

Regional cooperation in big data

Data collected in many less developed countries have serious problems and shortcomings. Big data offer tremendous opportunities to improve the effectiveness of public policies.

Zubair Faisal Abbasi

“Land was the raw material of the agricultural age. Iron was the raw material of the industrial age. Data is the raw material of the information age.”

The Industries of the Future by Alec Ross, p 152

Data collection and analysis is an established technique of understanding the world through research. South Asia has centuries-old tradition of data collection to gather information for government interventions. Data about prices, living conditions of people and also about social, political and economic histories of administrative territories were collected during the British Raj. These data were made part of official district gazettes, which were published periodically. With the passage of time, the spectrum of data collection expanded and many social and economic indicators were added in different kinds of surveys, such as integrated household surveys, demographic and health surveys, industrial and agriculture census and population census. Governments in South Asia also tried to increase both institutional and technological capacity of their statistical organizations and line departments with varying degrees of success and failure.

Less developed data

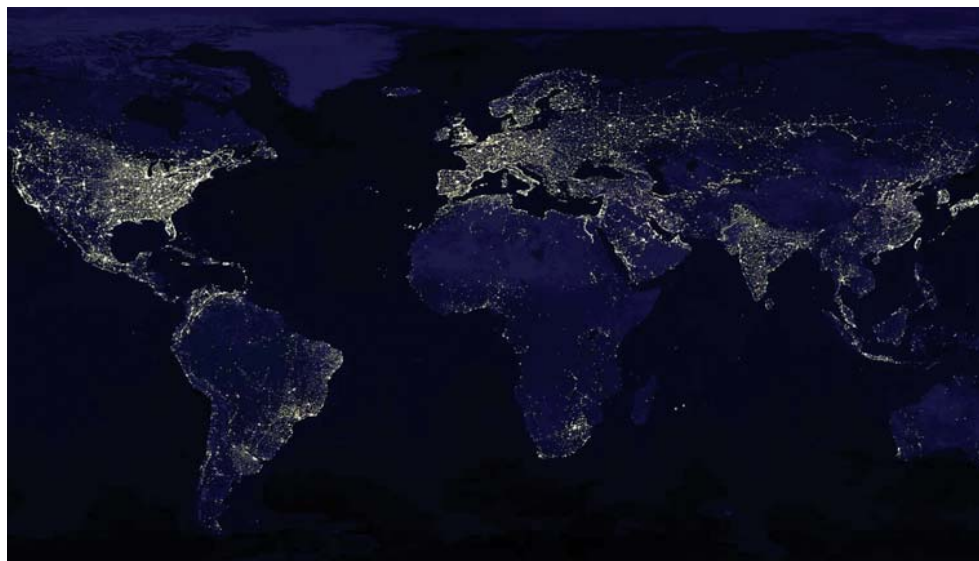
However, the data collected at the state level also attracted critique of public policy experts. It was argued that the data collected in many less developed countries have serious problems and shortcomings.¹ Some of the issues related to data are serious shortage of data for analysis and policy advice², and late data processing and dissemination, sometimes with a lag of two to three years, by which time issues have changed and the data do not reflect the true picture on the ground for policy action. Similarly, there are many issues like poor security situation, inaccurate responses, lack of logistical support to access remote locations, existence of informal economy, limited capacity of enumerators and quality assurance mechanisms, low budgetary allocations and a sheer lack of autonomy in the statistical offices. Moreover, most of the data which are collected anyway are not designed to inform policy analysis, and most of the data are not shared with public.

It has been argued that one of the key reasons for United States of America (USA) being an ideal country for economic and social research is the availability of “troves of good quality data”. The reverse is true for South

Asian countries, where there are significant internal and external problems pertaining to statistical offices.

With the passage of time, and with the introduction of modern information and communication technologies (ICTs), such as satellite images, videos, text messages, internet, emails, biometric systems and social media websites, a new source of data has emerged. These are a part of big data, which promise to show “how a large amount of data can now be used to understand, analyze, and forecast trends in real time”.³ One of the biggest and most successful uses of big data for politics was undertaken during the two presidential campaigns of President Obama. In the data analysis then, emails were analyzed in real time to understand which variant of email messages was able to generate more funding. It made the campaign one of the most effective.⁴

The World Bank, by using satellite to analyze the nightlights at different locations of South Asia, has shown a correlation between economic growth and nightlights. The World Bank’s South Asia Unit uses nightlight data to indicate how overall urban land use is increasing over time. Not only are Indian cities growing in size, the intensity of activity within certain



elektika.cz

urban regions, such as Coimbatore in Tamil Nadu, is growing as well. The areas registering the fastest rates of growth are those in close proximity to existing cities—Delhi, Hyderabad, Lahore, Mumbai, etc.⁵ Nightlight shows economic activity, prosperity and vehicular movements, thus portraying the level of production of goods and services. An interesting aspect of the nightlight indicator is that it is the same everywhere, hence, “an excellent indicator of economic activity”⁶.

There are many other sources of big data, such as cellular companies that have record of users’ text messages and calls data. There is whole range of social media providing an enormous amount of data on individual preferences and activities. All these sources of data can be and need to be utilized for economic measurement.

Pakistan has also started seriously thinking and using big data with various degrees of success. For example, there are evidences that telecommunication companies in Pakistan are investing in technologies associated with big data and deep analytics. The drive is coming from multinational companies that want to understand the markets and their customers in the face of market competition.⁷ In addition, the global market players are also picking the best brains from Pakistan for data science advancements.

Pakistan is also witnessing a government and private sector cooperation in big data. Teradata, a US-based firm, works closely with National Database and Registration Authority (NADRA) in the analysis of demographics, supports intelligence and provides help in crime investigations. The Government of Punjab, with the help of Teradata, also worked on a dengue fever epidemic and successfully used big data analytics to understand the situation and develop interventions to control the deadly menace. It has been reported that the interventions were successful. Since then, the results and methods of big data use have been shared with other countries too.

Another set of notable interventions in the pipeline is coming from Sindh Province, where land records and electricity metering systems are being digitized. It is being reported in the press that data on electricity use and land ownership shall help improve economic governance, as well as create better policy actions for tax reforms (working with the Federal Board of Revenue (FBR)) and economic incentives.⁸

Data gap

Interviews with officials of Pakistan Bureau of Statistics (PBS) have revealed that the Bureau was sending its employees for training to USA and

other countries to learn and implement new technologies. Some efforts have been made at the PBS to create a common database for different surveys, NADRA, Benazir Income Support Programme (BISP) and FBR. The effort is to remove data gaps and create both synergies and “single window” analytics. As a result, Pakistan, like other countries, is poised to have much more data available and also the ability to chew through the data.

A good example of the opportunities presented by good data is the Survey of Well-being via Instant and Frequent Tracking (SWIFT) Project of the World Bank, where small survey and big data are being simultaneously used for poverty estimation, viz:

“Like typical “Big Data” approaches, SWIFT applies a series of formulas/algorithms, as well as the latest ITS technology, to cut the time and cost of data collection and poverty estimation. For example, SWIFT does not estimate poverty from consumption or income data, which is time-consuming to collect, but uses formulas to estimate poverty from poverty correlates, which can be easily collected. Furthermore, by embedding the formulas into the SWIFT data management system, the correlates will be converted to poverty statistics instantly. To further cut the time for data collection and processing, SWIFT uses Computer Assisted Personal Interview (CAPI) linked to data clouds, and if possible, adopts a cell phone data collection approach.”⁹

These are projects to help South Asian countries complement their existing practices of poverty estimation. These advancements can help improve the effectiveness of poverty reduction programmes as well as increase the efficiency of policy actions. The most important point is that South Asian countries must learn to use big data in the context of measuring progress towards their Sustainable Development Goals (SDGs), not to mention developing targeted interventions. In this way, big data analytics can help governments execute midcourse corrections for better development outcomes.

There is a potential to learn from different country experiences, as

well. However, looking at the current state of cooperation in the South Asia Association for Regional Cooperation (SAARC) region, it seems advisable that multinational companies and international financial institutions, such as the World Bank, should step in. They can work closely with the governments, private sector and civil society organizations and help them in the use of big data analytics for social and economic development—health, education, gender-based violence, deforestation, crime control and what not. This should set a stage and create an eco-system of cooperation.

As data are increasing, so are the complexities. It is said that every nine seconds one petabyte of data is added in the global virtual repository. This shows the enormous rate of increase in both the volume and velocity of data. At the same time, variety in data has also increased manifold. There are emails, videos, pictures, text messages, social media content and calls, in addition to the existing databases of economic indicators. Most of the data are unstructured and multilingual, which need filtering and storage to make them usable for analytics and then retrieval and discussion.¹⁰ However, big data technologies are fast solving the problems of inter-language communication. Real time auto-translators have almost demolished the language barrier.

The good news is that the emerging technologies, both hardware and software, are not only capable but increasingly being enabled to neatly undertake analytics and generate reports needed for measurements. We do not need super computers to develop and use algorithms for big data analytics, because computing power and software functionality are increasing very fast.

In South Asia, the current state of statistical cooperation needs a thorough rehashing of policies and practices related to data management. National governments need to understand that governments and public policies now need rapid action to stimulate responses in economic and

social development fields. Time lags in stimulus and response create more problems than they solve. In fact, big city management in South Asia really needs smart data collection, interpretation and reporting platforms. Much of the problems, such as poverty and hunger, which has endangered South Asia, can be solved with financial technology and smart precision agriculture, which are nothing but big data-based interventions. Big data are said to be both a telescope and a microscope.

Regional data

Although there are serious security concerns over data protection and data governance, there are solutions in new technologies as well. Countries need to explore security options so that confidence level is achieved for further advancements. South Asian countries need to harness those technologies and bring in private, public and the third sector to rebuild and strengthen institutional arrangements for sound evidence-based and policy-relevant economic measurement systems.

South Asia's need for assistance from developed countries must be looked into and efforts made to keep brain circulation within, as opposed to brain drain from, the region for cross-border learning. A real South-South cooperation can help change the economic measurement and policy action landscape in the region. In this context, some quick and handy recommendations are listed below:

- SAARC countries should develop a programme for big data in their human resource development units;
- Scoping studies of regional advances are needed in big data uses, advancements and plans;
- Governments must jointly, and in a standalone fashion, engage in open dialogue with multinational private sector firms to assist their national statistics bureaus;
- There should be linkages with development agencies, such as the World Bank and social and economic research organizations,

to raise awareness about new technologies and new data; and,

- SAARC countries must organize serious dialogue and institute a high-level commission to report on advances in big data in the region and how intra- and inter-regional cooperation is possible. ■

The author is Director, Institute for Development Initiatives (IDI), Islamabad.

Notes

- 1 Elahi, Asad. 2008. "Challenges of Data Collection: with Special Regard to Developing Countries". In *Statistics, Knowledge and Policy 2007: Measuring and Fostering the Progress of Societies*. Paris: OECD Publishing. <http://www.pbs.gov.pk/sites/default/files/articles/Elahi%20Article%20-%20For%20FBS%20Homepage.pdf>.
- 2 Haider, Murtaza. 2014. "Desperately seeking data in Pakistan...". *Dawn* June 4. <https://www.dawn.com/news/1110537>
- 3 Ross, A. 2016. *The Industries of the Future*. London: Simon & Schuster UK.
- 4 *ibid.* p. 155.
- 5 Mukim, Meghna. 2013. "Tracking light from space: Innovative ways to measure economic development". *World Bank Blogs* December 11. <https://blogs.worldbank.org/category/tags/nightlight>.
- 6 Henderson, J. Vernon, Adam Storeygard, and David N. Weil. 2012. "Measuring Economic Growth from Outer Space." *American Economic Review*, 102(2): 994-1028. http://www.econ.brown.edu/Faculty/David_Weil/Henderson%20Storeygard%20Weil%20AER%20April%202012.pdf.
- 7 Javed, Hassan. "Big Data Challenges and Opportunities". *Pakistan Advertisers' Society*. <http://www.pas.org.pk/6375-2/>.
- 8 Haq, Shahram. 2016. "Tech expert dives into big data potential in Pakistan". *The Express Tribune* December 21. <https://tribune.com.pk/story/1151447/telecommunication-tech-expert-dives-big-data-potential-pakistan/>
- 9 Yoshida, Nuobo. 2014. "Revolutionizing Data Collection: From 'Big Data' to 'All Data'". *World Bank Blogs* November 12. <http://blogs.worldbank.org/developmenttalk/revolutionizing-data-collection-big-data-all-data>.
- 10 Nawsher Khan, Ibrar Yaqoob, Ibrahim Abaker Targio Hashem, et al. 2014. "Big Data: Survey, Technologies, Opportunities, and Challenges". *The Scientific World Journal* vol. 2014, Article ID 712826. doi:10.1155/2014/712826.