Working Paper

May 2018

Data Needs for SDGs

Navika Harshe

Navika Harshe is a consultant to the Bill and Melinda Gates Foundation through the Liverpool School of Tropical Medicine. She works in the Measurement Learning and Evaluation team. Prior to this she was Senior Manager, Research and Program Management at Operation ASHA. Navika was a Fulbright Scholar at the University of Chicago where she received her Masters in Public Policy. She also holds a Masters in Economics from the University of Hyderabad. Prior to joining Operation ASHA she worked with the federal government of India in both the Parliament and the Planning Commission. Her research interests include Health policy and its implementation, Economic development, Social and Public policy and Education policy.

Executive Summary: Reliable data and efficient data systems are vital to achieve the SDGs. They help in monitoring progress and identifying areas which need attention. The SDGs are global indicators but the data comes from local sources which means the countries will need to build and strengthen their data systems. Building data systems is expensive and has been estimated to cost a billion dollars a year. Other than the daunting costs some of the issues which arise are measuring the indicators itself some of which have never been measured. New methods of data collection such as mobile technologies, satellite imaging maybe the answer to more affordable and good quality data.

'*You cannot improve what you cannot measure*' is the cardinal rule for any good evaluation. It is only when something can be measured that its change can be monitored. The Sustainable Development Goals (SDGs)[1] have 17 main goals and 169 indicators to monitor these goals which cover different aspects of human development. The data required for this task both in terms of quantity and its quality is tremendous and hence will require a thorough understanding of what it entails.

The need for good data is recognized in Goal 17 of the SDG which states "To fully implement and monitor progress on the SDGs, decision makers need data and statistics that are accurate, timely, sufficiently disaggregated, relevant, accessible and easy to use"[2]. Catering to the data requirements for the SDGs will be critical in achieving these goals.

What do these terms actually mean for the collection of data. It means that countries need to invest in data systems which are capable of collecting data on each individual and various aspects of their life on a regular and timely basis . The data thus collected needs to be easy to use and accessible to the general public.

The SDG's indicators are a set of global indicators which were developed to achieve sustainable growth across the globe. **Even though the indicators are global the collection of this data will be done at national, sub-national and local levels which means the data systems at each of these levels need to be strengthened**.

One of the criticisms of the Millennium Development Goals, the predecessors to the SDGs was the gaps in data used to assess progress. The gaps were in the availability of data which impacted on the measurement of change for those indicators[3]. The MDGs were based on relative measure such as reducing poverty by half , but in many situations the actual number of poor people was not enumerated. This meant there was no accurate way of knowing what the decrease was.

As we move forward with new goals of development to be achieved by 2030 it is

[1] https://sustainabledevelopment.un.orgd/sdgs
[2] https://sustainabledevelopment.un.org/sdg17
[3]Stuart, E. et al. (2015). The Data Revolution: Finding the Missing Millions. Overseas Development Institute (ODI). Research Report 03, Development Progress, (https://www.odi.org/sites/odi.org.uk/files/odi-assets/publications-opinion-files/9604.pdf)

important to ensure that that the mistakes of the MDGs will not be repeated by the SDGs. The data gaps identified in the MDGs will need to be addressed through better quality data and measurement approaches . Such an enormous task will entail building stronger data systems across all levels of data collection in every country. The first step in this direction is understanding data and the ecosystem of data collection.

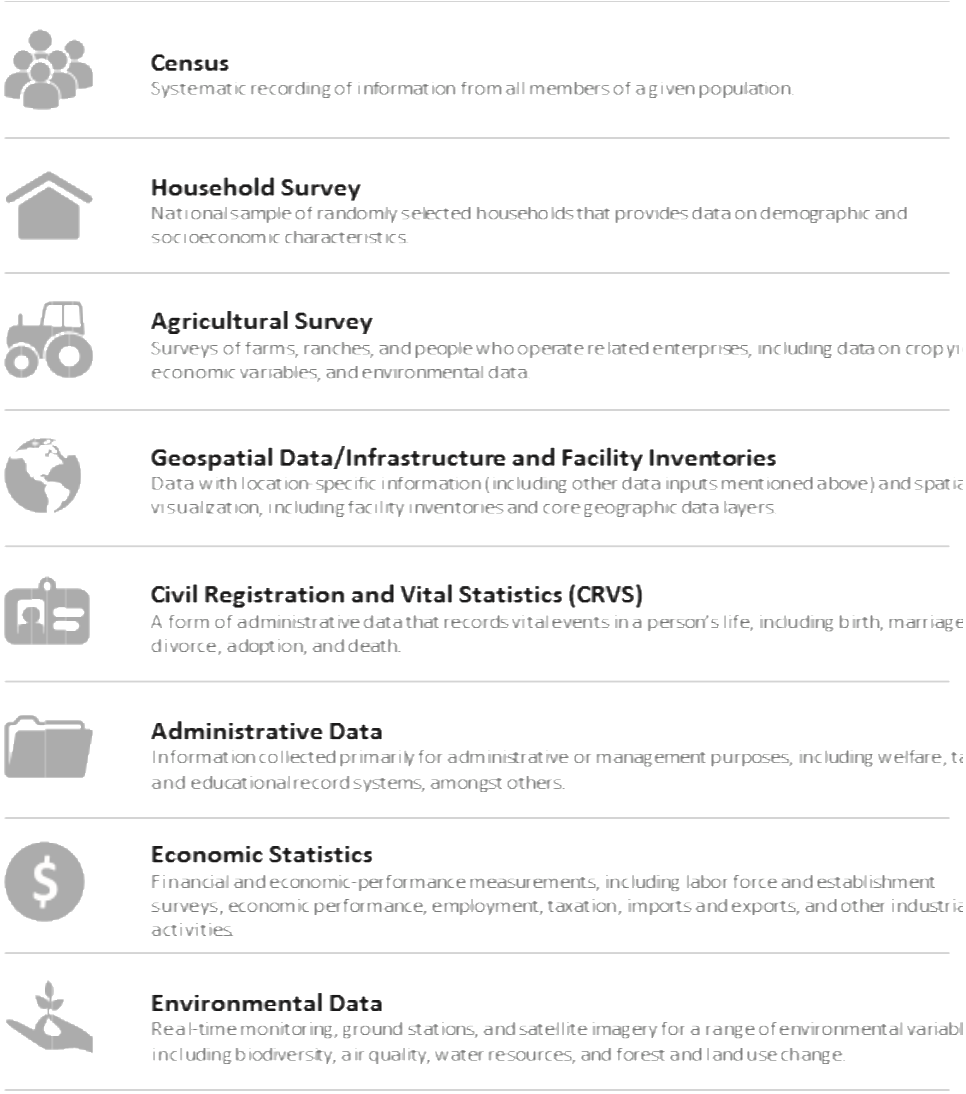
Data Ecosystem

Data does not come from one single source. There are numerous sources and methodologies of collecting it and data collected by each of these methods needs to be accurate at every level. To ensure a certain level of reliability of data the UN adopted the fundamental principles of official statistics. These are guidelines which were put in place to ensure that the data produced by national statistical systems were appropriate, reliable data and adhered to certain professional and scientific standards.[4]

When we look at countries and their current data systems what stands out is that of the 154 countries only 81 countries are implementing national statistical plans and, out of these 81 countries, only 34 comply with all the 10 fundamental principle of official statistics.[5]

Official data can be collected through different mechanism and each of these have specific methodologies to collect reliable data. The different methodologies for data collection are needed to access different types of data as not all data can be collected through one method. I present some of the main sources of data in the following table.

[4] https://unstats.un.org/unsd/dnss/gp/fundprinciples.aspx
[5] https://sustainabledevelopment.un.org/sdg17



Source: opendatawatch.com https://opendatawatch.com/strategy-guidance/data-for-development-action-plan/#textnote15

Information from all these data sources are fed in to track the SDG indicators. Each of these data systems gathers specific data and one method cannot replace the other methods of data collection. To understand the complexities around data systems we look at each of the sources in depth.

Census of a country is one of the major sources of data in which each individual is enumerated. Population and housing censuses are a primary source of disaggregated data – much needed to formulate, implement and monitor development policies and programmes. During the 10-year period from 2007 to 2016, 89 per cent of countries or areas around the world conducted at least one population and housing census, while 25 countries or areas did not have such a fundamental data source.

Census is an expensive process. For the census in a developing country, data for Development[6] estimates that the cost of data collection per individual can vary between $1 to a more realistic $3 per person. The cost gets higher in larger countries where it is more difficult to reach people. The 2000 census of the US costed $20 per person.[7]

The large scale of operations involved in the collection of census data also makes it difficult to collect information frequently. Hence, most countries do the census survey once in a decade. In the recent decade, about 89% of countries have done a population census. Yet, in the same period, 25 countries were not able to conduct a census because of financial constraints, natural disasters, social and political unrest.[8] Many of these countries were in sub-Saharan Africa, Northern Africa and Western Asia.

Another major issue which arises with using **household census** is that they measure data at the household level. This means homeless people and street children are left outside the enumeration.[9] And this population is usually the poorest, and the very people who need more help.

Administrative data is collected by the government departments and other organisations for administrative purposes.[10] Such data records daily activities and transactions. <For instance, school enrolment and drug supplies available in hospitals are all administrative data. At times, incentive for improvement are tied in with this data. For example, there might be additional funding available for the education department when enrolments increase. In this way, administrations, eager to get funds, may be tempted to show higher enrolment rates. This was one of UNESCO's main concerns while calculating the school enrolments for the MDGs.[11] Enrolment data was collated directly from schools' own administrative data, but schools tend to inflate their enrolment figures.

[6] Independent Expert Advisory Group on the Data Revolution for Sustainable Development. (2014). A World That Counts: Mobilising the Data Revolution for Sustainable Development. Available at http://www.undatarevolution.org/wp-content/uploads/2014/12/A-World-That-Counts2.pdf (https://www.odi.org/sites/odi.org.uk/files/odi-assets/publications-opinion-files/9604.pdf)
[7] Jerven, M. (2014). Benefits and Costs of the Data for Development Targets for the Post-2015 Development Agenda(Working paper). Copenhagen Consensus Centre (http://www.copenhagenconsensus.com/sites/default/files/data_assessment_-_jerven.pdf)
[8] https://unstats.un.org/sdgs/report/2017/goal-17/
[9] Stuart, E. et al. (2015). The Data Revolution: Finding the Missing Millions. Overseas Development Institute (ODI). Research Report 03, Development Progress (http://www.undatarevolution.org/wp-content/uploads/2014/11/A-World-That-Counts.pdf Accessed on 30th April 2018)
[10] https://stats.oecd.org/glossary/detail.asp?ID=6
[11] Ibid 6

Economic and budgetary data may not always be calculated using the same methodology across countries. For the calculation of economic indicators in the SDGs, countries are required to update the base year for their calculations every 10 years. Updating such a base has an effect on comparison across time.

Civil registration system for vital statistic measures all matters related to civil registrations such as birth, death, marriage. This is not well developed across countries. Between 2010-15, more than half of the countries in the world (138 of 246) had birth registration data at a 90% degree of completion.[12] Yet, there was a wide discrepancy. For example, among Sub-Saharan countries, only 8 of the 53 countries had birth registration systems. Moreover, data for death registrations systems is not functional in all countries. Only 144 countries had completed at least 75% of their death registration systems. In sub-Saharan Africa, 9 countries had death registration systems.

All these data systems are required to monitor the SDGs. These systems require a serious investment into their structure in order to ensure that they function well.

Is everything Measurable?

The indicators are created to ensure global comparison. But this effort does not acknowledge another fundamental problem: that is, **many countries do not collect data on these indicators.** Among other things, the World Bank Statistical Capacity Indicator[13] (SCI) measures the periodicity of data collection for each indicator. SCI finds that data on income poverty, which measures the number of people who live on less than $ 1.25 a day, has not been collected in Afghanistan since 2005. Afghanistan has also not had any agricultural census or population census in the last 10 years.

Countries like Brazil, China and India that have otherwise developed decent statistical systems, do not have a fully functional and proper vital registration system.[14] Health surveys in China have not been conducted since 2005. Brazil has not conducted any such survey since 2014. Data on HIV/ AIDS can at best be described as sketchy in India.

[12] https://unstats.un.org/sdgs/report/2017/goal-17/
[13] http://datatopics.worldbank.org/statisticalcapacity/SCIdashboard.aspx
[14] http://datatopics.worldbank.org/statisticalcapacity/SCIdashboard.aspx

For countries to ensure they have all the systems of data in place, including the right methodology and timeframe for collecting data regularly, it is expensive. The major costs will be monetary. However, these are not the only costs. A change in the methodology implies that those same numbers, that are used and understood internally within a country, will change. As simple as it is, when longitudinal surveys are designed, there needs to be consistency in the variables. Before any major change is made to a national data system, this aspect deserves further consideration.

Indicators for measuring gender equality also deserve deeper reflections. Goal Five aims to *Achieve gender equity and empower all women.*[15]More than 80% of indicators for SDGs 5 cannot be measured.[16] Data around measuring violence, especially intimate partner violence, violence against women with disabilities is difficult to collect[17] on account of the sensitive nature of these issues. Other problem which arise with these data are the methodology to study these issues are not well defined. The sensitive nature of these questions also requires data collectors to be qualified in handling these questions. All these factors make it difficult to collect this data regularly.

Apart from goal 5 which explicitly aims to end gender discrimination, ending discrimination is a prerequisite across the other 16 goals as well. When we look closer at each of the indicators the underlying need for gender equality becomes more apparent. For example in goal 4 which is to *Ensure inclusive and equitable quality education and promote lifelong learning opportunities for all*[18] the all includes both boys and girls. This goal cannot be achieved until all girls also have equal access to education. Similarly goals related to employment, justice and others require gender discrimination to end before they can be achieved.

Policy and legal frameworks can restrict measurability of indicators related to women. In countries such as Saudi Arabia it is illegal for women to be part of governing organisations. Till these restriction are eased out the indicators are meaningless. Alas, changing the legal framework of the country is beyond the scope of just good data.

[15] https://sustainabledevelopment.un.org/sdg5
[16] http://www.data2x.org/partnerships/gender-data-sdg-monitoring/
[17] https://sustainabledevelopment.un.org/sdg5
[18] https://www.un.org/development/desa/disabilities/envision2030-goal4.html

Monetary Cost of data

The Sustainable Development Solutions Network (SDSN),[19] in its report Data for Development, estimates that a total of US$1 billion per annum will be required to enable 77 of the world's lower-income countries to catch-up and put in place statistical systems capable of supporting and measuring the SDGs. This will require donors to maintain their current contributions to statistics, of US$300 (approx.) million per annum, but go further by leveraging US$100-200 million more in Official Development Assistance to support country efforts. These costs exclude the cost of monitoring and evaluation.

Other analyses on the SDGs, such as Data for Development,[20] use cost incurred for data needs of the MDG and, from these, extrapolate the cost for the SDGs between 2015 and 2030. MDGs, which were 18 indicators, cost $27 billion. Hence, the 169 indicators of the SDG will require about $254 billion over the 15 years. The cost, which were calculated for the MDGs, were based on the statistical cost of the survey. They did not include the cost of building data systems, hiring and training staff.

Both the calculations $15 billion for the period of 15 years in 77 low income countries and $254 billion for all countries show that monitoring and evaluation will be extremely expensive, especially considering the fact that money has always alternate uses. For many countries, it may be difficult to justify spending this amount on data and statistics instead of spending it on intervention which directly help people. It is a difficult decision to make for any country given its strong ethical conundrums.

Open data watch[21] used estimates from data for development to calculate how much some of the main data systems will cost for 77 countries eligible for International Development Aid. As stated earlier this would be close to a $1 billion every year, and it will cost between $13.5 to 14.2 billion for the entire period. Further break down of the cost for each of the data collections systems is provided in the following table.

[19] https://opendatawatch.com/wp-content/uploads/2016/09/development-data-funding-2016.pdf
[20] Jerven, M. (2014). Benefits and Costs of the Data for Development Targets for the Post-2015 Development Agenda(Working paper). Copenhagen Consensus Centre (http://www.copenhagenconsensus.com/sites/default/files/data_assessment_-_jerven.pdf)
[21] https://opendatawatch.com/strategy-guidance/data-for-development-action-plan/#textnote07

Statistical Instrument	Total cost for 77 IDA-eligible and blend countries, 2016 to 2030 (US$)	Annual cost for 77 IDA-eligible and blend countries (US$)	Source
National Survey Programs (including household, agricultural, and labor force surveys)	2.0 billion to 2.6 billion	134 million to 173 million	Costs reported on a sample of 30 countries, extrapolated to IDA-eligible/blend countries (PARIS21)
Census	4.8 billion	320 million	Per capita costs based on a sample of 26 countries, extrapolated to IDA-eligible/ blend countries (Morten Jerven)
Administrative Data			
• **Civil Registration and Vital Statistics** (CRVS)	3.3 billion	220 million (It should be noted that 80% of expenditures will take place in the first ten years)	Estimates for CoIA countries, extrapolated by population to IDA-eligible/blend countries (World Bank/WHO)
• **Education Management Information System** (EMIS)	1.4 billion	90.5 million	Estimates based on a sample of 60 countries (GPE, 2013)
Economic Statistics (excluding labor force surveys and trade statistics)			
• **Industrial Establishment Surveys**	289 million	19 million	Country-unit costs (UNIDO)
• **Improvements to Real Sector Statistics**	60 million	4 million	Country-unit costs (Morten Jerven)
Geospatial Monitoring	1.2 billion	80 million	Unit costs (CIESIN)
Environmental Monitoring (other)	514 million	34 million	Unit costs (CIESIN)
Total Costs	**13.5 to 14.2 billion**	**902 to 941 million**	

(Source:Openwatch.com https://opendatawatch.com/strategy-guidance/data-for-development-action-plan/#textnote15)

These are the estimated costs for data and statistical needs however the actual amount provided in terms of aid for statistical needs is very different. Developing countries received a total of $338 million in financial aid for statistics in 2014.[22] This was an increase of 2.9 per cent from 2010. It is estimated that to meet the data requirements of the SDGs, developing countries will need about $1 billion in statistical support annually. The deficit in the aid requirement is almost $662 million. Many countries may not have the resources to fund their data needs and may rely heavily on external funding to cover these costs. And this will require donor countries to step up and increase their funding.

All the cost estimates we have discussed here are the direct cost of strengthening the data systems, but many of the indirect cost, which feed into creating a healthy and strong data ecosystem, have bot be accounted for. The most important and crucial of

[22] https://sustainabledevelopment.un.org/sdg17

these is human resources. Creating data systems, collecting and analysing data all require people who are trained to these activities. Countries will need to invest in training people for these activities. Training will require establishing colleges, creating infrastructure for training which will include computers, software packages etc. finally the most import cost will be hiring competent people to teach others. Countries may not have competent staff who can train other people and may require assistance from other countries in building these capabilities.

Good data is expensive and requires resources which have alternate uses. As seen earlier, 25 countries failed to conduct a census survey in the last decade, but these countries were going through severe turmoil and would have required funds for equally necessary operations. When a country needs to decide how to spend its money, it is difficult to decide between a direct intervention or accurate data collection.

The final decision have to be on countries' shoulders. We cannot impose rules from outside. Yet, there is a need to have good and reliable data to measure change. How can these two needs be harmonised? One of the ways of doing this would be to emphasise the catalysing role that good quality data plays in achieving outcomes. This could be done by providing examples of successful data driven intervention within the country. In India, in the state of Maharashtra,[23] a data driven intervention, where every child was measured, was successful in reducing stunting. By measuring the height and weight of every child the health provider was able to identify children who were most vulnerable to stunting. These children were then tracked and their parents were advised on how to manage children's nutrition and additional medical help was provided when required. The systematic tracking of children allowed health providers to intervene at the right time and contain the instances of stunting.

In Uganda, health workers used SMS surveys to alert public health official on the outbreak of Malaria.[24] SMS surveys are faster and reach the concerned authorities on time. The surveys contained details of the needs and requirement of each Primary Health Facility including details about the area it serves. When there is a need for additional resources the local health provider is able to request for it through the SMS survey and

[23] http://www.mahnm.in/

[24] Independent Expert Advisory Group on the Data Revolution for Sustainable Development. (2014). A World That Counts: Mobilising the Data Revolution for Sustainable Development. Available at http://www.undatarevolution.org/wp-content/uploads/2014/12/A-World-That-Counts2.pdf

action on his request can be taken faster which is crucial for managing a situation before it gets out of hand.

Good data systems can be thought of as good blueprints for development. Unlike blueprints good data systems which are regularly updated are dynamic and keep track of changes across the system. They can help accurately identify vulnerabilities within the areas which need attention. Good data will help us target our interventions, and to ensure that those groups that need the most receive the most. Hence, it is of vital importance to advocate data needs as a critical component of development needs.

New Solutions to data collection

Methods of data collection have changed, and it would be unwise to think of data and data collection in only traditional forms. With mobile technologies, crowd sourcing and satellite data, the methods of data collection have changed, and this means that many of the traditional forms of data collections will be soon outdated. A notable example is the Economic survey of India 2016-17.[25] The survey used satellite image data to estimate the loss in property tax for the cities Jaipur and Bengaluru in India. Factors contributing to poor collection of property tax include inaccurate enumeration of building and undervaluation of the property. By using satellite images the Economics Survey helped identify the exact number of building which existed in the city and their locations. Thus identifying an alternate method for better enumeration and valuation of buildings thus plugging gaps in the standard method of estimation.

Mobile technology has also contributed to the growth of big data sources. Data from mobile phones are being used to identify migration patterns of people which helps to identify the path of a disease[26]. It is also being used to provide real time information on weather and price of agricultural commodities in different markets thus helping farmers make more informed decisions .[27]

[25] Economic survey 2016-17. (2017). New Delhi, Government of India (https://www.indiabudget.gov.in/es2016-17/echapter.pdf accessed on 2/5/2018)

[26] Milusheva Sveta (2016) "Less Bite for your Buck: Using Cell Phone Data to Target Disease Prevention" (Job Market Paper) (https://pdfs.semanticscholar.org/dab3/fc182a4699ba897d94b7b857d13bc2a398b7.pdf accessed on 2/5/2018)

[27] R. K. Lomotey and R. Deters, "Management of Mobile Data in a Crop Field," 2014 IEEE International Conference on Mobile Services, Anchorage, AK, 2014, pp. 100-107.doi: 10.1109/MobServ.2014.23

In Namibia, surveillance data, satellite imaging and mobile call logs were used to build targeted policy intervention in order to reduce Malaria.[28] These methods were cost effective and they helped in developing policies that targeted areas of high incidence, thus reducing the spread of the disease.

As more data becomes available through direct sources, it is critical to ensure the privacy of an individual. Data collected from mobile phones and other personal devices can lead to identification of the person, and in some case, it may also lead to the persecution of the person. If data is not properly protected, there is a high chance of identity theft. Individual data needs to be protected to ensure people have faith in providing their details.

In such scenarios how do we ensure that the privacy of the individual is maintained? How do we build secure data networks which ensure a person's confidentiality? These are some of the larger questions which will need to be answered as we move towards newer data collection methods. As a precaution it will serve as well to start incorporate these questions into our thinking on big data that way we will consciously try to find solutions for these problems in the design of our methodology.

Conclusions

The progress of the SDGs require good reliable data and data systems. There are different methods of collection official data, and indicators require data from all these sources. Hence, it is important to build good data systems.

Estimates of the cost of data for the SDGs are around $1 billion a year , but such estimates do not include the cost of training staff and physical infrastructure. Many countries have week data systems, where certain indicators have not been measured for a very long time and collecting data on other indicators may require a change in the legal framework of the country.

People who are the poorest of the poor, the homeless, the very people who the SDG's aim to help may get excluded hence collecting data at a granular level, which is able to capture even the most marginalised individual, is extremely important.

[28] 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), 52. doi:10.1186/1475-2875-13-52

These are serious and daunting tasks. What needs to be kept in mind is that good data is a major catalyst of change. With new methods of data collection, such as crowd sourcing, satellite imaging and mobile technologies, the cost of collecting good data will also decrease, thus leading to better measurement and ultimately achieving the goals

Reference:

1. https://sustainabledevelopment.un.orgd/sdgs
2. https://sustainabledevelopment.un.org/sdg17
3. Stuart, E. et al. (2015). The Data Revolution: Finding the Missing Millions. Overseas Development Institute (ODI). Research Report 03, Development Progress, (https://www.odi.org/sites/odi.org.uk/files/odi-assets/publications-opinion-files/9604.pdf)
4. https://unstats.un.org/unsd/dnss/gp/fundprinciples.aspx
5. https://opendatawatch.com/strategy-guidance/data-for-development-action-plan/#textnote15
6. Independent Expert Advisory Group on the Data Revolution for Sustainable Development. (2014). A World That Counts: Mobilising the Data Revolution for Sustainable Development. Available at http://www.undatarevolution. org/wp-content/uploads/2014/12/A-World-That-Counts2.pdf (https://www.odi.org/sites/odi.org.uk/files/odi-assets/publications-opinion-files/9604.pdf)
7. Jerven, M. (2014). Benefits and Costs of the Data for Development Targets for the Post-2015 Development Agenda(Working paper). Copenhagen Consensus Centre (http://www.copenhagenconsensus.com/sites/default/files/data_assessment_-_jerven.pdf)
8. Stuart, E. et al. (2015). The Data Revolution: Finding the Missing Millions. Overseas Development Institute (ODI). Research Report 03, Development Progress (http://www.undatarevolution.org/wp-content/uploads/2014/11/A-World-That-Counts.pdf Accessed on 30th April 2018)
9. https://stats.oecd.org/glossary/detail.asp?ID=6
10. http://datatopics.worldbank.org/statisticalcapacity/SCIdashboard.aspx
11. http://datatopics.worldbank.org/statisticalcapacity/SCIdashboard.aspx
12. http://datatopics.worldbank.org/statisticalcapacity/SCIdashboard.aspx
13. http://datatopics.worldbank.org/statisticalcapacity/SCIdashboard.aspx
14. https://sustainabledevelopment.un.org/sdg5
15. http://www.data2x.org/partnerships/gender-data-sdg-monitoring/
16. https://sustainabledevelopment.un.org/sdg5
17. https://opendatawatch.com/wp-content/uploads/2016/09/development-data-funding-2016.pdf
18. https://opendatawatch.com/strategy-guidance/data-for-development-action-plan/#textnote07
19. https://sustainabledevelopment.un.org/sdg17
20. http://www.mahnm.in/
21. Independent Expert Advisory Group on the Data Revolution for Sustainable Development. (2014). A World That Counts: Mobilising the Data Revolution for Sustainable Development. Available at (http://www.undatarevolution. org/wp-content/uploads/2014/12/A-World-That-Counts2.pdf accessed on 2/5/2018)
22. Economic survey 2016-17. (2017). New Delhi, Government of India(https://www.indiabudget.gov.in/es2016-17/echapter.pdf accessed2/5/2018)

23. Milusheva Sveta (2016) “Less Bite for your Buck: Using Cell Phone Data to Target Disease Prevention” (Job Market Paper) (https://pdfs.semanticscholar.org/dab3/fc182a4699ba897d94b7b857d13bc2a398b7.pdf accessed on 2/5/2018)
24. R. K. Lomotey and R. Deters, "Management of Mobile Data in a Crop Field," 2014 IEEE International Conference on Mobile Services, Anchorage, AK, 2014, pp. 100-107.doi: 10.1109/MobServ.2014.23
25. 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), 52. doi:10.1186/1475-2875-13-52