

Leveraging Big Data Sources to Support the Measurement of the Sustainable Development Goals

Thavisha Perera-Gomez, Sriganesh Lokanathan¹

ABSTRACT

The unprecedented data requirement for the SDGs, in terms of breadth, scope and indicators, coupled with a call to harness a data revolution drawing on existing and new data sources, underscores potential for big data sources (including mobile data, social media and satellites) to support 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 – thus a multisource approach leveraging both traditional and non-traditional data could make the use of big data more effective. However, concerns of privacy cloud the discussions around big data, as does the lack of a governance structure surrounding the use of big data such as mobile network big data. Moreover, there are other concerns of possible harms surrounding the use of big data for development. 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 governance and privacy.

INTRODUCTION

Adopted in September 2015, the SDGs or the global calls, represent a worldwide call to action to transform the world by 2030 through the achievement of 17 strategic goals that span from ending poverty to building global partnerships for the goals, and cut across six areas: people, prosperity, planet, peace, and partnership. The SDGs, which officially came into effect in January 2016, are unprecedented in terms of its scope and universal application to all countries. While its predecessor, the Millennium Development Goals covered 8 goals measured by 60 indicators, the SDGs cover 17 goals, 169 targets to be measured by around 232 unique indicators.

Monitoring the progress towards achieving these SDGs would require data available at global, regional national and further disaggregated geographic levels. This has placed increasing burden on the statistical communities of countries to generate timely, reliable, disaggregated (sex, age, gender, race etc.) and accurate data, with member states placing an emphasis on strengthening capacity building of their statistical systems to monitor progress. Moreover, in order to monitor progress, there is a need for data that is comparable over time. However, traditional measure such as surveys are costly and take a lot of time.

This is further complicated by the lack of data to measure some of these indicators. The Inter-agency and Expert Group on SDG Indicators (IAED-SDG) categorizes the indicators into three tiers based on availability of data and relevant methodology, with Tier III being those indicators that do not yet have set methodologies for measurement. As of April 2017, there were 84 Tier III Indicators. This has raised interest in new data sources to complement data gaps. The UN recommends “harnessing a data revolution for sustainable development” that integrates traditional data with new data sources. The UN defines the data revolution as:

“An explosion in the volume of data, the speed with which data are produced, the number of producers of data, the dissemination of data, and the range of things on which there is data, coming from new technologies such as mobile phones and the “internet of things”, and from other sources, such as qualitative data, citizen-generated data and perceptions data; A growing demand for data from all parts of society.” (A World that Counts, 2014)

Launched at the UN World Data Forum in January 2017, the Cape Town Global Action Plan for Sustainable Development Data outlines six strategic areas (as well as associated objectives) to better position countries to generate relevant statistics. One such objective calls for mainstreaming the use of new data sources and technologies whilst another specifically looks at the integration of geospatial data.

There is opportunity to capitalize on the use of non-traditional data sources such as mobile phone data, satellite-based technology and social media data among others to generate insights across a range of development issues to support the SDGs. While some of the big data can be leveraged to measure specific indicators, the greater opportunity lies in its use in contributing to the achievement of specific targets and in its use to complement traditional statistics, rather than replacing them by leveraging the various advantages afforded by the various sources: quicker updates, more granular detail and possibly lower cost than traditional surveys.

However, in spite of the interest garnered on the use of big data in the global statistical space, the discussion around big data is very much at an embryonic stage in the global south. For instance, while Sri Lanka has identified big data as a potential option for generating insights on the SDGs, to date there has been little traction on specific applications. Another issue is the fact that a lot of big data generated is in the hands of the private sector. While ‘data philanthropy’ has seen entities share their data for public purposes, a longer-term use of such would necessitate the development of more sustainable data sharing models. Moreover, concerns of privacy cloud the discussions around big data, as does the lack of a governance structure surrounding the use of big data such as mobile network big data. There are serious concerns about the possible harms of using big data, in relation to privacy, surveillance, identity, and among others. Thus, there is a need to have a balanced discussion on effective ways of leveraging big data within a framework that addresses concerns of privacy, marginalization and competition among others.

BIG DATA AND SDGS

While the initial focus on big data appears to be the volume (i.e. extremely large datasets), the term now captures a range of other features proposed by numerous players (For example, Laney, 2001; Gartner, 2011; Jones, 2012; IBM, 2013). Thus, the five Vs are typically used to describe the scope of big data—Volume: scale of data generated; Variety: data could be structured or unstructured and could take various forms; Velocity –speed of data generation and analysis; Veracity – quality of data generated and Value – impact on the economy and on society. Sources of big data could then include various categories of data that include, but are not limited to administrative data such as health records and census data, transaction-generated data such as credit card payments, mobile phone data and purchases at supermarkets, online data such as social media data and web-scraped data, sensor data such as satellite data and traffic data. While some of the big data can be leveraged to measure specific indicators, the greater opportunity lies in its use in contributing to the achievement of specific targets and in its use to complement traditional statistics, not replace them.

Mobile Phone Data

The use of mobile data for development has been garnering increasing attention in recent years with numerous studies looking at the applications of insights derived from mobile phone data in addressing development issues. One of the variables used is call detail records. Call Detail Records or CDRs are generated whenever a phone call is made and captures information such as duration of the call, number of the person called and more importantly, the cell tower the caller and the recipient were connected to. If each of the parties is on the move over the duration of the call, their call will be connected to different cell towers, providing important information on mobility that have a range of development applications. Thus, mobile phone data can be used to derive variables on mobility of the users.

In addition to mobility patterns, CDR data can also reveal social networks within communities by showing patterns of connectivity between people. Moreover, consumption of mobile services has the potential to provide insight into economic behavior of populations. Lokanathan and Gunaratne (2014) talk about three behavior variables with development applications: mobility, connectivity and consumption. Such behavior variables can be applied to various policy arenas, including but not limited to, transport/urban planning, disease propagation, disaster response, as well as economic development. These can feed into specific SDG targets and indicators. Moreover, given the frequency with which phone records are generated, they have the potential to provide ‘near real-time’ insights of the population being studied. Similarly, given the granular level of information captured by operators, the data can be fed into provide statistics are a much more disaggregated level than is possible through other means.

Goal 1: No Poverty. In addition to inferring patterns of socioeconomic status of populations, variables derived from mobile phone usage can also be used to promote great financial inclusion. Studies have looked at the correlation between a population’s socio economic levels and numerous variables including mobility (Frias-

Martinez et al. 2012; Gutierrez et al. 2013; Blumenstock et al. 2015), For instance, Blumenstock et al. (2015) sought to infer wealth at an individual level, based on that individual's mobile consumption patterns. In addition to socioeconomic status, (Kumar and Muhota, 2012) suggested that the analysis of mobile phone data could contribute to the assessment of the creditworthiness of the user. Understanding such patterns would open up avenues for reaching the unbanked.

Goal 2: Zero Hunger. In relation to goal 2, recent studies have looked at the potential to use the patterns of airtime credit purchases as a proxy indicator for food expenditure. Decuyper et al. (2014) found a relatively high correlation ($r > 0.7$) between purchases of airtime credit and spending on several market-dependant food items.

Goal 3: Good Health and Wellbeing. Human mobility is a key factor in the spread of mosquito-borne diseases, and with the increasing adoption of mobile phones, mobility data derived from mobile phone is well positioned to provide insights on the movement of population. Numerous studies have sought to understand the spread of mosquito-borne diseases such as dengue (Wesolowski et al. 2015) and malaria (Tatem et al. 2014) using CDR data to infer mobility patterns. In addition, researchers have also leveraged CDR data to understand the spread of other diseases such as Ebola (Wesolowski et al. 2014), Rubella (Wesolowski et al. 2015) and Cholera (Bengtsson et al. 2011).

Goal 4: Quality Education. While researchers (Sundsøy, P. 2016) have attempted to predict the rate of illiteracy using mobile phone data, the use cases of leveraging big data to further this goal appears to be limited.

Goal 5: Achieve Gender Equality and Empower all Women and Girls. While there don't appear to be significant inroads in predicting the gender of anonymized mobile phone subscribers, Sundsøy et al. (2015) sought to predict the gender of mobile phone users based on their phone usage. The identification of gender would enable the generation of gender-disaggregated insights across the spectrum of applications of mobile data for development.

Goal 8: Decent Work and Economic Growth. Studies have sought to understand the correlation between mobility and economic activity (Kreindler and Miyauchi, 2015) with others aiming to use mobile phones to predict aggregate rates of unemployment by identifying shocks in the workforce (Toole et al. 2015)

Goal 9: Industry, Innovation and Infrastructure. Mobility variables derived from mobile data and can be used for a range of applications including urban planning and traffic management. For instance, it can help identify origin-destination flows (Calabrese et al. 2011) and provide insights on the movement of population during given times of the day (Samarajiva et al. 2015). Moreover other studies have sought to identify road usage patterns by leveraging call detail records (Toole et al. 2014).

Goal 10: Reduced Inequalities. Mobile data can be used in inferring the socioeconomic status of populations as described in goal 1.

Goal 11: Sustainable Cities and Communities. The applications of mobile phone data to goal 9 can also feed into the development of sustainable cities and communities. Moreover, this goal also covers resilience to disasters and mobility data derived from mobile phones can be used in identifying the movement of people after disasters enabling responders to direct relief efforts to areas in need. Studies have looked at mobility after earthquakes (Wilson et al 2016) and cyclones (Lu et al. 2016) among others.

Goal 13: Climate Action. Apart from understanding the movement of people after a disaster, there appear to be limited applications of mobile data to measure/monitor this goal.

Goal 16: Peace, Justice and Strong Institutions. Mobile data can contribute towards the measurement of one of the goals under this target, namely the reduction of violence and related death, by helping in crime prediction. For instance research conducted by Bogomolov et al. (2014) leverage mobile phone data in conjunction with other data such as reported criminal cases, weather data etc. to predict areas that had a higher likelihood for crime.

Satellite Data

The use of satellite data for development purposes has traditionally been used as a means of identifying changes in topography. Remote-sensing based technology solutions have applications in numerous development areas including, but not limited to urban planning, natural resource management, disaster management, crop harvesting, monitoring effects of adverse weather conditions among others. In terms of the SDGs, they can contribute towards the measurement/monitoring of numerous targets and/or indicators.

Goal 1: No Poverty. Nighttime luminosity as observed through satellite imagery has been used to estimate poverty (Elvidge et al. 2009). However, a caveat of this has been the fact that this doesn't differentiate those in various positions within the poverty spectrum. To account for this, other studies have also leveraged daytime satellite imagery, applying machine-learning techniques to features of daytime satellite images that were predictive of poverty (Jean et al. 2016).

Goal 2: Zero Hunger. Remote sensing technology is well positioned to identify changes in topography and as such can be leveraged in monitoring drought conditions (Unganai & Kogan, 1998; Rhee & Carbone, 2010), particularly useful for food insecure regions. Similarly, data derived from remote sensing data can be used in conjunction with traditional vegetation indices to predict crop/yield (Kogan et al. 2011).

Goal 6: Clean Water and Sanitation. Satellite data can be used to observe changes in the water-related ecosystem with numerous studies (Haas et al. 2009; Pekel et al. 2014) using satellite data to map surface water.

Goal 7: Access to Clean and Affordable Energy. Night-time satellite imagery, can be particularly helpful in identifying electrification (Townsend & Bruce, 2010) as well as identifying areas that don't have access to electricity (Doll & Pachauri, 2010).

Goal 8: Decent Work and Economic Growth. Nighttime luminosity also been used as a proxy for economic development (Elvidge et al. 1997; Sutton et al. 2007). Similarly, studies have looked at estimating GDP at national as well as sub-national levels.

Goal 9: Industry, Innovation and Infrastructure. Moreover, daytime satellite imagery can be used as predictors of poverty (Jean et al. 2016).

Goal 11: Sustainable Cities and Communities. Satellite imagery can be used in the identification of slums (Kohli et al. 2012) helping to identify urban poverty.

Goal 12: Responsible Consumption and Production. Insights derived from targets in some of the other goals that have been discussed, can contribute towards the achievement of Goal 12. For instance, using satellite imagery for crop yield estimation can better enable policymakers to plan ahead accounting for shortages/surpluses.

Goal 13: Climate Action. The use of satellite imagery in understanding changes in water-related ecosystems and monitoring drought has been discussed under previous goals.

Goal 14: Life Below Water. Moreover, remote sensing data can be used to track the movement of marine vessels, which can then be used to track the movement of vessels and identify areas of illegal fishing.

Goal 15: Life on Land. Satellite imagery can be used to assess changes in forest cover (Hansen et al. 2014) helping to identify areas of deforestation. Similarly studies that help identify changes in vegetation (Hutchinson et al. 2013) could possibly be leveraged to better understand desertification.

Social Media Data/ Search Engine Data/Online Data

The analysis of social media usage helps to get a sense of the pulse of the users in relation to various issues. When applied to the development context such sentiment analysis, could be leverages to understand the effects of price shocks, labor shocks as well as other patterns gleaned through the consumer's use of the internet.

Goal 2: Zero Hunger. UN Global Pulse (2014) found a correlation between retrospective statistics on food inflation and the volume of tweets regarding increases in food prices.

Goal 8: Decent Work and Economic Growth. Research such as those conducted by Cavallo and Rogobon (2016), seek to construct daily price indices based on online

prices from large multichannel retailers (with both an online and offline presence) focusing on products that were included in a typical basket for official consumer price index and had consumer expenditure weights.

Similarly, Xu et al., (2013) sought to understand the relationship between unemployment rates and search engine query data with data mining methods used thereafter to forecast unemployment trends.

Goal 16: Peace, Justice and Strong Institutions. Gerber (2014) conducted linguistic analysis coupled with topic modeling to analyse twitter data as an input to a crime prediction model, with results being enhanced by the use of twitter data.

Postal Data

Goal 8: Decent Work and Economic Growth. The flow of global postal data has been used to act as proxy indicators to estimate socioeconomic indicators that can be used to gauge national wellbeing (Hristova et al. 2016).

CHALLENGES

While there is significant potential to leverage new data sources such as big data to contribute towards the achievement/measurement of the SDGs, it is not without a couple of caveats. Firstly, big data should not be thought of as a panacea for all the issues with data generation and measurement. Rather, at the moment, the true value of big data lies in complementing traditional measures. Big data sources, when used in conjunction with traditional data sources, have the potential to improve the measurement of traditional indicators and provide new insights in a particular domain.

The use of big data in official statistics is still in its embryonic stages and while numerous statistical institutions have been exploring the use of new data sources, as evidenced by the UN big data project inventory, very few of these have transitioned from exploratory research or pilot phase to the generation of statistics. At best most of the applications remain at pilot stage. There is a need to move beyond this and scale up to generate insights on a national level.

Even within national statistical offices, there is a need for capacity building, institutional and otherwise to better equip them to use big data. However, capacity building would need to extend outside the National Statistics Offices (NSOs) given the broad scope of data that needs to be gathered.

Moreover, a large share of new data is in the hands of the private sector, and there is potential to use that data for public good. Given that there is a general consensus that NSOs may not be able to shoulder the burden of measuring SDG alone, NSOs could seek to broaden their horizons by partnering with other data providers. In addition to strategic partnerships with the private sector, the achievement of the SDGs would benefit from the engagement of multiple stakeholders spanning from academia to civil society and multilateral organizations. While 'data philanthropy' has seen the private sector sharing data with research institutions and academia for

development purposes, a viable business model needs to be created if these solutions are scaled to provide statistics at a national level.

Whilst beyond the scope of this paper, it is important to note that the discourse on big data for development would not be complete without a discussion on the concerns around the use of big data that stem from issues such as privacy and marginalization. There are also concerns around the need for better calibration, for instance in testing the accuracy of algorithms developed. Marginalization in particular would be important for the purposes of the SDGs since it will be important as the UN report says to “count the uncounted.” (A world that counts, 2014).

Additionally, there is a lack of a governance framework that could take into consideration the concerns around big data and regulate the use of big data for development.

CONCLUSION AND POLICY RECOMMENDATIONS

While there is opportunity to capitalize on the use of 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 the measurement of the SDGs, the adoption of big data for development faces numerous challenges: data access, need for multidisciplinary teams, concerns of privacy, marginalization and a lack of governance among others.

Based on the discussion above, countries may consider the following policy actions:

Developing a multi-stakeholder data ecosystem. The full potential of new sources of data may only be realized through multi-stakeholder engagement and strategic partnerships with private sector, academia, and civil society among others. For instance, think tanks and civil society organizations have been active in exploring the use of big data in development contexts and leveraging their expertise can help governments further capacity building initiatives of their statistical institutions. There is also opportunity for intermediaries to bridge the gap between private sector operators and the public sector. Policymakers can also play an active role in fostering a conducive environment for data collaborations such as the Orange data for development challenges.

Engage in the global dialogue. Moreover, policymakers can take advantage of the growing global discourse on big data for development, in particular initiatives of the UN, such as the annual conference on big data for official statistics, as well as other initiatives such as the UN big data Project Inventory, a repository of current big data projects by country.

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. Policymakers would do well to explore 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.

Acknowledgments

The authors gratefully acknowledge the research assistance provided by Shazna Zuhyle.

REFERENCES

- A World that Counts (Rep.). (2014). Retrieved from <http://www.undatarevolution.org/wp-content/uploads/2014/11/A-World-That-Counts.pdf>
- Bengtsson, L., Gaudart, J., Lu, X., Moore, S., Wetter, E., Sallah, K., ... & Piarroux, R. (2015). Using mobile phone data to predict the spatial spread of cholera. *Scientific reports*, 5.
- 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.
- Calabrese, F., Di Lorenzo, G., Liu, L., & Ratti, C. (2011). Estimating Origin-Destination flows using opportunistically collected mobile phone location data from one million users in Boston Metropolitan Area. *IEEE Pervasive Computing*, 99.
- Cavallo, A., & Rigobon, R. (2016). The Billion Prices Project: Using online prices for measurement and research. *The Journal of Economic Perspectives*, 30(2), 151-178.
- Decuyper, A., Rutherford, A., Wadhwa, A., Bauer, J. M., Krings, G., Gutierrez, T., ... & Luengo-Oroz, M. A. (2014). Estimating food consumption and poverty indices with mobile phone data. arXiv preprint arXiv:1412.2595.
- Doll, C. N., & Pachauri, S. (2010). Estimating rural populations without access to electricity in developing countries through night-time light satellite imagery. *Energy Policy*, 38(10), 5661-5670.
- 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.

Gartner. (2011). Gartner Says Solving “Big Data” Challenge Involves More Than Just Managing Volumes of Data. Press Release. Retrieved from www.gartner.com/it/page.jsp?id=1731916

Gerber, M. S. (2014). Predicting crime using Twitter and kernel density estimation. *Decision Support Systems*, 61, 115-125.

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.

Hansen, M. C., Potapov, P. V., Moore, R., Hancher, M., Turubanova, S. A., Tyukavina, A., ... & Kommareddy, A. (2013). High-resolution global maps of 21st-century forest cover change. *science*, 342(6160), 850-853.

Hansen, M., Potapov, P., Moore, R., & Hancher, M. (2013, November 14). The first detailed maps of global forest change. Retrieved from <https://research.googleblog.com/2013/11/the-first-detailed-maps-of-global.html>

Hristova, D., Rutherford, A., Anson, J., Luengo-Oroz, M., & Mascolo, C. (2016). The International Postal Network and Other Global Flows as Proxies for National Wellbeing. *PloS one*, 11(6), e0155976.

IBM (2013), The Four V's of Big Data. Retrieved from: http://www.ibmbigdatahub.com/sites/default/files/infographic_file/4-Vs-of-big-data.jpg.

Jean, N., Burke, M., Xie, M., Davis, W. M., Lobell, D. B., & Ermon, S. (2016). Combining satellite imagery and machine learning to predict poverty. *Science*, 353(6301), 790-794.

Jones, S. (2012, December 27). Why “Big Data” is the fourth factor of production. *Financial Times*. Retrieved from www.ft.com/intl/cms/s/0/5086d700-504a-11e2-9b66-00144feab49a.html.

Kogan, F., Adamenko, T., & Kulbida, M. (2011). Satellite-based crop production monitoring in Ukraine and regional food security. In *Use of Satellite and In-Situ Data to Improve Sustainability* (pp. 99-104). Springer Netherlands.

Kohli, D., Sliuzas, R., Kerle, N., & Stein, A. (2012). An ontology of slums for image-based classification. *Computers, Environment and Urban Systems*, 36(2), 154-163.

Kreindler, G. & Miyauchi, Y. (2015). Commuting and Productivity: Quantifying Urban Economic Activity using Cell Phone Data. *LIRNEasia*

Kumar, K. and Muhota, K. (2012), Can Digital Footprints Lead to Greater Financial Inclusion? (pp. 1–4). Washington DC.

Laney, D. (2001), 3D data management: Controlling data volume, velocity and variety. Gartner, 6.

Lokanathan, S., & Lucas Gunaratne, R. (2014). Behavioral insights for development from Mobile Network Big Data: enlightening policy makers on the State of the Art

Lu, X., Wrathall, D. J., Sundsøy, P. R., Nadiruzzaman, M., Wetter, E., Iqbal, A., ... & Bengtsson, L. (2016). Unveiling hidden migration and mobility patterns in climate stressed regions: A longitudinal study of six million anonymous mobile phone users in Bangladesh. *Global Environmental Change*, 38, 1-7.

Pekel, J. F., Vancutsem, C., Bastin, L., Clerici, M., Vanbogaert, E., Bartholomé, E., & Defourny, P. (2014). A near real-time water surface detection method based on HSV transformation of MODIS multi-spectral time series data. *Remote sensing of environment*, 140, 704-716.

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.

Samarajiva, R., & Lokanathan, S. (2013). Using behavioral big data for public purposes: Exploring frontier issues of an emerging policy arena.

Samarajiva, R., Lokanathan, S., Madhawa, K., Kreindler, G., & Maldeniya, D. (2015). Big Data to Improve Urban Planning. *Economic & Political Weekly*, 50(22), 43.

Sundsøy, P. (2016). Can mobile usage predict illiteracy in a developing country?. arXiv preprint arXiv:1607.01337.

Sundsøy, P. R., Bjelland, J., Iqbal, A., D, Pentland, A., Jahani, E., & De montjoye, Y. (2015, January). A cross country study of gender prediction using mobile phone metadata. Retrieved from https://www.researchgate.net/publication/270960813_A_cross_country_study_of_gender_prediction_using_mobile_phone_metadata

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.

Toole, J. L., Colak, S., Alhasoun, F., Evsukoff, A., & Gonzalez, M. C. (2014). The path most travelled: mining road usage patterns from massive call data. arXiv preprint arXiv:1403.0636.

Toole, J. L., Lin, Y. R., Muehlegger, E., Shoag, D., González, M. C., & Lazer, D. (2015). Tracking employment shocks using mobile phone data. *Journal of The Royal Society Interface*, 12(107), 20150185.

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.

UN Global Pulse. (2014). Mining Indonesian Tweets to understand food price crises. Jakarta: UN Global Pulse.

UN Global Pulse. (2015). Using mobile phone data and airtime credit purchases to estimate food security. Retrieved from www.unglobalpulse.org/sites/default/files/UNGP_ProjectSeries_Airtimecredit_Food_2015.pdf

Unganai, L. S., & Kogan, F. N. (1998). Drought monitoring and corn yield estimation in Southern Africa from AVHRR data. *Remote Sensing of Environment*, 63(3), 219-232.

Wesolowski, A., Buckee, C. O., Bengtsson, L., Wetter, E., Lu, X., & Tatem, A. J. (2014). Commentary: Containing the Ebola outbreak—the potential and challenge of mobile network data. *PLOS currents outbreaks*.

Wesolowski, A., Metcalf, C. J. E., Eagle, N., Kombich, J., Grenfell, B. T., Bjørnstad, O. N., ... & Buckee, C. O. (2015). Quantifying seasonal population fluxes driving rubella transmission dynamics using mobile phone data. *Proceedings of the National Academy of Sciences*, 112(35), 11114-11119.

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.

Wilson, R., zu Erbach-Schoenberg, E., Albert, M., Power, D., Tudge, S., Gonzalez, M., ... & Pitonakova, L. (2016). Rapid and near real-time assessments of population displacement using mobile phone data following disasters: the 2015 Nepal earthquake. *PLoS currents*, 8.

Xu, W., Li, Z., Cheng, C., & Zheng, T. (2013). Data mining for unemployment rate prediction using search engine query data. *Service Oriented Computing and Applications*, 7(1), 33-42.

