

# Report of the Planning Meeting: Forecasting Propagation of Dengue in Sri Lanka with Mobile Network Big Data

---

May 06<sup>th</sup>, 2016  
Renuka City Hotel, Colombo, Sri Lanka



## 1. Background and Objectives of the Meeting

LIRNEasia has been at the forefront of big data analysis for development in Sri Lanka, conducting in-house analysis to generate actionable insights across a range of policy domains.<sup>1</sup> The planning meeting on forecasting disease propagation in Sri Lanka with Mobile Network Big Data (MNBD) was convened by LIRNEasia and the Health Informatics Society of Sri Lanka to lay the foundation for a multi-disciplinary collaboration engaging health informatics specialists, epidemiologists, and data scientists to identify research priorities and opportunities.

This document summarizes the presentations of each session and presents the key points of discussion and next steps coming out of the meeting.

The planning meeting, held on 6<sup>th</sup> May 2016 in Colombo had the following participants:

- Madhushi Bandara, Junior Researcher, LIRNEasia
- Prof Vajira Dissanayake (Health Informatics Society of Sri Lanka, Biomedical Informatics Programme - Postgraduate Institute of Medicine)
- Dr.M.B.A. Ghouse (Epidemiology Unit)
- Thavisha Gomez, Research Manager, LIRNEasia
- Dr. Roshan Hewapathirana (Health Informatics Society of Sri Lanka, Biomedical Informatics Programme - Postgraduate Institute of Medicine)
- Dr.B.D.W.Jayamanne (National Dengue Control Unit)
- Isuru Jayasooriya, Researcher, LIRNEasia
- Sriganesh Lokanathan, Team Leader – Big Data Research, LIRNEasia
- Danaja Maldeniya, Senior Researcher, LIRNEasia
- Dr. Shehan Perera, PhD, Senior Lecturer, University of Moratuwa
- Prof. Rohan Samarajiva, Chair, LIRNEasia
- Dr. Nishan Silva (Ministry of Health)

The sessions focused on (a) an overview of current big data for development research at LIRNEasia; (b) state of the art in forecasting infectious disease propagation with big data, as well as preliminary work undertaken in Sri Lanka in collaboration with University of Moratuwa; and concluded with an (c) open discussion to identify research questions and relevant data sources.

The participants were welcomed by Mr. Sriganesh Lokanathan, Team Leader – Big Data Research at LIRNEasia and by Professor Vajira H.W. Dissanayake, President of the Health Informatics Society of Sri Lanka.

## 2. Key Points of Discussion

- **There is potential for big data to be used in conjunction with traditional statistics to address disease control and prevention measures.** Big data is not a replacement for official statistics or information sources; rather it is a value addition that can complement traditional statistics. To this end, the Epidemiology Unit expressed interest in partnering with LIRNEasia and University of Moratuwa to conduct a monthly/weekly study of mobility in Sri Lanka, (in particularly the Western province, which has the highest incidence of reported dengue cases). Moreover, the Epidemiology unit was interested in exploring the opportunity to obtain real-time mobility data centered on

---

<sup>1</sup> For more details refer to <http://lirneasia.net/projects/bd4d/>

key festivals/places of worship as part of their efforts to contain disease outbreaks, not limited to dengue, associated with such gatherings.

- **There is a strong case for leveraging big data in forecasting the propagation of non-vector borne diseases.** According to the healthcare professionals, the spread of vector borne diseases were dependent on a range of parameters making the development of big data solutions that much more difficult, an issue discussed by Danaja Maldeniya in his presentation. Dr. Dr.M.B.A.Ghouse from the Epidemiology unit pointed out that despite the inherent importance of outcome prediction of vector-borne diseases, its multidisciplinary nature made it challenging to capture all the relevant parameters. For example, with regards to the dengue virus, an individual's movement is just one parameter. Others including mosquito population and biting patterns, vector population, virus strains need to be considered. Overall, there appeared to be a view among the participants that, at present, big data analytics would be more feasible when considering the impact of non-vector borne diseases.
- **Big data analytics can help resource-strapped health departments prioritize the implementation of disease prevention and control measures.** Lack of resource personnel is a key constraint for healthcare professionals engaging in disease prevention and control and the health sector is open to innovative approaches that would help improve their resource prioritization so they can deliver targeted healthcare solutions and more effectively manage their limited resources. Big data analytics can assist these teams by offering solutions that enable targeted prevention efforts, allowing departments to allocate resources to areas they are needed the most.
- **There is a strong need for multidisciplinary coordination to strengthen the big data ecosystem in Sri Lanka.** There are many stakeholders in the big data environment and it is essential to develop and strengthen linkages between these stakeholders to build networks that facilitate the sharing of information, technical expertise and ideas. The need for increased collaboration is highlighted by the fact that data that are needed for such analyses are being generated by a range of stakeholders, including, but not limited to, telecommunications operators and government health departments. Moreover, this is an evolving field and technical expertise is a constraint. Thus, third party providers with data analytics capabilities are well positioned to play a pivotal role in bridging this information and skill gap. There is significant opportunity for big data stakeholders such as health departments to partner with analytics providers such as LIRNEasia to address disease propagation and develop viable models that would strengthen the case for government partnerships with telecommunication operators.
- **The representivity of Mobile Network Big Data was a key concern highlighted by health professionals.** The health professionals were concerned about the ability to leverage MNBD to address disease propagation in children, a large segment of the Sri Lankan population with no access to mobile phones. Moreover, concerns were raised regarding marginalized groups such as Pakistanis in refugee camps in Sri Lanka, who are more vulnerable to HIV and Malaria and who may not have mobile phones.
- **Issues at the data gathering and reporting stage (of disease incidence) could potentially constrict the effectiveness of big data in the healthcare sector.** A key frustration among healthcare professionals was the lack of accurate, real-time granular data regarding reported disease case. Despite the implementation of an online system across selected key hospitals to capture details of dengue cases, its adoption has had limited success, with the traditional paper-based reporting system being more widely used. Moreover, further challenges arose in the loss of granularity of data when it reached the national level.

### 3. Next Steps

Participants at the meeting suggested a number of potential applications of big data in the local health sector. Three key points of action were put forward:

- **Strategic partnership between the Epidemiology unit, LIRNEasia and University of Moratuwa to cross-utilize their capabilities.** LIRNEasia and Dr Perera (University of Moratuwa) accepted an invitation by the Epidemiology unit to explore opportunities to leverage the unit's historical data in conjunction with data available to LIRNEasia to map reported dengue cases against mobility data based on MNBD. LIRNEasia further offered to support the development of a viable proof of concept that would better position the Epidemiology unit to initiate conversations with key telecommunications operators.
- **Contribute to the Master of Health Information program.** LIRNEasia offered to share some of their insights on the use of big data in development with students enrolled in their Master of Health information program with the aim of instilling in students interest regarding the possibilities of big data analytics for the health sector.
- **Further discussion with the Ministry of Health/Health Informatics Society on the development of a malaria detection model.** The participants agreed to continue the discussion on potential opportunities to develop a malaria detection model.

### 4. Proceedings

#### Welcome Message: Setting of Objectives

Mr. Lokanathan highlighted the dual purpose of the meeting in his welcome remarks summing it up as (a) an effort to get a sense of the potential of MNBD to build stronger models of disease propagation and (b) to brainstorm research parameters for concrete research ideas to work on in partnership with Sri Lanka's healthcare community. Mr. Lokanathan highlighted the need for multidisciplinary effort, with the initial meeting being the basis for some concrete research ideas that LIRNEasia, the University of Moratuwa and the Department of Health could work collaboratively on.

In his opening remarks, Dr. Dissanayake spoke on the efforts of the Health Informatics Society of Sri Lanka to grapple with the challenges of dealing with big data and stated his aim to share some of their ideas with the participants. He called out Dr. Roshan Hewapathirana (Biomedical Informatics Programme - Postgraduate Institute of Medicine) who is leading the effort from the Health Informatics Society and who was able to offer a comprehensive overview of the data they are working with.

Dr. Dissanayake spoke of their current engagement with the UNICEF, which in a couple of years could potential generate one of the larger datasets about children and their development. Moreover, he touched on their involvement with the private sector in leveraging big data analytics. Dr. Dissanayaka further expressed optimism in capitalizing on the opportunities for LIRNEasia and the Health Informatics Society to work together as well and cross-utilize their expertise. Dr. Dissanayaka who would be taking over as President of the Commonwealth Medical Association (CMA) in October 2016 (a three-year term) explained that a conference for all the stakeholders in the Commonwealth is scheduled to be held in Colombo, with one of the proposals in the pipeline being the development of a Commonwealth Health Information Network.

Mr. Lokanathan built on the introduction by providing a quick overview of big data and some prior big data-related engagements undertaken by LIRNEAsia. Given the multidisciplinary nature of the work undertaken, LIRNEAsia has assembled a group of local and international experts to be part of its advisory group in helping refine their ideas. Mr. Lokanathan, in particular called out the work of Dr. Linus Bengtsson (MD, Ph.D.) who is the CEO of Swedish-based Flowminder. Flowminder has conducted various engagements in relation to MNBD (for example, understanding how to track the spread of Cholera in Haiti, understanding displacement after Nepal earthquake to better target relief efforts).

LIRNEAsia's primary source of big data has been pseudonymized call detail records from mobile operators in Sri Lanka, which it uses to provide insights across a range of domains of relevance such as transportation and urban planning.<sup>2</sup> Mapping mobile big data with other data sources offered significant opportunity for development purposes. Participants heard of projects in Sri Lanka, that were spear-headed by LIRNEAsia in which mobile communications data were used to assess population density, track changes in density, mobility patterns of the population with some of the work used by the Western Region Megapolis Planning Project.

Participants were also given a synopsis of other ongoing research including modeling infectious disease propagations, measuring impact of a transport shock, traffic analysis using CCTV footage, and land use prediction. Mr. Lokanathan also commended the efforts of Dr. Shehan Perera, PhD, Senior Lecturer, University of Moratuwa in partnering with LIRNEAsia to drive big data initiatives within student groups.

Mr. Lokanathan dwelled into the issue of dengue in Sri Lanka and in conclusion threw the participants a question in terms of whether such mobility-based forecasting models could be used for other development/health-related purposes, or as preparation for any other communicable disease where humans were vectors.

## **Session 2: State of the Art in Forecasting Disease Propagation using Big Data**

The objective of the session was to showcase some of the current studies/projects that were supported by big data technologies and gauge how such tools may be leveraged by Sri Lankan healthcare authorities to support disease control. The session also highlighted the limitations and potential problems if such a study was replicated in Sri Lanka. It featured presentations from Mr. Maldeniya from LIRNEAsia, and Dr. Shehan Perera, University of Moratuwa.

Mr. Maldeniya discussed research by Amy Wesolowski and others in Pakistan to showcase the potential of big data analytics in identifying disease propagation.<sup>3</sup> He laid out the rationale for the case study: dengue mosquito has a very small life span, has a small range of activity (spanning around 1 kilometer) and the spread of dengue beyond the natural area of a mosquito's activity being was by the movement of infectious hosts. As a result, if there is

---

<sup>2</sup> The data from the mobile operators are pseudonymized i.e. all phone numbers were replaced with a unique identifier. LIRNEAsia does not have access to any mapping from the identifiers to actual phone numbers. LIRNEAsia was provided access to this data under non-disclosure agreements that prevent LIRNEAsia from sharing the base data with others.

<sup>3</sup> Wesolowski, A., Qureshi, T., Boni, M. F., Sundsøy, P. R., Johansson, M. A., Rasheed, S. B., Engø-Monsen, K., & 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. Available at <http://www.pnas.org/content/112/38/11887.full.pdf>

detailed, high-resolution data available on the patterns of human mobility, it would be possible to derive insights as to the spread of dengue spatially.

The study sought to identify the role of human mobility in spreading dengue. The multidisciplinary research leveraged data from Telenor Pakistan (which provided aggregated results), with multiple parties engaging in the analytical work with regards to dengue including Howard T. H. Chan School of Public Health, Center of Disease Control, Oxford University and the University of Peshawar. The primary outcome of the study was to predict how dengue would spread throughout the country.

The researchers developed a model that predicted the timing of importation of dengue from endemic regions (such as Karachi) to the rest of the country. This approach coupled with a model of human and mosquito population dynamics in a localized space allowed them to predict spatial spread of dengue with human mobility across Pakistan as the primary driver.

Based on the mobility data, dengue case data, distribution of human population and the suitability of the climate for dengue mosquito, the researchers developed a map of epidemic risk and sought to predict importation of dengue to naive regions from endemic regions through travel. Mr. Maldeniya explained the methodology leveraged for the study; particularly how mobile network data was used for this analysis.

While this study demonstrates great potential for the use of high-resolution human mobility estimates to forecast dengue propagation patterns, there are a number of limitations and issues that need to be considered. There is a disconnect between the core epidemiological model and prediction of dengue propagation based on human travel. The analysis assumes that once dengue has been introduced to a naive region, the local host-vector dynamics take over disease propagation within the region unlike other models that attempt to account for the continuous effect of inter-regional mobility patterns.

Existence of multiple dengue strains and the buildup of immunity to different strains within populations due to previous epidemics are ignored primarily due to lack of data. Immunological covariates are particularly important in regions that suffer regular epidemics.

A number of critical entomological parameters related to the life cycle of the dengue mosquito were estimated using temperature-based equations derived in controlled experiments. The applicability of these values to specific real region such as Pakistan or Sri Lanka needs to be investigated. Where available, estimates for these parameters that are specific to the region of analysis are always preferable. It may be useful to evaluate if estimating region specific values for these parameters may be worthwhile.

A number of other parameters such as the maximum sustainable mosquito population (carrying capacity), mosquito biting rate, dengue case reporting rate etc. are estimated by brute force search, a technique which carries with it the risk of settling parameter values that are local optimum with the effect of the model not being generalizable.

The weather data used in the study is based on just 38 weather stations meant to cover the entire landmass of Pakistan, a region almost 14 times the size of Sri Lanka. Local mosquito populations are affected by microclimates, which are unlikely to be captured at this level of spatial resolution.

The meeting also heard from Dr. Perera who shared preliminary findings on the work conducted in Sri Lanka to date. The study conducted by students from the University of Mortuwa and supervised by Dr Perera and LIRNEasia, leveraged call record data from telecommunication operators, population data from the census department, meteorological data and past dengue cases (at the MOH levels). He touched on three prediction models that

were leveraged by the students: Metapopulation model, artificial neural network, and a Hidden Markov Model approach.

Dr. Perera in conjunction with Mr. Maldeniya also addressed the fact that this initial exercise has served to highlight a number of issues specific to the Sri Lankan context beyond the more general considerations discussed in the Pakistan study

The different datasets that will need to be used for the analysis such as the mobility estimates from CDR data, climate data from weather stations and reported dengue cases from government sources have different spatial resolutions and reference frames. While from a policy perspective the most appropriate spatial frame of reference may be the set of Ministry of Health regions, there was difficulty in acquiring details on the boundaries of these regions, which might not be well defined. In addition transforming values from all data sources from one spatial frame of reference to another inevitably would lead to some errors and a careful study of techniques that could minimize such errors should be conducted.

Models that attempt to predict the spatial spread of dengue normally require dengue incidence rates, which are based on resident human populations. Dengue incidence rate data used in the preliminary analysis was an aggregation of data reported by hospitals at the Ministry of Health regions. It was impossible to disentangle the number of dengue cases based on resident populations from those that may live elsewhere from this data. However it is quite possible that individual case reports may contain additional details about the locations frequented by a patient.

While available weather data acquired from the Meteorology department had much better spatial resolution compared to what has been used in the Pakistan study, missing data was a significant problem and required significant effort to impute.

Similar to the Pakistan case study, the analysis ignored the presence of different dengue strains and the effect of immunity due to the lack of data.

Given the prevalence of dengue in Sri Lanka, concentrated preventive efforts are the norm. There may be a material effect on dengue propagation from these actions that will be reflected in the reported dengue incidence rates. Adjusting the model parameters for this effect may make it possible to more accurately capture dengue dynamics. However detailed spatio-temporal information regarding these activities may not exist

### **Session 3: Open Discussion to Identify Research Questions and Relevant Data Sources**

The objective of the session was to facilitate group discussion to garner recommendations of potential opportunities for strategic partnership between development-focused big data analytics providers such as LIRNEasia and government ministries, institutions and other stakeholders. Prof. Rohan Samarajiva, Chair, LIRNEasia moderated the session.

Dr. Ghouse shared details on the current stratification system used by the government to identify high-risk areas in the country. The system, which relied on two key sources: 2-3 years of historical data and predictions by local health professionals, categorized all the districts in Sri Lanka MOH-wise into high risk, medium risk and low risk areas. National funding, available for dengue control activities, was based on this risk category assigned. The relevant MOH and Public Health Inspectors (PHI) were empowered to initiate predetermined dengue control activities based on the risk level assigned to each area.

Under an NDCU proposed stratification system, a composite scoring factor has been developed taking into account population density, number of construction sites, cases of children under 15 years, number of schools. The classification would change every three months, as the scoring will be revised with the NDCU focusing on MOHs to apply this at the PHI level.

The health professionals outlined a couple of challenges in terms of data gathering and compilation that may inhibit the effectiveness of big data.

Dr. Ghouse from the Epidemiology unit outlined the challenges of resource constraints, which drove them to rely on external resources such as the military and police support the implementation of dengue prevention programs. The need to better manage their limited resources, underscored the importance of the stratification system that helped prioritize resources to high-risk areas.

A further challenge raised was the impact of migrant population, including daytime migrant population and the resulting distortion in data captured. For example, the information of a worker admitted to the Kalubowila hospital who provides contact details of his hometown in Hambantota will be captured in the relevant Hambantota MOH, despite the fact that the worker has been employed in Dehiwala, thus distorting the results.

The non-availability of granular level national data was another concern raised by the health professionals. Although clinicians and house officers are encouraged to capture detailed information regarding dengue cases, the level of detail is significantly shaved off when it reaches the national level. The issue is further exacerbated by the fact that the Epidemiology unit is not able to trace the location of each dengue case beyond the resident's town.

Moreover, the Epidemiology unit, which is the central point for dengue data collection in Sri Lanka, finds that traditional paper-based reporting system to be the more favored option despite the implementation of an online system 3-4 years ago. Under a World Bank project, 50 sentinel hospitals were selected and an infection control nurse in each of the hospital is required to visit each ward, identify the dengue patients and update the online system with patient information. However, the usage of this system has been below expectations. In addition to the challenges posed by the granularity and quality of data, health departments also struggle with the availability of data. For example, the Health Informatics Society outlined their issues when compiling a data set for nutrition which cuts across various disciplines: health, financial, social and education.

Despite the perceived challenges presented at the data gathering/reporting stage, participants were optimistic of the potential opportunities for the use of big data (human mobility) to achieve government health goals. A summary of the key points discussed is included below:

Dr. Ghouse was interested in the opportunity to leverage real time mobile network data during key festivals to predict/track the movement of people for a week or so after festivals, since such seasonal events impose a huge strain on local health authorities. Mr. Lokanathan invited Dr. Ghouse to meet again to continue the dialogue regarding the possibility of using MNBD data. Mr. Maldeniya ventured that since such patterns were roughly the same over the years, the previous years' data could potentially be used as a proxy for real-time data.

In addition, Dr. Ghouse expressed interest in a monthly/weekly study of mobility in Sri Lanka, (with a particular focus on the Western province, which had the highest incidence of dengue cases.) if the mobility data could be correlated with monthly reported cases. LIRNEasia agreed to partner with the Epidemiology unit to use their historical data in conjunction with the data accessible to LIRNEasia to develop a viable proof of concept, that

would better position the Epidemiology unit to have a conversation with major telecommunications operators.

Dr. Nishan Silva from the Ministry of Health highlighted interest from the Anti-Malaria Campaign (AMC) in developing a risk projection model as part of its efforts to prevent a Malaria outbreak in the country. With the last indigenous case of malaria reported in 2012, there has been concern that should an outbreak occur, there could be a lag in diagnosis as the specific tests required by the hospitals have been on the decline.

Moreover, Dr. Hewapathirana from the Health Informatics Society raised a question with regards to the possibility of filtering MNBD. His proposition being that given the fact that Malaria is imported, if SIM cards activated at airports and/or SIM cards that were reactivated after being dormant can be tracked, it may be possible to identify the spread of Malaria if it enters the country and also the people. With filtering they would thus be able to identify hotspots.

Although LIRNEasia said that this is theoretically possible, if the purpose was to identify specific people with malaria then there would be privacy concerns. Without legal requirements in the case of emergencies, operators would be reluctant to share un-anonymized data. In addition to this, further issues that were discussed revolved around the question of inbound travellers using roaming services from their data service provider. Local operators would therefore not own this data. Thus, eliminating a potentially significant share of inbound travelers. Moreover, unless the travelers purchase their SIM cards at the airport, it would be difficult to track them.

In addition to this, Dr. Hewapathirana spoke about an initiative undertaken by the Health Informatics Society to generate what could potentially be one of the larger datasets focusing on children. The program, rolled out in 3 districts leverages data from 600 mid wives with the ultimate aim of expanding this to the whole country. In addition to demographic data, family factors, financial, educational and behavioural factors will also be included.

Moreover, the Health Informatics Society offers a Masters in Health Informatics program at the University of Colombo, which has currently produced 90 masters' level medical officers. The Health Informatics Society has introduced big data analytics into the program. LIRNEasia extended an offer to share some of their insights on big data with the students.

## 5. Wrap up and Closing Remarks

The last session also provided time to raise any questions/issues that had not been discussed earlier. Participants expressed their interest to work collaboratively with LIRNEasia to better address health issues in Sri Lanka. Prof. Samarajiva emphasized LIRNEasia's focus as seeking research that has policy applicability and expressed their willingness to initiate discussions between government departments and mobile operators.

A recap of the key outcomes of the meeting:

- Availability of LIRNEasia to share big data insights with students at the Master of Health Information program
- Partnership opportunity for the Epidemiology unit, LIRNEasia, and University of Moratuwa (UoM) to cross-utilize their capabilities
- Further discussion with regards to the malaria detection model