to inform effective trade and competitiveness strategies. While national statistical capacities are improving, alternative datasets and analytical techniques enabled by the rise of big data are fundamentally transforming traditional approaches to economic data collection and analysis. Indeed, big data has the capacity to supplement, or even replace, previously cumbersome and expensive traditional data collection with new, alternative sources of information. The rapid growth of advanced data science techniques and technologies also is transforming the range of evidence that can be derived from big data to inform decision making in trade and competitiveness policy.

At its core, through the formulation of economic insights that were previously inaccessible, big data's promise in the realm of economic trade and competitiveness is to help formulate the components of effective trade and competitiveness policies and interventions. These components include the understanding of economic activity and linkages, the fostering of a favorable investment climate, the optimization of logistics and supply chain management, the elevation of the poor, and the competitiveness of cities.

Promising big data applications in international development are already defining a departure from traditional approaches to formulating trade and competitiveness strategies. For example, performing advanced analytics on large and disparate trade datasets makes global trade insights more accessible; using real-time, auto-generated data such as satellite imagery informs the pricing of agricultural insurance products; and the ever-expanding Internet of Things (IoT)* enables remote management of a hospital's Radio Frequency Identification (RFID) tagged supplies.

This knowledge note presents three case studies representative of the Trade & Competitiveness Global Practice's innovations in big data solutions. The studies demonstrate the promising potential for big data in trade and competitiveness policymaking. The first case study demonstrates how machine learning and web scraping techniques can help competition authorities identify cartels and other anti-competitive practices. The second case study demonstrates how reliable, comprehensive, and comparable information assessing the innovation capacities of cities can be derived from open-source data to assist the policy making of city governments. The third case study demonstrates how machine learning and text mining techniques can accelerate the collection and analysis

* The International Telecommunication Union defines the IoT as a global information infrastructure, enabling advanced services by interconnecting physical and virtual things based on interoperable technologies.

of data on non-tariff measures to support timely, evidence-based policy decisions to reduce barriers to trade. In addition to these three cases, this knowledge note also highlights additional promising applications big data throughout the text.

While big data applications are promising, their scalability for economic development efforts can be challenging. Datasets essential to this type of operationalization may often be private or proprietary. Furthermore, big data requires data science talent to manipulate and draw out meaningful insights to guide appropriate decision making. Lastly, there remains some skepticism in the economic development community of big data's ability to accurately assess economic indicators and reconcile them with on-the-ground truths. While these challenges represent some hurdles to the operationalization of big data in trade and competitiveness interventions and policymaking, promising public and private sector solutions are emerging and are presented in this note.

Despite these concerns, this Practice believes that harnessing big data can help policymakers take an innovative approach to spur economic development in low- and middle-income countries by helping them to better understand local economic activity, foster increased investment, resolve logistical barriers and promote competitive supply chain management, and increase the competitiveness of cities and the poor. For policymakers operating in a world of increasingly large and disparate datasets, the challenge now is to make use of these new data sources in ways that maximize the impact of their policies and interventions to reduce poverty and increase shared prosperity.

SECTION 1

BIG DATA: THE FUTURE OF COMPETITIVE DECISION MAKING

"Big data" is broadly used to define the techniques for collecting and analyzing large datasets to gain productive insights and inform decision making. As the sheer quantity of global data skyrockets, so do the operational uses for it. For those poised to embrace it, big data is quickly becoming a new foundation for data-driven competitive decision making, challenging traditional approaches through innovative solutions to complex problems such as economic development. To explore the potential for big data to inform competitive decision making for those in low- and middle-income countries, imagine the scenarios in figure 1.

These decisions take place in different contexts and pertain to different economic issues, but they all have been informed by big data. Despite its name, big data is about more than just the quantity of data available for use. It is about diversity, speed, resourcefulness, and creativity in collecting, analyzing, and operationalizing that data to inform decisions by leveraging all information available. It's also about devising new products and services that take large, unstructured data inputs and derive actionable insights.

In this report, we describe how big data innovations are revolutionizing the way we think about international trade integration, promoting competitiveness and investments, and fostering entrepreneurship in developing economies. We also discuss how big data has the potential to shape the future of policymaking in economic development.

Figure 1: How Big Data Could Influence Competitive Decision Making

Source: World Bank.

SECTION 2

THE CHALLENGE: ENABLING EVIDENCE-BASED TRADE AND COMPETITIVENESS POLICY

Over the past century, globalization has created unprecedented opportunities for individuals, businesses, and governments to integrate with global markets. In the fight against global poverty, acquiring an understanding of global economic activity and how it affects markets in low- and middle-income countries is a precursor to the success of economic development interventions. Policymakers should focus on acquiring quality data to inform effective trade and competitiveness policies aimed at fostering investment, optimizing supply chains, and increasing competitiveness for cities and the poor. However, information gaps in low- and middle-income countries sometimes hinder the achievement of these objectives and continue to pose great challenges to policymakers. On the other hand, these information gaps also represent opportunity areas for innovative big data solutions to inform trade and competitiveness policy.

The objectives for effective trade and competitiveness interventions covered in this note – understanding economic activity and linkages, fostering investment, optimizing logistics and supply chains, increasing urban competitiveness, and improving the competitiveness of the poor – are representative pathways for effective policies to spur economic growth in low- and middle-income countries. Trade, for example, is a fundamental component to understanding global economic interconnectivity. From 1990 to 2013, the share of exports in global gross domestic product (GDP) rose from less than 20 percent to over 30 percent. The global economic landscape is also growing more complex: trade in

goods is expected to level off with a shift toward services, for which data plays a critical component in tracking.[1] For low- and middle-income countries, trade is an especially valuable tool in improving growth and competitiveness. For example, well-informed policies that reduce trade costs and improve trade facilitation have been shown to increase inflows of foreign direct investment (FDI).[2] Improved trade facilitation also helps the rural poor by empowering farmers to move perishable goods to market with greater reliability.[3]

However, without the economic data to enable effective policy making, these avenues of growth may not be achievable. For low- and middle-income countries, the consequences of economic data gaps can be dire. The World Economic Forum (WEF) reported in 2014 that developing economies, which have historically relied heavily on investment and commodity-driven revenue, faced tighter capital markets and lower commodity prices, diminishing their prospects for future growth.[4] Furthermore, the 2015-2016 WEF report showed that less competitive economies are less resilient to external shocks, and were consistently outperformed by competitive economies following the 2008 Global Financial Crisis.[5]

The objectives that follow are several illustrative focus areas in which economic development institutions can inform evidence-based policymaking by addressing the need for data:

Understanding Economic Activity and Linkages

Every year, the world produces greater quantities and varieties of data measuring economic integration, productivity, and competitiveness. Data ranges in complexity from descriptive figures to statistical indices providing deeper insight into economic trends. Two traditional sources of economic data for policymakers are the World Bank Group's Ease of Doing Business Index,[6] which assesses business regulations and enforcement, including taxes, trade, and contract enforcement; and the WEF Global Competitiveness Index,[7] which captures economic fundamentals and assesses an overall index for competitiveness based on basic requirements, efficiency enhancers, and innovation and sophistication factors.

As the global economy becomes increasingly interconnected and complex, tracking, understanding, and responding to economic trends in low- and middle-income countries through traditional lenses will become increasingly challenging for policymakers. This presents an opportunity for innovative big data solutions to provide governments, businesses, and individuals with difficult-to-access but pivotal information on economic activity in developing economies. For example, policymakers could use supplementary evidence acquired through utility, phone, and transaction records to understand how market access and firm networks affect productivity and competition to formulate effective economic interventions. Businesses can harness aggregated data on the dispersion of prices across retailers selling a similar good to better understand their competition. Among individual producers, a smallholder farmer may use Geographic Information Systems (GIS) precision agriculture tools, for example, to optimize crop production.[8]

Fostering Investment in Developing Economies

Investment is a central component of growth for developing economies to attract capital, technology, infrastructure, and human capital. Building a business climate that attracts investment is an important goal for economic policymakers in low- and middle-income countries. To this end, they should work to ensure that regulations are efficient, transparent, and fair to promote investment in priority industries across their economies.

Policymakers are heeding these directives. In the twelve years since the World Bank Group launched its *Doing Business Report,* which provides a comprehensive analysis of the barriers, procedures, and costs associated with starting and operating private firms in 189 economies, there have been 2,600 regulatory reforms conducted globally. The results of these reforms are equally impressive, with the average length of time required to establish a business dropping from 51 days in 2003 to 20 days in 2015, for example.[9]

Despite this progress, investment flows to developing economies are subject to volatile factors, including commodity price fluctuations and investor sentiments. For policymakers, acquiring actionable information that allows them to understand investment flows and volatilities is critical to fostering a strong and resilient investment climate. Investors also need economic data that allows them to assess risk accurately. Big data solutions that help both parties acquire actionable insights into the investment climate would accelerate and revolutionize the ability of and range of tools policymakers can use to draw investor capital to low- and middle-income countries.

Optimizing Logistics and Promoting Competitive Supply Chain Management

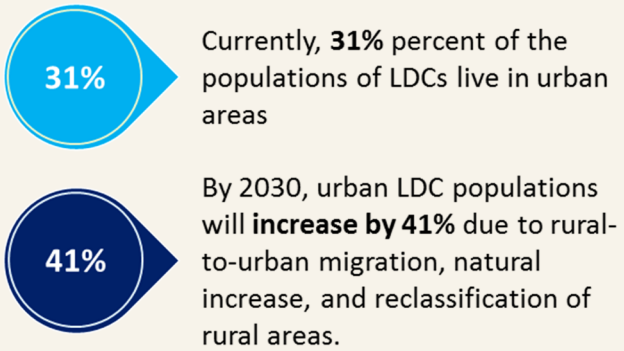
To capitalize on global interconnectivity, producers in low- and middle-income countries need access to markets, but there are often logistical, economic, and political barriers to this seemingly simple objective: developing economies consistently score lower than developed countries on the World Bank Group's Logistics Performance Index. In 2014, developing economies scored an average of 2.41 against the Organisation for Economic Co-operation and Development (OECD) member country rating of 3.70 (out of 5), illustrating the opportunity for logistics performance improvement as a factor for economic growth.[10] These countries may also face infrastructure-related obstacles, including lower road density and poor road conditions. Other roadblocks include problematic governance: for example, smallholder farmers may be coerced by government officials to pay bribes before transferring and selling their produce across borders.[11]

Ensuring efficient, fair, and competitive supply chains is pivotal to achieving valuable and inclusive growth policies. Policymakers should identify barriers to effective supply chains by gathering detailed information on how goods and services are being exchanged. If innovative big data solutions leveraging the Internet of Things, for example, illuminated supply chain networks and barriers for the policymaker, it could help them improve trade competitiveness by informing interventions that optimize logistics in low- and middle-income countries and eliminate barriers to inclusive growth.

Increasing the Competitiveness of Cities

The world is urbanizing at an unprecedented rate. The majority of the world's population already lives in cities, with that proportion expected to grow as high as 66 percent by 2050.[12] Industrialized countries have historically led this trend. In the Least Developed Countries (LDCs), the proportion of the population living in cities is just 31 percent.[13] However, the urban population in LDCs is projected to increase by as much as 41 percent by 2030[14] (See figure 2). This explosion has critical implications for the development of competitive urban economies.

Figure 2: Increasing Demand for Urban Economic Data in Least Developed Countries (LDCs)



Source: World Bank Group and UN Population Fund.

Policymakers need data to understand these burgeoning urban economies and develop effective policies that increase their competitive capacity. For example, accurate and timely poverty distribution figures can help policymakers effectively target economic interventions in slums.[15] Innovative big data approaches can also help policymakers acquire hard-to-access measures, such as capacity for innovation, skills gaps, and the size of the informal economy.

Improving Competitiveness for the Poor

Even with significant gains in human capital and investment, much work remains to improve the competitiveness of entrepreneurs and businesses in low- and middle-income countries. Factors that can boost or impede economic competitiveness among the poor range from macroeconomic stability to currency crises, but a particularly challenging barrier to competitiveness is insufficient access to capital. Roughly four billion people — over half of the world's population — lack access to traditional financial services.[16]

Policymakers can leverage innovative big data solutions to derive insights on local catalysts and inhibitors to individual competitiveness in developing economies. This would help policymakers formulate interventions to improve the competitiveness of entrepreneurs in low- and middle-income countries by attracting and expanding access to capital, thereby fostering inclusive growth.

SECTION 3

INNOVATIVE BIG DATA SOLUTIONS OFFER NEW SOURCES OF INFORMATION

Figure 3: Defining Big Data

Big data can be understood through the following five characteristics which describe its departure from traditional data: Volume, Variety, Velocity, Veracity, and Value.

Volume: Large quantities requiring advanced analytical methods to generate meaningful information

Variety: Drawn from a broad range of often non-traditional sources

Velocity: Data is drawn from sources available in near real-time

Veracity: High-quality data must be drawn from credible sources

Value: Analysis yields meaningful insights that inform decision making

Source: Gartner.

Achieving the objectives presented will require creativity, collaboration, and most importantly, information. Enter big data. Experts have put forth countless definitions of big data over the last twenty years, but a widely accepted version comes from Doug Laney, Vice President of U.S. research firm, Gartner's. He posited that big data can be captured by three "Vs": Volume, Variety, and Velocity as defined in figure 3. Experts have advocated for the inclusion of a fourth "V," Veracity, acknowledging that uncertainty could remain as a result of poor source data quality. In practice, innovators are crafting big data solutions rooted in various combinations of the three "Vs." What truly defines big data is the collection and analysis of information in innovative ways to derive productive insights and aid effective decision making. This defining purpose is often captured as a fifth "V," for Value.

Big data solutions have a substantial role to play in enabling evidence-based trade and competitiveness policy formulation. As a general introduction to the concept of big data solutions in development contexts, consider the following three strategies for deriving evidence-based insights from big data to formulate trade and competitiveness policy.

Advanced Analytics for Large and Disparate Datasets

Whereas national and market-specific economic activity and linkage data have traditionally been presented and accessed in static tables, data scientists are now creating tools that empower policymakers, businesses, and individuals to easily access and manipulate data, and to use the insights gathered to make effective economic policy decisions. The development of these innovative big data decision-making tools often relies upon the application of advanced data science methods and technologies to perform analytics on large and disparate datasets that were difficult for policymakers to leverage in their original forms.

Global customs data analytics for policy insights

Global trade in goods, as a critical component to economic growth, has averaged 7 percent growth annually. The share of exports originating in developing economies rose from 34 percent in 1980 to 47 percent in 2013.[17] However, global customs data is often difficult to navigate, limiting the ability of policymakers to gather insights to make informed

trade policy decisions. Companies such as U.S.-based Panjiva, which analyzes global trade, are employing machine learning-based intelligent feature extraction to aggregate and derive key customs transaction information (e.g., source, destination, types of goods) through analytics, ultimately to support decision makers on trade policy. Using customs data from eight governments, including Chile, China, Mexico, and the United States, Panjiva's data analytics can collect, clean, and process customs data detailing trade information from 190 countries comprising over 450 million records — despite inconsistency in documentation conventions – and provide the aggregated analysis to users. Actionable trade information enables policymakers to better understand how firms in low- and middle-income countries are impacted by trade reforms or economic shocks, or how they are shifting their product mix in response to new competitors, for example.

Advanced text analytics for regulatory compliance

Deloitte and IBM are partnering to create software that performs advanced text analysis on companies' financial management frameworks, and evaluates them against relevant government regulations to identify compliance risks. In the future, incorporating text analytics into regulation enforcement mechanisms could help governments save millions of dollars and increase compliance rates.[18] For policymakers, it could help increase their visibility in regulatory enforcement in developing economies, where they are planning trade and competitiveness interventions.

Labor market monitoring using internet data

Online job portals such as LinkedIn feature a wealth of information on the supply and demand of skills in the labor market that may offer policymakers extractable insights. Indeed.com, for example, currently displays the ratio of job postings-to-unemployed persons in cities across the United States, thereby affording policymakers a quick view into relative shortages or surpluses of labor demand.[19] Similar sites in low- and middle-income countries, such as India's Babajob, present opportunities for detailed analysis, and could potentially inform the design of skills development programs in these countries.

Creative Uses for Real-Time, Auto-Generated Data

Variety (the second "V") reflects a particularly innovative paradigm that big data solutions rely upon, which is the technological advancements that have enabled the collection of data – and the production of actionable information – from creative, non-traditional sources. A prime example of this resourcefulness is the application of mobile phone data. The everyday actions of individuals can be collected and analyzed through their use of mobile phones, which automatically generate real-time data through call detail records (CDRs) recording call duration, source and destination numbers, geographic location data using cell tower activation, in addition to the sensor and Global Positioning System (GPS) mobility data produced by most smartphones today. This data can also be used to map human movement and create statistical models that infer socioeconomic measures, such as a self-learning algorithm that identifies that people who make calls during typical "work hours" are less likely to be formally employed. Moreover, access to mobile phones is no longer limited to the wealthy – 5 billion of world's estimated 7 billion mobile phone subscriptions originate from developing countries.[20] Given the range of data mobile phone technology produces, and its increasing penetration in low- and middle-income countries, it has the potential to inform trade and competitiveness policy in contexts where traditionally used data may not be as readily available.

Mobile data facilitates financial access for the poor

For example, mobile phone data can be used to facilitate access to banking and capital for low-income individuals. Access to finance is inequitable in low- and middle-income countries, where Gallup estimates that the wealthiest members are banked at a rate of 64 percent versus 24 percent for the poorest members.[21] For the rural poor, it can be logistically difficult to travel to banks and establish accounts, even mobile money accounts. For others, banking fees and other costs may be prohibitive: 75 percent of the poorest individuals cited "Not Enough Money" [to open an account] as a reason why they did not have a formal bank account, according to a 2012 Gallup survey.[22] Furthermore, even when poor individuals, families, and entrepreneurs can access financial services, the costs associated with financial services,

such as interest rates, are prohibitively high. This is largely due to the difficulty of assessing risk for loans to the poor – only 27 percent have a traditional credit score, as fewer poor people have formal financial histories.

First Access, a U.S.- and Tanzania-based financial services company, capitalizes on growing mobile penetration rates in developing economies to analyze pre-paid phone records and produce loan assessments to lenders through a self-learning algorithm. The algorithm uses the loan recipient's pre-paid mobile phone records to derive information, including the user's age, gender, geographic proximity to urban areas, and where and how frequently they move. Mobile phone records also provide analysts with data on the financial capabilities of the user, such as how often they are buying minutes, sending remittances via mobile money, and so on. Individually, these derived components do not provide enough information to generate an accurate risk assessment. However, First Access uses a self-learning algorithm, which is informed by hundreds of thousands of prior assessments and behavioral patterns, to assess the recipient's creditworthiness.

Although alternative credit analyses such as those provided by First Access support achieving universal financial access, they may not become a central feature in the intervention portfolios of government policymakers and development institutions. The methodology does, however, represent a fundamental shift in how financial inclusion can be extended to the global poor.

Just as uncertainty is a barrier to finance, it can also inhibit the development of an attractive investment climate. Investors shy away from projects in countries experiencing high levels of political or economic risk. Fortunately, big data's potential transcends individual and small business loans. Innovators are leveraging a variety of big data sources to paint a more accurate, real-time picture of investment risk landscapes, which represents a promising opportunity for policymakers seeking to attract investment to developing economies.

Remote sensing measures economic productivity

Potential applications for big data in trade and competitiveness also extend to remote sensing data. Satellite imagery analysis, for example, could produce proxy measures for manufacturing sector productivity. Consider SpaceKnow, a U.S. satellite imagery company that turns satellite imagery into actionable business intelligence. One of SpaceKnow's products, the China Satellite Manufacturing Index, uses a 1-100 score that reflects data gathered from 2.2 billion satellite photos taken over 14 years covering 500,000 square kilometers to capture whether the Chinese manufacturing sector is expanding or contracting. Analyzing changes across 6,000 industrial sites and incorporating features such as the number of trucks in industrial parking lots and frequency of turnovers allows SpaceKnow and its customers to monitor manufacturing sector size and competitive capacity in real-time. Policymakers can leverage these kinds of alternative measures to improve their ability to balance sectors and craft policies.[23]

Crowdsourcing economic competitiveness measures

The Trade & Competitiveness Global Practice has investigated opportunities for big data to inform measures of economic competitiveness. In 2014, the World Bank Group (WBG) conducted three pilot studies to assess intra-regional trade in Africa through SMS-based crowdsourcing. The first study examined the price of fertilizer and farmer satisfaction with seed quality. The study found that the price of urea was 16 percent higher in Kenya than in Tanzania, which is a substantial cross-border difference for homogenous products. The second study gathered quantitative information about trade in health and educational services, such as costs, market size, and quality, across nine countries. It found that that the quality and availability of services were more important determinants of trade flows than the cost. The third study surveyed official and unofficial barriers for cross-border traders from four countries. Analyzed by gender and the type of goods carried, the study concluded that traders were forced to pay unofficial fees, contend with verbal insults, and even experience physical abuse.

Embracing the Internet of Things Revolution

The Internet of Things (IoT) is a revolutionary idea powered by exponential technologies: as more devices become digitally interconnected, they increasingly capture, produce, and exchange data on the physical world. This data can be leveraged to inform real-time decision making to improve efficiency, accuracy, and utility. IoT devices boast a huge variety of uses, including accelerometers in smartphones, sensors monitoring soil quality, and


IMPSA
QC 03
IMPSA
40T QC 02
NYK
45
OOCL
tex
02

SECTION 4

INNOVATIONS IN BIG DATA FROM THE TRADE & COMPETITIVENESS GLOBAL PRACTICE

The following three cases (box 1) exemplify some of the Trade & Competitiveness Global Practice's innovative big data solutions projects that have been developed and/or are currently under further development. These applications are representative of the transformative potential for big data to change the way governments and international development institutions formulate trade and competitiveness policy.

Box 1: Innovations in Big Data from the Trade & Competitiveness Global Practice

Case Study #1
Leveraging Big Data to Help Competition Agencies Tackle Anticompetitive Behavior

Using machine learning and web scraping techniques, this project helps competition detect cartels and other anti-competitive practices in 16 pilot countries.

T&C Working Area:
Competitive Sectors

Case Study #2
Using Big Data to Measure City Innovation Capacity

The Start-Up City Dashboard provides city governments with a tool to help measure and compare the health, diversity, and scale of innovative economic activity in 22 pilot cities.

T&C Working Area:
Innovation & Entrepreneurship

Case Study #3
Classifying Non-Tariff Measures using Machine Learning

To accelerate the collection and analysis of non-tariff measures (NTMs), this project used machine learning and text mining techniques to help governments and businesses assess the impact of NTMs on the broader economy.

T&C Working Area:
Trade Competitiveness

Leveraging Big Data to Help Competition Agencies Tackle Anticompetitive Behavior

By Georgiana Pop, Andrew Whitby

Summary:

Anticompetitive practices have been found to yield negative effects on productivity growth in developed and developing economies. Cartels across the world, for instance, negatively impact consumer welfare through price overcharges in the order of billions of dollars. But detecting these practices requires a better understanding of the nature of such anticompetitive practices by market operators across various markets and jurisdictions. This is not an easy task. Currently, competition authorities and researchers who study effects of anticompetitive behavior may have to gather this data manually, which is both time-consuming and inefficient.

In response, this project aimed to develop a database of key decisions by the competition authorities relating to anti-competitive practices. Using machine learning and web scraping techniques, the project automates the collection and organization of data from sixteen pilot countries. The database would serve as an essential infrastructure for future visualization and analyses to identify signals of anti-competitive behavior.

Challenge:

Data on anticompetitive practices is an integral part of Competition Authorites' work. For example, in the case of cartels, such data may cover data on market characteristics, specific regulations, conduct, decisions, and sanctions imposed. But many times competition authorities do not currently have ready access to cross-market, cross-jurisdictional data on such anti-competitive practices. This is because this type of data is often collected manually from public sources such as competition authorities' websites, media, and specialized international organizations and consultancies. Moreover, building and maintaining a repository of this data "by hand" is time consuming--and can yield imperfect results.

Innovation:

Using machine learning techniques and web scraping, this project aimed to automate the collection and organization of key information on anti-competitive practices from sixteen pilot countries.

The database would allow competition authorities to better understand, detect, and take actions against actors that systematically engage in anti-competitive practices. At the World Bank Group, it would allow the Competition Policy Team to further develop its analytical tools and systematize its corpus of evidence on the effects of anticompetitive practices, including cartels in developing economies.

Further, this initiative could potentially be expanded to the analysis of various anticompetitive practices as well the implementation of other policies, e.g., investment policy and state-aid control policy.

Process and Results:

The team initially started out with a set of countries where competition authorities had published documents-- in either English or Spanish-- relating to anti-competitive practices. This initial set was narrowed down to sixteen pilot countries where a comprehensive set of records on past decisions was available. The pilot focused on the following sixteen countries: Albania, Botswana, India, Moldova, Romania, Uruguay, Argentina, Chile, Macedonia, Pakistan, Serbia, Bosnia, Costa Rica, Malaysia, Peru, and Seychelles.

In the first phase of the project, the team worked with a partner organization with technical expertise in machine learning and web-scraping to scan thousands of pages of documents and gather semi-structured data on anti-competitive practices. This involved creation of an algorithm that picked out relevant information from 5000 documents-- including the actors involved, type of anti-competitive practice, products/markets affected, type of

anticompetitive practices, decision taken, and sanctions, if any. This phase required the imposition of both automated and manual checks and troubleshooting to ensure credibility and quality of data collected.

The second stage, which is in progress, involves the extraction and structuring of this information so that it is ready for analysis.

In the third phase, the team hopes use this database to gather insights on anti-competitive behavior. This could potentially include: (1) snapshots of anti-competitive behavior by product/market, (2) network mapping and analysis to identify companies linked to previous record-holders across different markets, jurisdictions, and geographies, and (3) likelihood analysis of potential anticompetitive behavior for companies with links or previous records, etc.

Lessons learned:

1) *Allocate time and resources to communicate domain knowledge with technical experts:* Machine learning and web scraping expertise needs to be complemented by transfer of domain-specific knowledge to technical experts. For this, the team found it important to allocate enough time and resources to make sector-specific terms and analytical approaches/ practices explicit to technical experts.

2) *Human oversight is crucial in ensuring quality:* There is no perfect analytical technique when it comes to unstructured data, because so much depends on the quality of documents available online. Since there is a lot that needs to be taught to the machine, the human factor is very crucial in ensuring quality of outputs.

For example, while a human expert may intuitively place the terms “cartel”, “price agreement”, and “horizontal agreement” in the same category of anti-competitive practice, a machine has to be explicitly taught that this is the case. Breaking down what a human does into concrete steps and feeding it to the machine takes time—both in terms of algorithm design and in terms of cross-checking, so the process cannot be entirely automated at this time.

3) *Anticipate adjustments to time-frame:* From database creation to completing complex analyses, each phase of the project builds on the previous one. The team found it important to factor in additional time so that quality can be assured adequately throughout the project lifecycle.

Using Big Data to Measure City Innovation Capacity

By Megha Mukim and Juni Zhu

Summary

City leaders around the world have been grappling with economic development challenges in the face of slowing growth, changing demographics, and increasing unemployment rates, especially among youth. Task teams at the WBG are searching for ways to better understand these challenges and find solutions to help their clients.

Private-sector firms are the main drivers of job creation, productivity, and wage increases; they also drive much innovation. Despite the role that privately-held start-ups play in innovation-led growth, cities currently lack reliable, up-to-date, and comparable data necessary to understand and inform policy decisions that affect start-ups.

The Start-Up City Dashboard aims to (1) provide reliable, comprehensive, and comparable data on start-up activity and innovation ecosystems in data scarce environments; (2) provide a better understanding of start-up activity drivers to guide more targeted policy; and (3) demonstrate the use of big-data tools for more standardized WBG data and analysis. The Start-Up City Dashboard is comprised of three interactive visual diagnostic tools that help measure and compare the health, diversity, and scale of innovative economic activity in 22 pilot cities: the Health of the Innovation Ecosystem Tool, Industry Benchmarking and Uniqueness Dashboard, and the Innovation Archetype City-by-City Comparison. These tools allow city governments to obtain an up-to-date and accurate picture of their innovation ecosystems and to learn from other cities that are operating differently.

Challenge

The Trade & Competitiveness Global Practice's report, "Competitive Cities for Jobs and Growth: What, Who and How," aims to help cities understand how to facilitate private sector growth to create jobs, raise productivity, and increase incomes. The report identified four enabling factors for growth: (1) institutions and regulations; (2) infrastructure and land; (3) enterprise support and finance; and (4) skills and innovation. The findings also suggested that the creation of innovative small firms and the displacement of incumbents was one of the main sources of innovation and – according to the team's experience – a topic of interest for many city governments. City leaders asked three key questions related to the role start-ups play in innovation-led growth: "Who are the entrepreneurs and start-ups in my city?"; "What industries do they focus on, and are these unique to my city?"; and, "How is my city doing compared to others?"

Cities lack reliable, up-to-date, and comparable data on their innovation ecosystems that would help them answer these questions. Moreover, readily-available data sources are not much help for a variety of reasons. For example, most available data is often aggregated at the national level. Even when sub-national data is available, it is often limited to industrial sectors. It is difficult to find data necessary to assess the factors that contribute to a successful start-up ecosystem in cities, especially those that are intangible in nature, such as networking assets to help entrepreneurs get connected, or a city culture that tolerates failures and encourages collaboration.

Innovation

To provide cities with a reliable diagnostic tool, the team gathered data on each of the following factors that contribute to a successful start-up ecosystem: human capital, financial infrastructure, urban amenities, collaborative culture, and networking assets. They did this by identifying proxy indicators for each of the factors for which open-source data could be updated frequently and rapidly, and by using a combination of data science tools including R, Python, STATA, Excel and Tableau for data collection, transformation, analysis, and visualization. For example:

- Strength of human capital in a city is determined by the number of universities as obtained in real-time from Open Street Maps.

- Financial infrastructure is captured by the number of banks or ATMs in a city as obtained in real-time from Open Street Maps.
- The nature of networking assets can be assessed through information on available networking activities, for example, incidence of mentee-mentor relationships and presence of serial investors. The dashboard used AngelList to determine the percentage of entrepreneurs who are well-connected as a proxy for the strength of a city's networking assets.
- Urban amenities is approximated by the ubiquity of coffee-shops, pubs and restaurants as obtained in real-time from Open Street Maps.
- Lastly, collaborative culture is assessed by examining the percentage of technicians who are active on collaborative online platforms, such as Stack Overflow.

By virtue of being problem- and demand-driven, the project benefited from unique approaches in both design and process. Rather than being a purely academic endeavor, the project had a concrete goal of addressing concerns that clients consistently brought up in consultations, such as whether start-ups in their cities are creating competitive jobs for young people. In addition to speaking with World Bank clients to inform the design process, the team also consulted other Global Practices and Regions within the WBG. These consultations lent relevance to the project and led to the enthusiastic reception from clients.

Results

Three tools comprise the Start-Up City Dashboard, whose prototype has data on twenty-two pilot cities ranging from Dar es Salaam to New York:

1) The Global Start-Up City Snapshot provides a snapshot of the innovation ecosystem's health, including an overall rank and a breakdown score for each of the five factors that contribute to a start-up ecosystem.

2) The Industry Benchmarking and Uniqueness Tool allows for the identification of the industrial mix of the city's start-ups and how these industries compare with two to three similar cities.

3) The Archetypes of Innovation Activities Tool allows for one-to-one comparison of cities against the four innovation archetypes identified by the consulting firm, McKinsey & Company: science-based, engineering-based, customer-focused, and efficiency-based. This tool indicates a city whether it is strong or weak in a particular type of innovative activity compared to competitor cities.

Sub-national clients have responded enthusiastically to the prototype dashboard. For example, city leaders in Shanghai were able for the first time to compare Shanghai's performance to Seoul, Tokyo, New York, and even other Chinese cities, disaggregate the data by sector, and ask what other cities might be doing differently. The Dashboard is now being piloted in Tanzania as part of a broader initiative to understand entrepreneurial ecosystems, leading to the design and preparation of a US$100 million lending operation.

The team is keen to build upon its success. First, by working with a capstone group, the team is focusing on understanding the direction and the magnitude of the possible bias using these new sources of data obtained through web scraping methods as compared to data obtained from other traditional sources in select OECD and upper middle-income cities. Second, the team is exploring opportunities for corporate partnerships with IBM and LinkedIn to further the work on gathering reliable data on entrepreneurship and to deepen and scale the Dashboard to include additional variables. Third, the team is looking to scale the project to include up to 600 cities worldwide, including many in low-income countries.

This project also illustrates that similar web scraping methods from open-source websites to obtain national or sub-national proxy data can be employed to develop monitoring and diagnostic tools for other projects within the Trade & Competitiveness Global Practice and the WBG. Once these tools are established, efforts to maintain and scale them could be marginal.

Lessons Learned

1) *Design based on a solid analytical foundation:* This project was built on two years of initial research to understand the importance of helping clients with economic development challenges. The project brought

together team members with skill sets in urban and private sector development, which was critical to developing a broad-based tool for task teams operating across different thematic areas.

2) *Demand as a foundation for design:* This project proved the importance of aligning design in response to demand. The initial thinking underlying the tool was based on increasing demands from clients to understand start-up activity, particularly from a project in China. Strong demand kept the project focused and ensured its outputs were impactful for and responsive to its end-users (i.e. city leaders). In addition, the team continually reached out to regional task teams and experts in other Global Practices (GP) (Information and Communications Technology (ICT) & Transport GP, Social, Urban, Rural, and Resilience GP) to solicit feedback, which made the tool flexible to the needs of different clients and users.

3) *Seek help on technical expertise:* This project was fairly new for the team and the WBG. As a result, the team faced many challenges, including finding the right technical skills in the absence of a standardized Terms of Reference for the required expertise. For this, the team turned to other advisors both within the Trade & Competitiveness Practice and the Big Data team, and even to private sector firms, for advice. It helped to have a clearly identified knowledge lead in the GP to provide guidance and regular feedback and support.

Resources

Start-Up City Index, Health of the Innovation Ecosystem Tool: https://public.tableau.com/profile/romulo.cabeza#!/vizhome/DashboardDraft/WholeDashboard

Industry Benchmarking and Uniqueness Dashboard: https://public.tableau.com/profile/romulo.cabeza#!/vizhome/UniqueMarketsDashboardFinalversion/UniqueMarketsInformation

Innovation Archetype City-by-City Comparison: https://public.tableau.com/profile/romulo.cabeza#!/vizhome/InnovationArchetypeCity-by-CityComparison/SequentialPresentation

Competitive Cities for Jobs and Growth: What, Who, and How: http://documents.worldbank.org/curated/en/902411467990995484/pdf/101546-REVISED-Competitive-Cities-for-Jobs-and-Growth.pdf

Classifying Non-Tariff Measures Using Machine Learning

By Michael J. Ferrantino and Siddhesh V. Kaushik

Summary

Non-tariff measures (NTMs) are defined as policy measures other than tariffs that could impact the prices or quantities of goods traded. NTMs are of particular concern to exporters and importers in low-income countries, as they impede international trade and can prevent market access. Systematic collection of NTM data continues under a multi-agency process coordinated by the UN Conference on Trade and Development, but the process of collecting and classifying NTM data is cumbersome, time consuming, and heavily dependent on consultant skills.

As a response to this challenge, this project sought to automate the manual process of classifying NTM data. By using available data on Malaysia, this pilot project illustrates how machine learning and text mining techniques can be used to automate and accelerate the NTM classification process and improve data quality. This solution can help governments, international agencies, businesses, and researchers get a better sense of which NTMs are in place in a given country, and in turn, to assess the impact they have on the wider economy.

Challenge

Consider the following cases: Country A imposes a restrictive licensing system on imports of noodles to boost local manufacturing and agriculture. Country B bans the import of a particular chemical compound used by paint manufacturers, citing health dangers linked to the compound.

These are both examples of NTMs. NTMs can fall into various categories, from sanitary or environmental protection measures to other restrictions such as quotas and price controls. Their impact can reach beyond the policy or regulation's original intentions. In country A, consumers may end up paying significantly higher prices than if the goods could be imported from a neighboring country. In country B, despite the import ban on the chemical, import of paints containing the same compound continues to be allowed. This may protect factory workers in the paint industry from harm, but does nothing to protect consumers.

Availability of comprehensive data on NTMs is crucial for governments to make informed decisions on these issues. Data on NTMs allows policymakers to accurately assess the impact of policies and regulations that affect trade; data even enables the calculation of the dollar equivalent impact of NTMs. Accurate information on NTMs is also necessary to negotiate modern trade agreements. It is equally important to allow private sector firms to avoid uncertainties in conducting cross-border business. It familiarizes them with requirements and levels of compliance, such as labeling and certification, that trading with a particular country would require.

However, collecting data on policy measures and classifying them as NTMs is a laborious and manual process. To begin with, data collection relies heavily on consultants who have to be trained to read thousands of pages of regulatory documents and identify trade-related regulations as well as the countries and products they affect. Data collectors must also correctly classify a regulation as a NTM. For some countries, it can take up to nine months to fully classify the NTMs. In addition, many low-income countries often maintain this information in hard copy, and even when digitized, they are often stored as picture files. Data quality assessments of consultants' output face similar problems.

Process and Innovation

Through machine learning and text mining techniques, the project sought to address the challenge of gathering data classifying NTMs in an efficient and accurate manner. For the pilot initiative, the team chose Malaysia. Since NTMs have already been classified for Malaysia, the pilot's data could be evaluated for accuracy. The project used machine learning and text mining, in particular, a Support Vector Machine to train the algorithm to identify patterns in existing documents and replicate the process. First, 60 percent of the data from Malaysia was used as training data in which 8,400 paragraphs of text were evaluated. The algorithm achieved 92 percent accuracy in identifying whether paragraphs contained NTMs or not, and 85 percent accuracy in identifying

whether the NTMs fell into categories A or B under the NTM rules. The extracted information was categorized as: Source, Document Title, Regulation Title, Regulation Agency, Regulation Date, Regulation Description, Regulation URL, and Regulation Text. The algorithm can also help identify affected countries and products.

Results

First, the pilot significantly reduced the time and resources required to identify, assess, and classify documents relating to NTMs. Second, it allowed for more accurate and consistent classification of NTMs by reducing human intervention and error. Third, it will provide better, timely and more comprehensive data to inform policy decisions and reduce barriers to trade.

The pilot paves way for classification of NTMs in more countries. Further functionality can also be added to the product to allow for automatic periodic updates or to develop the ability to process documents in languages other than English. The same process can be borrowed to gather data on other trade-related issues, such as intellectual property and rules of origin.

SECTION 5

SCALING BIG DATA FOR TRADE AND COMPETITIVENESS: CHALLENGES AND OPPORTUNITIES

The world faces multiple obstacles to tackle if big data is to become an integral part of trade and competitiveness policy formulation. Many challenges pertain to big data's scalability, as big data for economic development is still an emerging field. These challenges should be viewed as *opportunities* as much as *challenges.* Government and development institutions have a critical role to play in learning from the solutions in these proofs-of-concept.

Challenges to Accessing Sources of Big Data

Big data shows great promise as a tool to inform trade and competitiveness policy, but there are challenges to its operationalization that must be addressed. These challenges represent barriers to accessing the data sources needed for the work based on valid privacy or proprietary concerns. A big data environment can encompass a broad spectrum of sensitive information, including datasets of proprietary research information or intellectual property, datasets requiring regulatory compliance, and datasets of personally identifiable information (PII), which are all sources of information that are important to secure in a big data environment.[27]

Access and Privacy

Concerns about an individual's privacy, and compliance with policies governing individual privacy, present the greatest challenges to big data solutions. Though terms of service agreements are ubiquitous, individuals supplying the information used in big data solutions may be unaware that they are creating a digital trace. Thus, big data users must work to safeguard the security of an individual's privacy by ensuring anonymity in the generated information.

For example, Call Detail Records, a frequently utilized source of big data, include call time, call duration, caller and recipient cell tower locations, and most critically, Subscriber Identity Module (SIM) card identification data.[28] When used in the development context, this data is anonymized with random number translators. However, critics point out that this information, coupled with growing re-identification capabilities, could be harmful in the wrong hands. CDRs are but one example of the privacy concerns surrounding the rise of big data; social media, Internet search histories, and medical records sharing all create similar concerns.

Big data interventions must be implemented together with governance mechanisms to ensure that the interventions are ethical and respect the privacy of all those involved. Privacy in the age of big data will be a critical pillar of global policymaking, and it will be necessary to ensure privacy laws commute across national borders.

1. *Solution Spotlight:* Researchers at AT&T, a U.S. telecommunications corporation, collaborated with several universities to create an algorithm that injects structured "noise" into CDR models to mathematically obscure the data and mask individual data points, thereby protecting privacy without sacrificing fundamental insights.[29]

2. *Solution Spotlight:* U.S. President, Barack Obama chartered the President's Council of Advisors on Science and Technology to conduct an in-depth exploration of the intersection of big data and privacy to identify technologies that may disrupt current U.S. policies and jeopardize the privacy of citizens. Staying ahead of privacy breaches will be critical for a successful policy transition.[30]

Access to Proprietary Information

While big data solutions to inform trade and competitiveness policymaking are best served through the open access and integration of a multitude of datasets from reliable sources, there is a significant amount of proprietary data that is closely held by governments and corporations and that is often inaccessible. The apprehension of governments and corporations toward sharing proprietary information for public use is not without merit: the loss of this information may diminish security or a business's competitive advantage. However, these institutions hold big data sources that may be pivotal to enhancing the effectiveness of interventions, such as CDRs, satellite imagery, demographic information, and social media analytics.

For example, the real estate industry relies on privileged access of its employees to transaction data and insights into buyer behavior that is not openly available to consumers. Big data, however, has threatened their competitive advantage by democratizing this information through user-submitted information and parallel sources for real-estate data, thereby expanding the consumer's once limited access to real estate cost and pricing data.[31]

Business and government holders of useful data sources need to be convinced that the contribution of their proprietary information to big data solutions benefiting economic development efforts will not affect: (i) the security of that information; (ii) any competitive advantage they derive from being an exclusive user of that information; and (iii) their regulatory compliance with any applicable policies.

1. *Solution Spotlight:* Three popular travel navigation applications, Waze, Moovit, and Strava, collaborated with transportation planners for the 2016 Summer Olympics in Rio de Janeiro, Brazil, by providing user-generated data. The GPS data provided by Waze about drivers' movement histories, for example, separated users' names from their 30-day driving information using aliases to allay privacy and security concerns.[32]

Data Quality Still Affects Big Data Conclusions

Non-traditional data sources, including CDRs and satellite imagery, can create proxies for wealth distribution, unemployment, and other economic indicators. However, these proxies are often created by comparing alternative data to traditionally-sourced data (for example, consumption surveys), and using machine-learning algorithms to derive connections and construct a predictive model. The new model's predictive capacity is contingent upon the accuracy and quality of the original "ground-truth" data. Therefore, big data innovators in the public and private sector should not neglect their traditional data sources altogether, as they remain critical components to the predictive and prescriptive capacity of some big data solutions.

1. *Solution Spotlight:* Premise is a startup company that pays its users to send pictures of retail

products plus price information, and then aggregates the data into inflation indices and other alternative measures of economic activity. This simple application of analytics to traditional data enabled Premise to create a Brazilian Food Staples Index that predicts the inflation rate 25 days ahead of the release of official Brazilian statistics.[33]

Building Trust is Critical to Big Data-Driven Policy

Big data solutions leverage innovations from various fields, including information from agricultural science, geology, finance, economics, and rigorous data analytics. This multi-dimensionality is undoubtedly a strength, but it also means that solutions are limited by their weakest componential input. For example, tools that employ satellite data to map manufacturing activity may feature incredibly adept algorithms, but the ultimate applicability of these tools can be limited by insufficient satellite coverage in remote areas. Moreover, transforming insights drawn from big data solutions into policy requires a knowledge of the economic, political, and social context. For example, research being done to map poverty with mobile phone data is only actionable if policymakers trust that the implications of this data are grounded in reality.

The interdisciplinary nature of big data solutions necessitates close stakeholder collaboration, not just in terms of data sharing, but also in terms of goal-setting, project planning, and solution scaling. The rise of data cooperatives is a first step in moving beyond simply sharing information to a livelier exchange across public/private boundaries to solve the toughest public sector problems.

2. *Solution Spotlight:* In 2015, the United States National Oceanic and Atmospheric Administration, a U.S. government scientific agency, enlisted the support of U.S. technology companies Amazon, Google, IBM, and Microsoft, to make its data, which amount to about 20 terabytes daily, publicly available.[34]

3. *Solution Spotlight:* In 2015, the U.S. announced the creation of the Precision Medicine Initiative Cohort Program, a data collaborative, which aggregates volunteered electronic health records of individuals, research organization findings, and clinical trial results to improve approaches to disease prevention and treatment.[35]

Drawing Investment in a Data Science Workforce

In a survey of almost 500 data scientists, the EMC Corporation, a U.S. technology firm, found that 64 percent believed that the demand for big data skills will outpace the supply of relevant talent in coming years.[36] Much of the big data work that is currently underway, both in research and public and private sector enterprises, is conducted in industrialized economies. Claire Melamed, Director of Poverty and Inequality at the Overseas Development Institute, has expressed concern that, "At the moment, the explosion of big data has far-outpaced our ability to make sense of it in all countries, but most of all in poorer nations that already lack human and technical capacity."[37] The data science skills gap for low- and middle-income countries must be addressed and will require planning, resource allocation, and ongoing commitment on the part of both governments and the private sector.

Implementing big data solutions in development contexts must be complemented by policy initiatives, including data science skills development, that build the capacity to sustain the solutions. This will ensure ownership for local and national actors and ensure that big data-driven insights remain substantial and beneficial in the long term.

Solution Spotlight: The World Bank Group Institute, in partnership with the African Media Initiative, hosts "data bootcamps" around the world, including in Malawi, Tanzania and South Africa. During these sessions, members of civil society organizations and governments come together to learn basic data analysis techniques, including how to contextualize raw data and apply it to solve real-world challenges.[38]

SECTION 6

LOOKING AHEAD: THE FUTURE OF BIG DATA FOR TRADE AND COMPETITIVENESS

Despite the challenges, there are plenty of reasons to be hopeful about the evolving role of big data solutions in trade and competitiveness policy. Through innovative approaches to collecting and analyzing economic data, big data has opened new avenues for policymakers to gather the information necessary to understand a developing economy's local economic activity; foster increased investment; resolve logistical barriers and promote competitive supply chain management; and increase the competitiveness of cities and the poor. As the methods to harness big data grow more sophisticated, so too will the quality of information that can be gathered to identify economic opportunity areas, formulate effective trade and investment policy, and design development interventions that ultimately improve conditions for the world's most disadvantaged populations.

The following are considerations to ensure that big data continues to grow as a promising catalyst of inclusive economic growth and effective trade and competitiveness policymaking.

Supporting Future Solutions for Development

The utility of big data solutions is widely recognized. However, remembering that these techniques are

rendered far more powerful when applied collaboratively, including with traditional economic data sources.

Consider a 2015 initiative undertaken by the New York City (NYC) Association for Neighborhood and Housing Development to create an interactive analysis of key economic indicators, including figures related to poverty, infrastructure, and the business climate (for example, the percent of at-risk small businesses). The data was drawn from a combination of traditional and big data sources, including the U.S. Census Bureau's 2013 American Community Survey, longitudinal employment statistics, the NYC Open Data portal, and the NYC Department of Finance figures.[39]

For future solutions for trade and competitiveness policy, one can imagine more sophisticated tools drawing on diverse data sources, including:

Leveraging geo-tagged jobs data to examine the distribution of skills supply and demand

By comparing the distribution of skills associated with both job postings and worker profiles, governments and development organizations could map the human capital landscapes of cities. They could then design policies and programs to bridge any gaps by *facilitating* connections between employers and potential employees or by targeting technical and vocational training programs, for example.

Using on-the-ground imagery to gather income, capital, and innovation distribution data

Researchers with the National Bureau of Economic Research are exploring the use of Google Street View to predict household income in New York City based on data captured in the form of textures, colors, and shapes in its 360-degree pictures. The algorithm has been relatively successful: in 2015, it predicted 77 percent of income variation, whereas a combination of race and education predicted just 25 percent. In the future, by combining this ground-level analysis with other proxies for innovation (for example, business records, electrification, LinkedIn data), policymakers could keep a more vigilant watch over where individuals and enterprises are thriving and why.[40]

A New Role for Governments in Big Data

To bring about a shift towards data-driven decision making processes, governments must begin to think of data not just as a tool for competitive analysis, but as an asset in and of itself. What this means is that governments must work to facilitate policy environments that enable the use of big data as a productive tool for economic competitiveness and for society generally. This necessitates careful policy considerations across several factors:

1. **Enabling Skills Development:** Governments should facilitate workforce development initiatives in low- and middle-income countries that contribute to the data science skill sets necessary to operationalize big data for insight-driven decision making. Initiatives to increase information and communications technologies (ICT) penetration into these economies would also foster big data solutions.

2. **Making Big Data Available:** Governments should work to facilitate open, responsible access to data. Creating robust data governance frameworks in which big data for public good is accessible to both public and private sector users would enable the greatest opportunity for applications that significantly impact low- and middle-income countries.

3. **Collaborating with the Private Sector:** Governments should explore and if possible, foster, collaborative efforts between businesses and institutions using big data for development. For example, credit card and healthcare companies discard roughly 80 percent of their data. Although there may be privacy reasons for this, anonymization techniques could turn this data into actionable information for governments to inform socio-economic policies.[41]

While some of these solutions for trade and competitiveness policy may seem distant today, they can be made attainable by governments taking action to facilitate the foundational policies, skill sets, and collaborative environments needed to craft creative big data solutions to trade and competitiveness challenges.

REFERENCES

1 Sam Ro, "Chart of the Day: The World is More Open Than Ever," *Business Insider*, November 7, 2013, http://www.businessinsider.com/world-exports-to-gdp-ratio-2013-11.

2 Yann Duval and Chorthip Utoktham, "Impact of Trade Facilitation on Foreign Direct Investment," United Nations ESCAP Trade & Investment No. 4 (2014), http://www.unescap.org/sites/default/files/Staff%20Working%20Paper%2004-14_0.pdf.

3 World Trade Organization, "World Trade Report 2015," WTO (2015), https://www.wto.org/english/res_e/booksp_e/world_trade_report15_e.pdf.

4 Klaus Schwab, "Global Competitiveness Report 2014–2015," *World Economic Forum* (2014), http://www3.weforum.org/docs/WEF_GlobalCompetitivenessReport_2014-15.pdf.

5 "The Global Competitiveness Report 2015-2016: Report Highlights,'" World Economic Forum (2015), http://reports.weforum.org/global-competitiveness-report-2015-2016/report-highlights/.

6 World Bank Group, "Economy Rankings," http://www.doingbusiness.org/rankings

7 "The Global Competitiveness Index 2015–2016," World Economic Forum (2015), http://reports.weforum.org/global-competitiveness-report-2015-2016/the-global-competitiveness-index-2015-2016/.

8 D.V. Tran and N.V. Nguyen, "The Concept and Implementation of Precision Farming and Rice Integrated Crop Management Systems for Sustainable Production in the Twenty-first Century," Integrated Systems, http://www.fao.org/3/a-a0869t/a0869t04.pdf.

9 World Bank Group, "Doing Business Report Finds More than 60% of World's Economies Improved Their Business Rules in Past Year," October 27, 2015, http://www.worldbank.org/en/news/press-release/2015/10/27/doing-business-report-finds-more-than-60-of-worlds-economies-improved-their-business-rules-in-past-year.

10 Ancor Suarez-Aleman, Javier Morales Sarriera, Tomas Serebrisky, and Loudres Trujillo, "When it Comes to Container Port Efficiency, Are All Developing Regions Equal?" Inter-American Development Bank (2015), https://publications.iadb.org/bitstream/handle/11319/6788/IDB-WP-568.pdf.

11 World Bank Group, "Great Lakes Project to Help African Traders Get Their Goods and Services to Market," September 25, 2015, http://www.worldbank.org/en/news/feature/2015/09/25/great-lakes-project-to-help-african-traders-get-their-goods-and-services-to-market.

12 United Nations, "World's Population Increasingly Urban with More than Half Living in Urban Areas," July 10, 2014, http://www.un.org/en/development/desa/news/population/world-urbanization-prospects-2014.html.

13 World Bank Group, "Urban Development," http://data.worldbank.org/topic/urban-development.

14 United Nations Population Fund (UNFPA), "Population Dynamics in the Least Developed Countries: Challenges and Opportunities for Development and Poverty Reduction," https://www.unfpa.org/sites/default/files/pub-pdf/CP51265.pdf.

15 Kasper Worm-Petersen, "Data as a Mitigator in Slum Development," June 23, 2014, http://graspmag.org/urbanism/urban-informality/data-mitigator-slum-development/.

16 August Sjuaw-Koen-Fa and Inez Vereijken, "Access to Financial Services in Developing Countries," Rabobank (2015), https://economie.rabobank.com/PageFiles/3584/access_tcm64-75165.pdf.

17 Keith Soura and Peter Goodings Swartz, "Big Data & International Trade: Creating Transparency through Information," October 17, 2014, https://www.youtube.com/watch?v=HZlDw6eU55Y.

18 Joab Jackson, "IBM and Deloitte Bring Big Data to Risk Management," *Computer World, May* 18, 2015, http://www.computerworld.com/article/2923150/big-data/ibm-and-deloitte-bring-big-data-to-risk-management.html.

19 Adarsh Desai, "Meet the Winners and Finalists of the First WBG Big Data Innovation Challenge!", December 18, 2014, http://blogs.worldbank.org/voices/meet-winners-and-finalists-first-wbg-big-data-innovation-challenge.

20 Marzia Rango, "How Big Data Can Help Migrants," last modified October 5, 2015, https://www.weforum.org/agenda/2015/ 10/how-big-data-can-help-migrants/.

21 Leora Klapper, Douglas Randall, and Jenny Marlar, "Income Biggest Barrier to Banking in Developing Countries," December 19, 2012, http://www.gallup.com/poll/159380/income-biggest-barrier-banking-developing-countries.aspx.

22 Ibid.

23 Bloomberg News, "Hedge Funds Look to Space with New China Economy Gauge," March 13, 2016, http://www.bloomberg.com/news/articles/2016-03-13/hedge-funds-look-to-space-with-new-china-economy-gauge.

24 Peter Buxbaum, "Modern Supply Chains Benefit from the Internet of Things," *Global Trade Magazine,* January 30, 2016, http://www.globaltrademag.com/global-trade-daily/commentary/modern-supply-chains-benefit-from-the-internet-of-things.

25 Lucas Mearian, "Self-Driving Cars Could Create 1GB of Data a Second," *Computer World,* July 23, 2013, http://www.computerworld.com/ article/2484219/emerging-technology/self-driving-cars-could-create-1gb-of-data-a-second.html.

26 William Herkewitz, "Just a Few Self-Driving Cars on the Highway Could Cut Random Traffic Jams by Half," *Popular Mechanics,* October 8, 2015, http://www.popularmechanics.com/cars/a17718/just-a-handful-of-self-driving-cars-on-the-highway-could-cut-traffic-jams-by-half/.

27 Jeff Markey, "How to Manage Big Data's Big Security Challenges," May 13, 2014, http://data-informed.com/manage-big-datas-big-security-challenges/.

28 Emmanuel Letouzé and Patrick Vinck, "The Politics and Ethics of CDR Analytics," December 10, 2014, http://static1.squarespace.com/static/531a2b4be4b009ca7e474c05/t/54b97f82e4b0ff9569874fe9/1421442946517/WhitePaperCDRsEthicFrameworkDec10-2014Draft-2.pdf.

29 David Talbot, "How to Mine Cell-Phone Data Without Invading Your Privacy," *Technology Review,* May 13, 2013, https://www.technologyreview.com/s/514676/how-to-mine-cell-phone-data-without-invading-your-privacy/.

30 John Podesta, "Big Data and the Future of Privacy," January 23, 2014, https://www.whitehouse.gov/blog/2014/01/23/big-data-and-future-privacy.

31 Tim McGuire, James Manyika, and Michael Chui, "Why Big Data is the New Competitive Advantage," *Ivey Business Journal,* (July/August 2012), http://iveybusinessjournal.com/publication/why-big-data-is-the-new-competitive-advantage/.

32 Parmy Olson, "Why Google's Waze Is Trading User Data with Local Governments," *Forbes,* July 7, 2014, http://www.forbes.com/sites/parmyolson/2014/07/07/why-google-waze-helps-local-governments-track-its-users/

33 "Pioneering a Set of Alternative Economic Indicators in Global Markets," 2016, http://www.premise.com/case/#2.

34 Office of Public Affairs, "U.S. Secretary of Commerce Penny Pritzker Announces New Collaboration to Unleash the Power of NOAA's Data," April 21, 2015, https://www.commerce.gov/news/press-releases/2015/04/us-secretary-commerce-penny-pritzker-announces-new-collaboration-unleash.

35 National Institutes of Health, "About the Precision Medicine Initiative Cohort Program," https://www.nih.gov/precision-medicine-initiative-cohort-program.

36 "Data Science Revealed: A Data-Driven Glimpse into the Burgeoning New Field," http://www.emc.com/collateral/about/news/emc-data-science-study-wp.pdf.

37 Jan Piotrowski, "Big Obstacles Ahead for Big Data for Development," April 15, 2014, http://www.scidev.net/global/data/feature/obstacles-big-data-development.html.

38 oAfrica, "Malawi's First Data Literacy Bootcamp Tackled Voter Participation, Healthcare Ratings, Food Security, and More," June 25, 2013, http://www.oafrica.com/education/malawis-first-data-literacy-bootcamp-tackled-voter-participation-healthcare-ratings-food-security-and-more/.

39 Edward L. Glaeser, Scott Duke Kominers, Michael Luca, and Nikhil Naik, "Big Data and Big Cities: The Promises and Limitations of Improved Measures of Urban Life," The National Bureau of Economic Research (2015), http://www.nber.org/papers/w21778.

40 Bourree Lam, "Can Google Street View Images Predict Household Income?", *The Atlantic,* December 10, 2015, http://www.theatlantic.com/business/archive/2015/12/big-data-google-street-view-income/419214/.

41 Martin Hilbert, "Big Data for Development: From Information to Knowledge Societies," Social Science Research Network (2013), https://papers.ssrn.com/sol3/papers2.cfm?abstract_id=2205145