



Big Data and the Modernization of Statistical Systems in Developing Countries

Olivier Dupriez*

World Bank, Washington, DC, USA– odupriez@worldbank.org

Grant J. Cameron

World Bank, Washington, DC, USA– gcameron1@worldbank.org

The demand for socio-economic data in developing countries is constantly growing, driven by societies requiring greater accountability from their leaders. Much of the produced data come from surveys conducted by government agencies. Availability and public access to these data is improving; but not quickly enough. Data gaps, as well as issues of relevance, timeliness and accuracy remain daunting for many low-income countries. This has prompted a call for a “Data Revolution”. The 2014 *Report to the United Nations Secretary General by the Independent Expert Advisory Group on a Data Revolution for Sustainable Development* called for exploring and exploiting new technologies and data sources to transform the way we measure and monitor development. In this context, the use of big data has raised much interest and high expectations from official statisticians.

There is strong evidence of the value of big data for policy planning and monitoring. But reports of successful use of big data typically originate from data scientists who sought to answer questions and address issues that are out of the current mandate of official statisticians. The use of big data in statistical organizations—usually in high income countries—is still in a piloting stage. Some of these organizations have started sharing their experiences, and are providing critically needed information not only on successes, but also on less-encouraging results, on the risks and costs involved, on skills and technological constraints, and on legal and ethical hurdles. Their reports confirm the potential of big data to supplement traditional data collection methods and to derive new indicators, but often point to caveats and a significant entry cost.

The paper draws lessons from experiences, with a focus on the needs and constraints of official statisticians in developing countries. It calls for a modernization of statistical systems more than for a revolution. It argues that improving the quality, timeliness, accuracy and public availability of traditional data collections – surveys, censuses – should remain a priority, as these datasets are key inputs for big data applications that rely on machine learning algorithms. Use of big data for the production of official statistics, and possibly for the creation of new services that statistical agencies could provide to their stakeholders, will require a coordinated incubation phase in which development partners will play an important role. And a long-term capacity building program should contribute to creating a new cohort of data scientists.

Keywords: data revolution, capacity building, official statistics