Applications of Big Data for the Sustainable Development Goals: Enlightening Policy Makers on the State of the Art

CPRsouth2017 **Policy Brief**

The unprecedented data requirement for the Sustainable Development Goals (SDGs), in terms of breadth, scope and indicators, coupled with a call to harness a data revolution that leverages new data sources in addition to existing ones, underscores potential for big data sources (including mobile data, social media and satellites) to support the measurement of SDGs. The full potential of big data would lie in its use along with more traditional measures of statistics, looking at situations more from the perspective of what needs to be measured rather than what data is to be used. However, there are concerns around privacy, provenance, data access, representativity and harms. This policy paper serves to better inform policymakers of the range of insights that can be extracted from big data sources, and how they can be leveraged to better position countries to measure and monitor the SDGs. It does this by reviewing the current body of literature that has explored the utilization of various big data sources for specific developmental applications. The paper also touches on the discussion around concerns of the use of big data including privacy.

APPLICATIONS OF NON-TRADITIONAL DATA SOURCES

Mobile Phone Data: Variables relating to mobility, connectivity as well as economic status that can be derived from mobile network data have important development applications that range from identifying the movement of population, to the spread of diseases as well as estimating the socioeconomic status of a population.

Remote Sensing Data: Data derived from satellite imagery are particularly useful to identify changes in topography over space and time. Variables derived from the analysis of this data have applications in early crop yield estimation, drought monitoring, monitoring changes in deforestation, desertification and changes in water-related ecosystems, all relevant to various SDG targets.

Social Media Data: The analysis of social media engagement of various populations can provide a pulse of that group in relation to various issues of interest.

THE RESEARCH

I LITERATURE/PRIOR WORK

There are numerous studies on that explore the application of big data insights for development. These range from the analysis of mobile network big data for applications related to poverty (Frias-Martinez et al. 2012; Gutierrez et al. 2013; Blumenstock et al. 2015). In addition with regards to health, studies have sought to understand the spread of mosquito-borne diseases such as dengue (Wesolowski et al. 2015) and malaria (Tatem et al. 2014) while others have sought to predict crime (Bogomolov et al. 2014).

Moreover, variables derived from the analysis of remote sensing data can be used to provide another angle of insights that range from leveraging nighttime luminosity to estimate poverty (Elvidge et al. 2009), economic development (Elvidge et al. 1997; Sutton et al. 2007) and access to electricity (Townsend & Bruce, 2010). Additionally, identifying changes in topography can provide valuable insights on the change in environment over space and time which can be applied to monitor drought conditions (Rhee & Carbone, 2010), and changes in surface water (Haas et al. 2009).

II METHOD

This paper is based on a review of existing literature on non-traditional Big Data. As such, the method of study has been desk research. Some of the main sources have been referenced here.

IV OBSERVATIONS AND POLICY RECOMMENDATIONS

There is a need for timely, disaggregated and accurate data to measure and monitor the progress towards achieving the targets associated with the SDGs. There is potential for statistics office to leveraged non-traditional data sources such as mobile phone data, satellite-based technology and social media data among others to generate insights that can be used to support this. Some policy recommendations to facilitate this includes:

Focus on what needs to be measured, rather than on the data source. Instead of replacing traditional sources, the greatest potential for big data at the moment lies in complementing traditional data, exploring solutions from an issue perspective rather than a data source perspective – thus, measuring an indicator may entail the use of multiple data sources, both traditional and big data.

View insights from big data as providing a sense of directionality as opposed to stand-alone measures. For instance, the insights generated from big data sources could be compared to low-resolution video while more traditional methods may produce a high-resolution photograph, both useful.

Expand the data ecosystem to include other entities that produce/generate data (including private sector, academia, and civil society) outside the traditional system. For example, big data could be held in the hands of the private sector, for instance mobile network operators.

III SOURCES

Blumenstock, J., Cadamuro, G., & On, R. (2015). Predicting poverty and wealth from mobile phone metadata. Science, 350(6264), 1073-1076.

Bogomolov, A., Lepri, B., Staiano, J., Oliver, N., Pianesi, F., & Pentland, A. (2014, November). Once upon a crime: towards crime prediction from demographics and mobile data. In Proceedings of the 16th international conference on multimodal interaction (pp. 427-434). ACM.

Elvidge, C. D., Baugh, K. E., Kihn, E. A., Kroehl, H. W., Davis, E. R., & Davis, C. W. (1997). Relation between satellite observed visible-near infrared emissions, population, economic activity and electric power consumption. International Journal of Remote Sensing, 18(6), 1373-1379.

Elvidge, C. D., Sutton, P. C., Ghosh, T., Tuttle, B. T., Baugh, K. E., Bhaduri, B., & Bright, E. (2009). A global poverty map derived from satellite data.Computers & Geosciences, 35(8), 1652-1660.

Frias-Martinez, V., Virseda-Jerez, J., & Frias-Martinez, E. (2012). On the relation between socio-economic status and physical mobility. Information Technology for Development, 18(2), 91-106.

Gutierrez, T., Krings, G., & Blondel, V. D. (2013). Evaluating socio-economic state of a country analyzing airtime credit and mobile phone datasets. arXiv preprint arXiv:1309.4496.

Haas, E. M., Bartholomé, E., & Combal, B. (2009). Time series analysis of optical remote sensing data for the mapping of temporary surface water bodies in sub-Saharan western Africa. Journal of Hydrology, 370(1), 52-63.

Rhee, J., Im, J., & Carbone, G. J. (2010). Monitoring agricultural drought for arid and humid regions using multi-sensor remote sensing data. Remote Sensing of Environment, 114(12), 2875-2887.

Sutton, P. C., Elvidge, C. D., & Ghosh, T. (2007). Estimation of gross domestic product at sub-national scales using nighttime satellite imagery. International Journal of Ecological Economics & Statistics, 8(S07), 5-21.

Tatem, A. J., Huang, Z., Narib, C., Kumar, U., Kandula, D., Pindolia, D. K., ... & Lourenço, C. (2014). Integrating rapid risk mapping and mobile phone call record data for strategic malaria elimination planning. Malaria journal, 13(1), 1.

Townsend, A. C., & Bruce, D. A. (2010). The use of night-time lights satellite imagery as a measure of Australia's regional electricity consumption and population distribution. International Journal of Remote Sensing, 31(16), 4459-4480.

Wesolowski, A., Qureshi, T., Boni, M. F., Sundsøy, P. R., Johansson, M. A., Rasheed, S. B., ... & Buckee, C. O. (2015). Impact of human mobility on the emergence of dengue epidemics in Pakistan. Proceedings of the National Academy of Sciences, 112(38), 11887-11892.

AUTHORS

Thavisha Perera-Gomez | Research Manager, LIRNEasia | 12 Balcombe Place, Colombo 00800, Sri Lanka | Tel Fax +94 773 163 398 | thavisha@lirneasia.net | www.lirneasia.net

Sriganesh Lokanathan | Team Lead – Big Data Research, LIRNEasia | 12 Balcombe Place, Colombo 00800, Sri Lanka | Tel Fax +94 77 762 6902 | sriganesh@lirneasia.net | www.lirneasia.net