Open Data & Big Data: their relevance and impact on Myanmar's humanitarian/development sectors

Phandeeyar

Cing Don Nuam & Siddharth Chaudhary

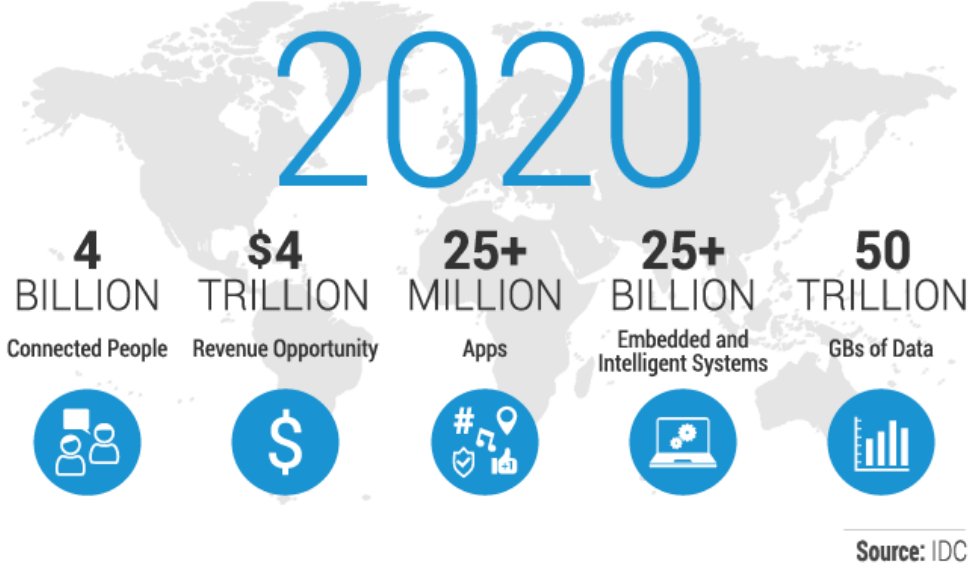
What is Open Data? What is Big Data?

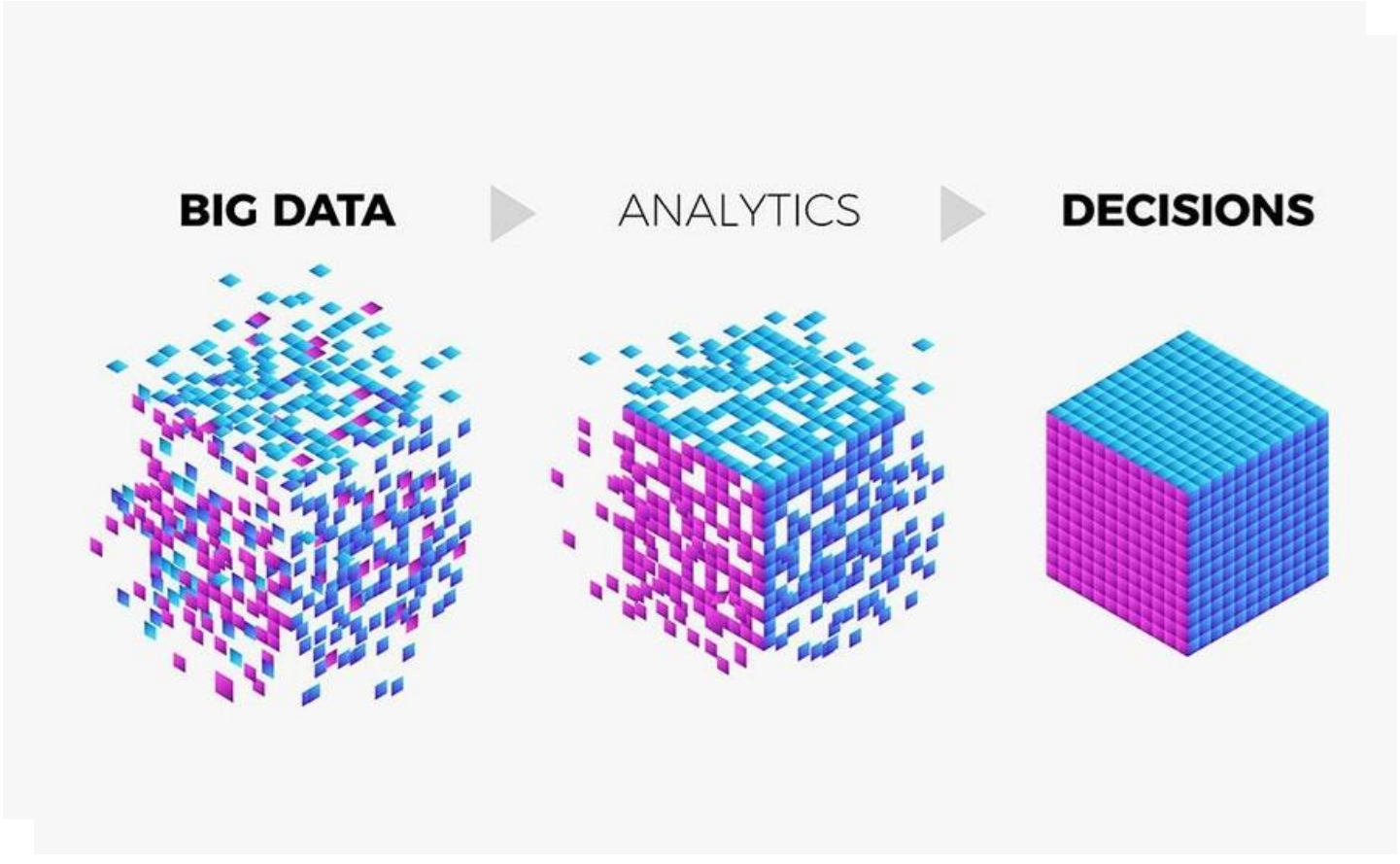
Open Data is data that can be freely used, shared and built-on by anyone, anywhere for any purpose.

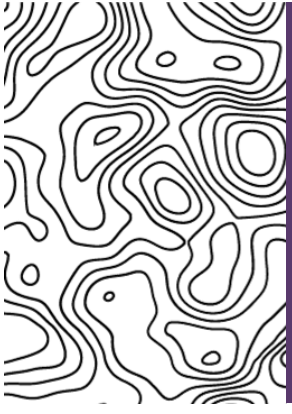
- Defining Open Data - Open Knowledge Foundation Blog

Big data is a term applied to datasets whose size or type is beyond the ability of traditional relational databases to capture, manage and process the data with low latency.


VOLUME
SCALE OF DATA
VARIETY
DIFFERENT FORMS OF DATA
BIG DATA
VELOCITY
ANALYSIS OF STREAMING DATA
VERACITY
UNCERTAINTY OF DATA


2020
4 BILLION
Connected People
$4 TRILLION
Revenue Opportunity
25+ MILLION
Apps
25+ BILLION
Embedded and Intelligent Systems
50 TRILLION
GBs of Data
Source: IDC


BIG DATA
ANALYTICS
DECISIONS



Open Data

The Role of Open Data in Myanmar's Humanitarian/ Development Sectors

- More local CSOs and NGOs are starting realize the value of data in their work.
- Thanks to a more accessible internet service, more humanitarian and development sectors (including local CSOs) are able to browse the web for more information.
- Open data is vital for customized mapping - e.g. Open Street Map, crisis mapping such as Myanmar Flood Map 2019
- Open data helps start the conversation on social issues between the public and policy makers, and collaboration between local CSOs, NGOs, INGOs and tech experts, resulting in fewer silos.

Challenges of Using Open Data by Humanitarian/ Development Sectors in Myanmar

a) Lack of either one or more of the following data literacy skills -
   - Reading data: Understanding what data is and what aspects of the world it represents.
   - Working with Data: Creating, acquiring, cleaning and managing the data.
   - Analyzing data: Filtering, sorting, aggregating, comparing and performing other such analytics operations on the data.
   - Arguing with data: Using data to support a larger narrative intended to communicate some messages to a particular audience.

Challenges of Using Open Data by Humanitarian/ Development Sectors in Myanmar

b) Lack of internet in rural and conflict areas where open data cannot be easily accessed by the local humanitarian/development organizations for their work.

c) Language barrier: While many understand English, difficult terminologies and vocabularies may strike fear in some locals working in the humanitarian and development sectors to effectively take advantage of existing open data sources.

As a Result: Data Inequalities

Open data: Empowering the empowered or effective data use for everyone? | Gurstein

Data Inequalities, if not checked, can exacerbate in the future which will put certain group in an advantageous position while others will lose out. Therefore, diminishing the gap of information inequalities is vital for an informed decision making process.

Our Approach

A crucial component of open data in humanitarian/development sector is the capacity to build a data-driven story (i.e. creating a storytelling component to communicate your data to the wider public)

- Data journalism
- Data analysis using visualization (interactive dashboards)
- Infographics
- Animation/ Videos
- Gamification (Quizzes, tests)

A Data-Driven Story Can Help

- Anchor the conversation in facts
- Generate widespread public debate
- Overcome indifference by the stakeholders
- Engage public, politicians and donors
- Grow demand for more data from the government by citizens for better governance

Data Driven Stories from Certification Program Participants

LESS THAN THE COST OF A CUP OF TEA: HOW LOW PROPERTY TAXES ARE STILL HOLDING BACK CITIES ACROSS MYANMAR

LATEST BLOGS

MYANMAR'S ECONOMIC FUTURE HINGES ON...
29-08-2018

NEW REPORT: WHERE THE RUBBER HITS THE...
02-07-2018

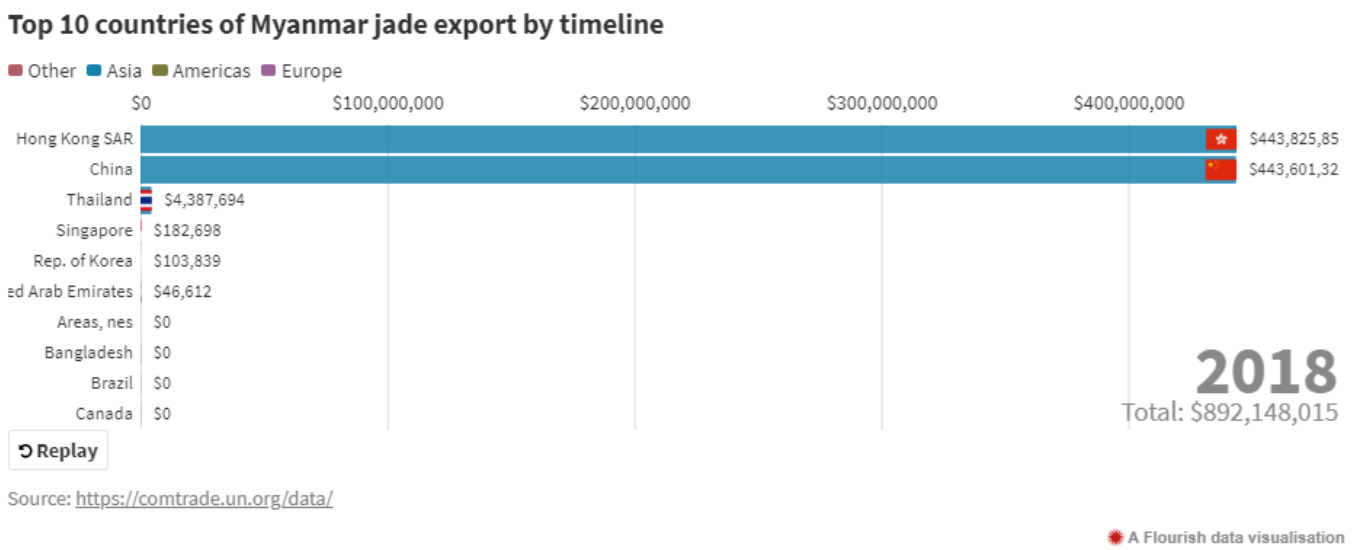
မြန်မာနိုင်ငံ၏ ဘတ်ဂျက်ပွင့်လင်းမြင်သာမှု
30-03-2018

BUDGET TRANSPARENCY IN MYANMAR
28-03-2018

SPEECH: MANAGING THE CHALLENGES OF...

Data Driven Stories from Certification Program Participants

Myanmar jade export values

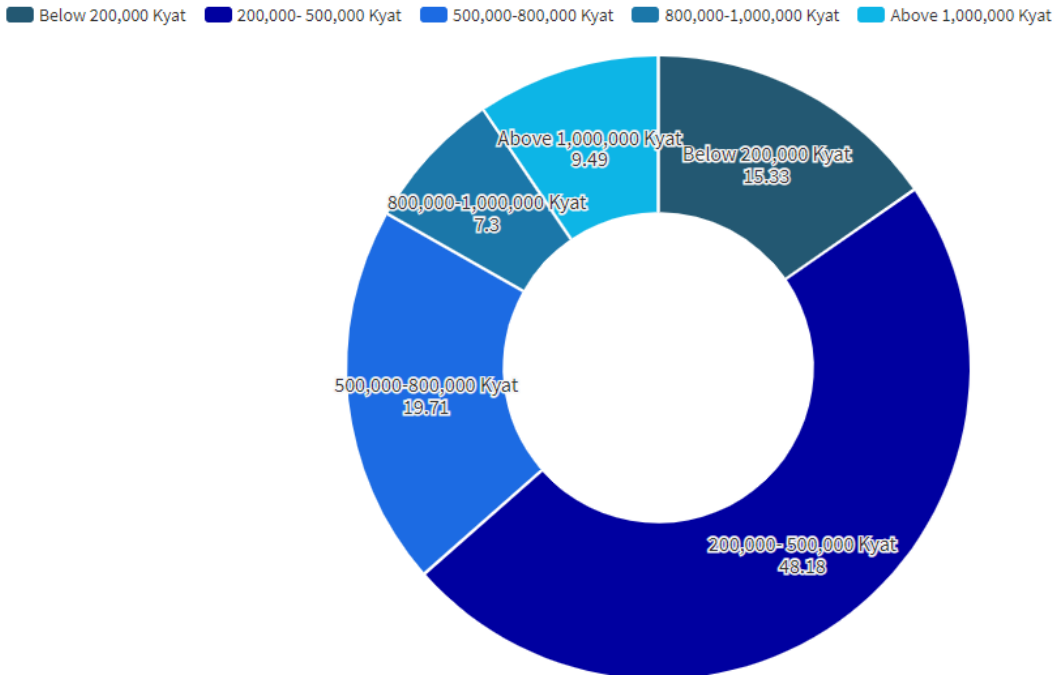


Data Driven Stories from Certification Program Participants

1 of 3

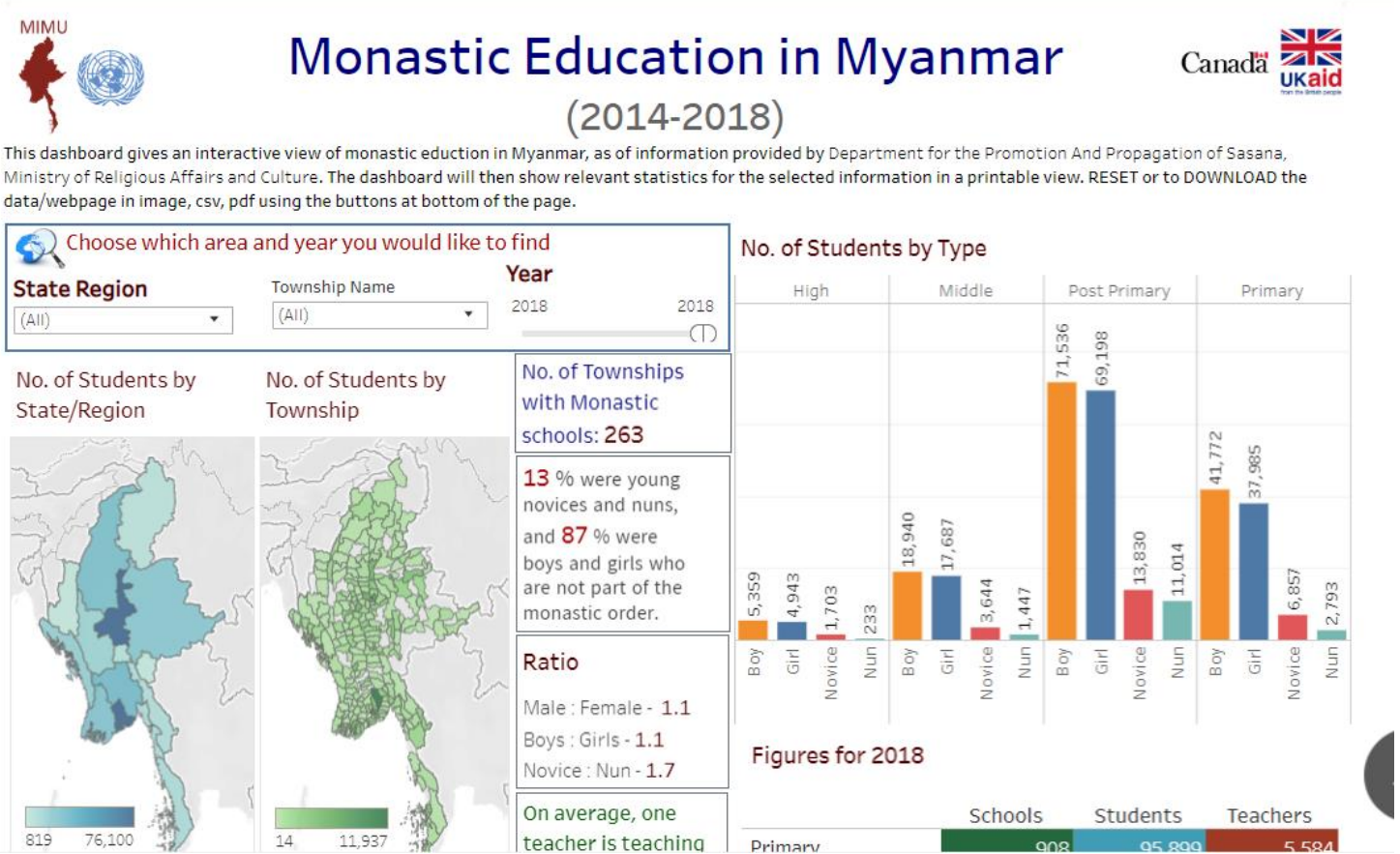
Nearly 50 percent of journalists receive the monthly salary between 250,000-500,000 Kyat

This data is based on 139 journalists

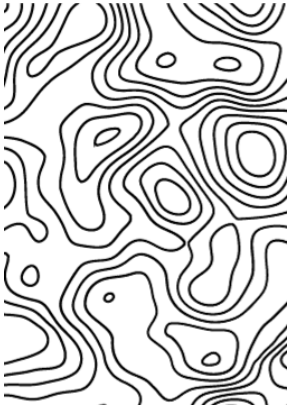


Data Driven Stories from Certification Program Participants

Data Driven Stories from Certification Program Participants



Big Data



Big Data Examples

https://app.sli.do/event/m7udqinf

K816

Big Data-driven digital humanitarianism

Big Data-driven digital humanitarianism is changing the face of humanitarian aid and development in a wide variety of fields:

1. **Migration**/refugee crisis (e.g., the Syrian migrants crisis)
2. **Epidemic crisis** (such as the dengue crisis in Punjab, Pakistan)
3. **Natural disaster** (such as the Haiti/Kashmir Earthquakes)
4. **Crowd control problems** (such as the Mina crush crisis in Hajj 2015)
5. **Terrorist attack** (such as the Peshawar, Pakistan school attack)
6. **Civil wars** (such as in Iraq and Syria)
7. **Public violence** (such as post election violence in Kenya)

Crisis Cycle

Pre-Crisis Preparedness: advance preparation of a potential crisis

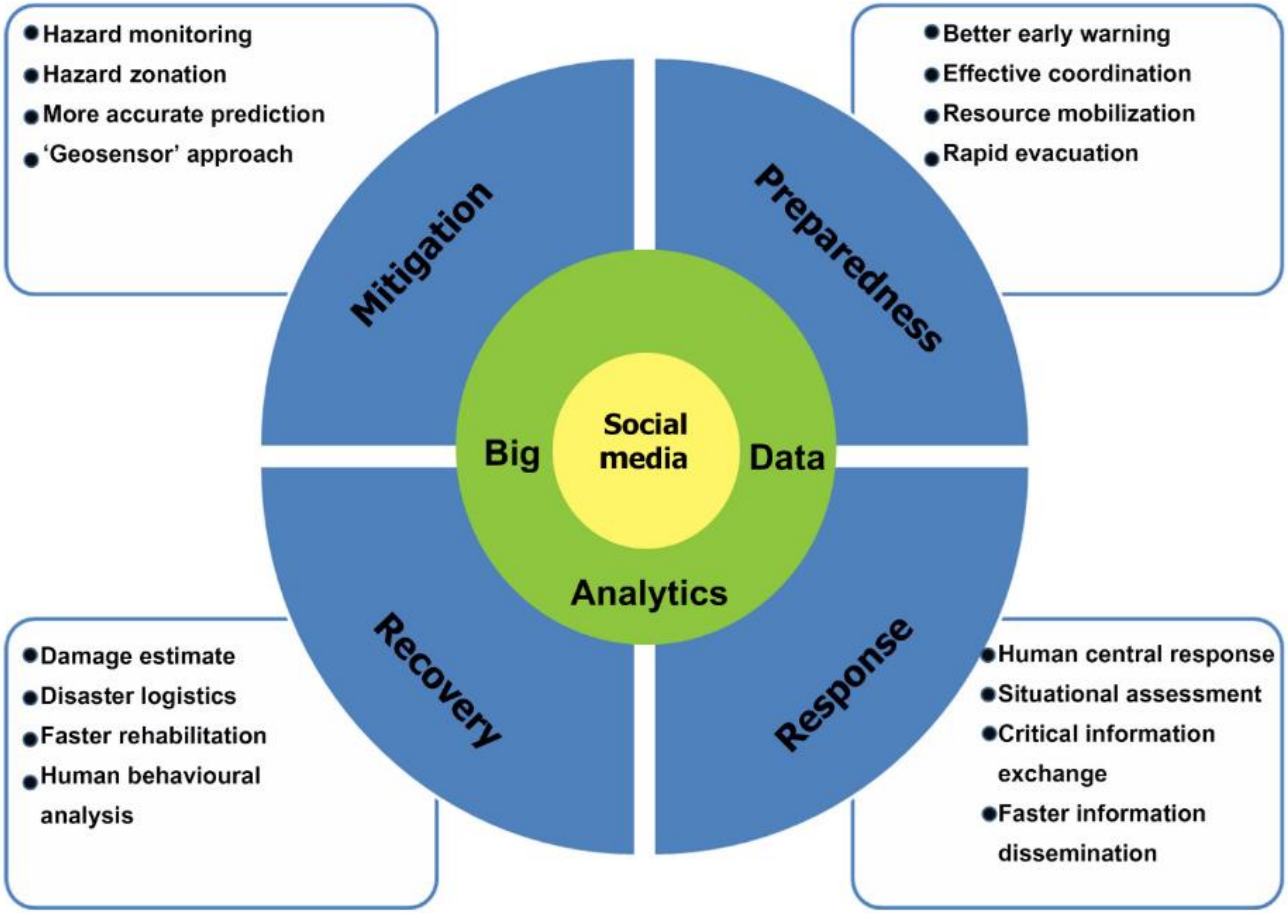
Big data analytics can be useful in this stage for emergency prediction either before the initiating event (such as an earthquake/tsunami) or as part of the unfolding of an ongoing crisis (such as prediction of the refugees' flow in the aftermath of a natural disaster such as the Haiti earthquake).

Crisis Response: coordination (among aid agencies) in the aftermath of a crisis to address any gaps or overlaps that may be affecting the effectiveness of the response.

This stage also involves adjusting and monitoring aid to ensure accountability and effective dispatch of resources so that it is available at the right time and at the right place.

Post-Crisis Response: deals with reporting, auditing, and accounting of the overall humanitarian response.

How can Big Data Help in Disaster?



Also relevant to big crisis data analytics are the many big data for development initiatives that focus on

Data mining, Data understanding, and harnessing data

Improving the world in both crisis settings (e.g., by providing **data-driven assistance**) and non-crisis settings (e.g., for **battling hunger and poverty**).

Such initiatives overlap with the digital humanitarian initiatives and include

1. **Social computing groups** (such as the UN Global Pulse, and QCRI);
2. **Groups initiated by for-profit companies** (such as Google.org—through it's Google Crisis Response, Google People finder, and Google Crisis Map projects, and Microsoft Research);
3. **Academic humanitarian initiatives** (the MIT Humanitarian Response Lab, and, Harvard Humanitarian Initiative and Peace Informatics Lab in The Hague, Netherlands);
4. **Non-profits** such as DataKind (formerly, Data without Borders).

Why big crisis data analytics is challenging?

The technical challenges associated with processing big data have traditionally been summarized using the **V**s:

1. **Volume** (large amounts of data that cannot be processed on a single machine or with traditional database tools);
2. **Variety** (data in many formats: structured, semi-structured, and unstructured—with mostly the data being unstructured);
3. **Velocity** (streaming data, with milliseconds to seconds to respond); and
4. **Veracity** (uncertainty in the data being "true" or consistent with reality). With big crisis data (such as social media), the list of "vexing Vs" is even longer: we also have to deal with
5. **Vagueness** (dealing with natural language);
6. **Virality** (understanding cascading crisis information);
7. **Volunteers** (motivating and coordinating digital volunteers);
8. **Validity** (mitigating the biases and pitfalls of social media);
9. **Values** (ensuring privacy and ethical use of crisis data); and finally,
10. **Visualization** (how to best visualize big crisis data such as crisis maps). We will cover some of these vexing Vs next; for the sake of brevity, we focus only on a subset of these challenges.

Challenges that we need to overcome

1. **Identifying** the right problems where new data sources can help
1. **Data** on its own will not yield insights
1. **Finding** data translators and data therapists
1. **Validating** big data architectures with small data foundations
1. **Reverse engineering** representativeness from big data sources
1. **The divide** is coming
1. **Finding** the right moment to introduce data innovations during emergencies
1. **Data** does not always tell you what you want to hear
1. **The impact** of new big data innovations has not yet been measured

Examples of Big Data for Humanitarian

When these sectors come together, great things can happen.

1. **Big data systems following a natural disaster was in the aftermath of the 2010 Haiti earthquake**.

1. The public sector and mobile service provider Digicel used **cellular data to track displaced populations movements both pre and post-quake to help locate** and deliver **necessary food and medical supplies** to those in need.

1. The Standby Volunteer Task Force and Humanity Road **used social media data to create a UN crisis map following the 2012 typhoon** in the Philippines.

Examples of Big Data for Humanitarian

When these sectors come together, great things can happen.

5. The map gave a detailed report on infrastructure, crop damage, housing, population movements using data derived from Twitter, **helping decision makers tailor plans to the situation**

5. **Middle Eastern refugee crisis**, seeing over **4 million displaced people** fleeing war torn countries.

   When the refugees arrive in a country to apply for asylum, their **details are entered into the UN ProGres database and from this they are issued encoded cards** and vouchers for food and other basic resources. The coding **allows spends to be tracked and monitored so that better forecasts can be made and resources made available**.

Examples of Big Data for Humanitarian

DETECTING STRUCTURES IN SATELLITE IMAGES WITH A.I. DURING HUMANITARIAN EMERGENCIES

Satellite image analysis is one of the most effective means to provide accurate data to understand the situation on the ground. The manual analysis of satellite imagery has thus far been the most reliable method of mapping structures in settlements built to house displaced populations

IMPROVING TRANSPORT PLANNING THROUGH REAL TIME DATA ANALYTICS

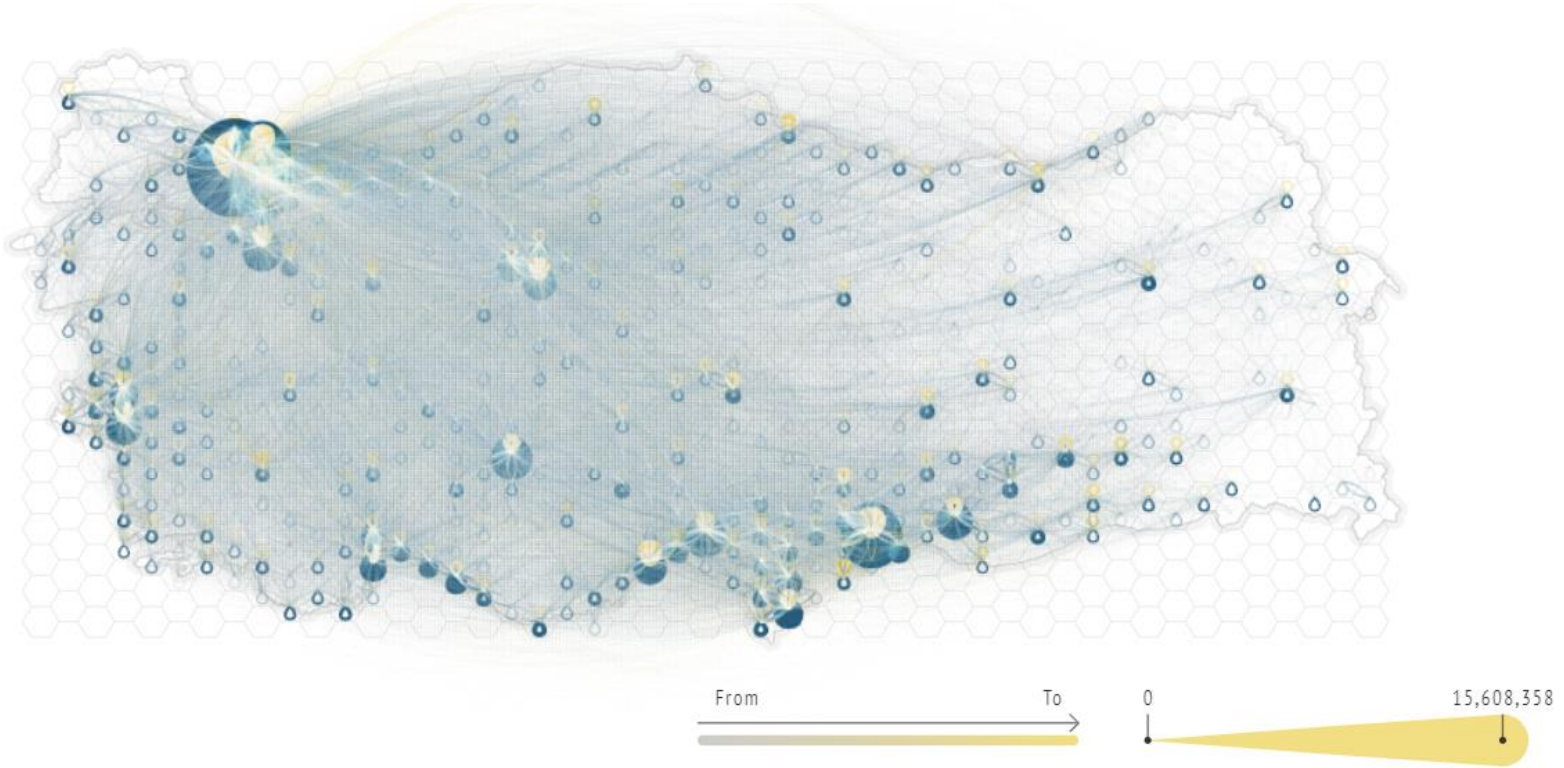
Jakarta is well-known for its traffic jams, which, among other effects, prolong commutes and complicate the scheduling of public transport. In an effort to improve the timing of busses, the Jakarta Government is looking to big data analytics for insights.

USING CALL DETAIL RECORDS TO UNDERSTAND REFUGEE INTEGRATION IN TURKEY

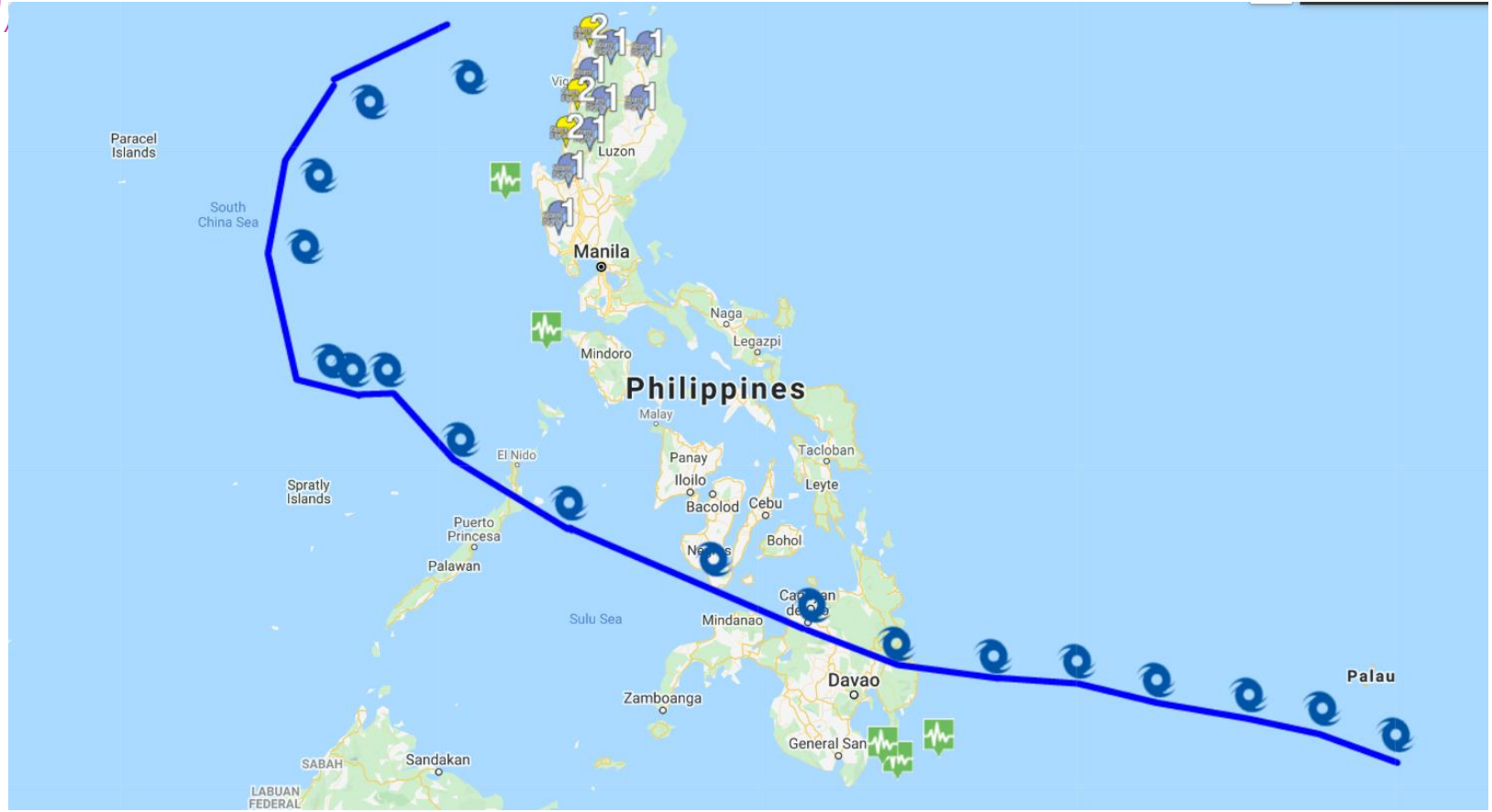
HAZE CRISIS ANALYSIS AND VISUALIZATION TOOL

Tracking the impact of Indonesia’s forest and peatland fires

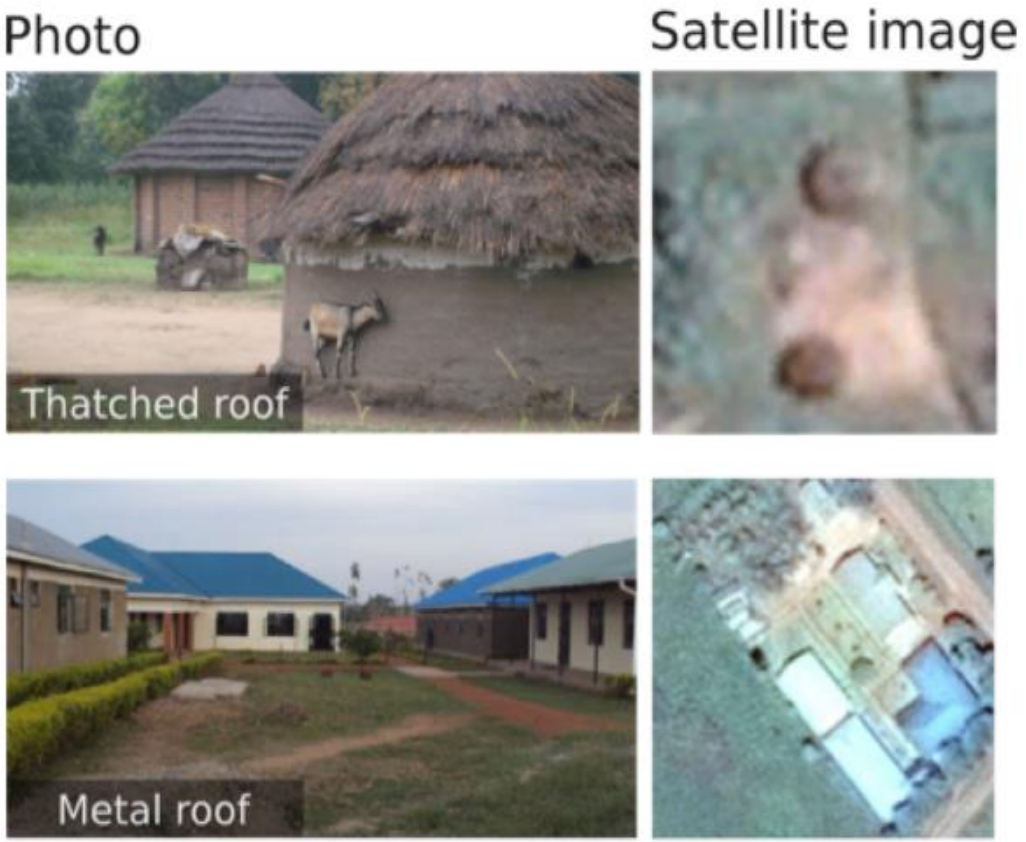
USING CALL DETAIL RECORDS TO UNDERSTAND REFUGEE INTEGRATION IN TURKEY



2012 typhoon in the Philippines



MEASURING POVERTY WITH MACHINE ROOF COUNTING



HARNESSING MOBILE BIG DATA TO IDENTIFY TUBERCULOSIS HOTSPOTS IN INDIA

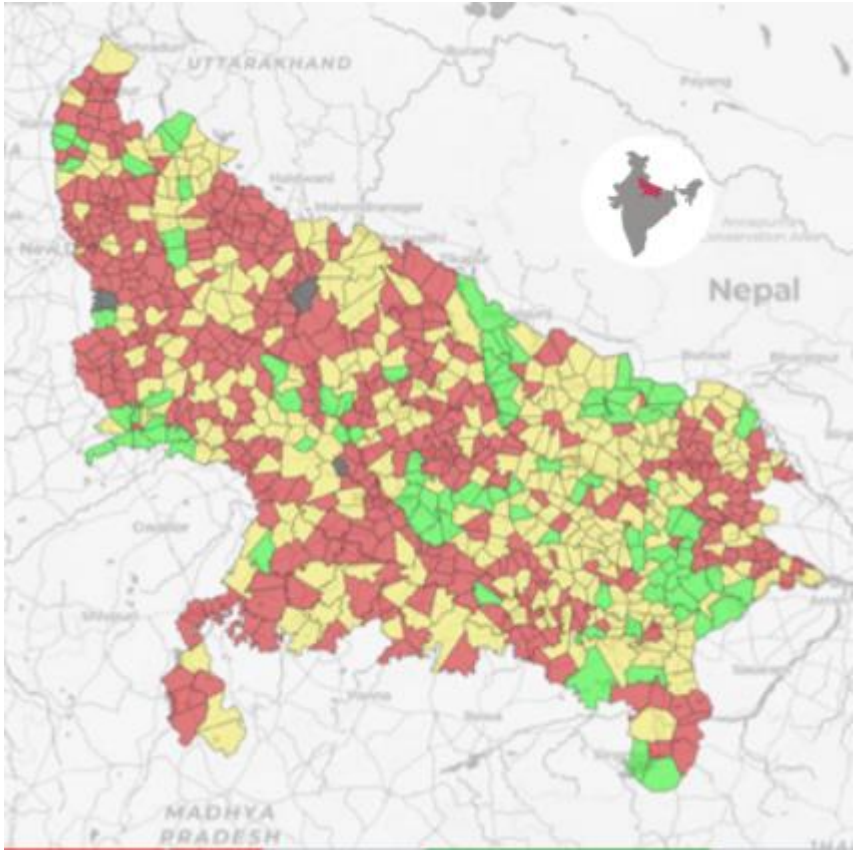
Based on the analysis of the specificities of TB, insights were established through three key stages:

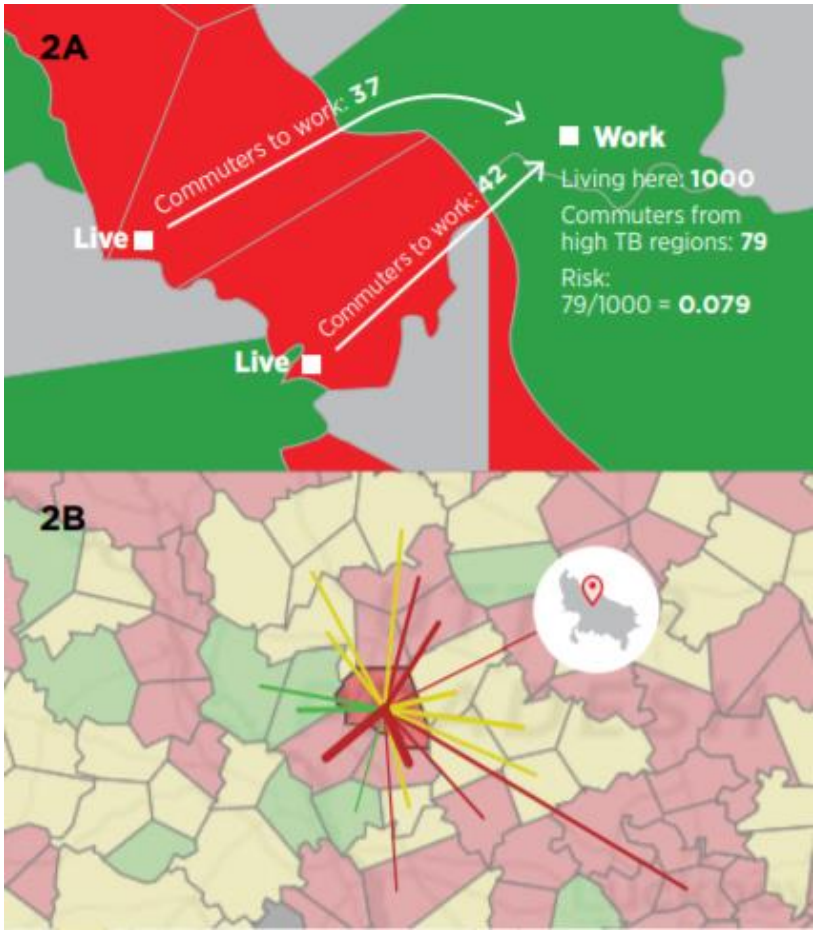
1. **Identifying** the relationship between regular movement patterns and reported TB notifications

1. **Mapping geographies** where the reported TB cases are lower than expected according to the movement patterns and TB notifications

1. **Mapping geographies** that could be at risk of increasing TB rates due to regular movement activity

HARNESSING MOBILE BIG DATA TO IDENTIFY TUBERCULOSIS HOTSPOTS IN INDIA





ESTIMATING SOCIOECONOMIC INDICATORS FROM MOBILE PHONE DATA IN VANUATU

Data from mobile phones, in particular from **Call Details Records (CDR)** and **airtime credit purchases**, can be used to understand socioeconomics conditions, especially in the absence of official statistics.

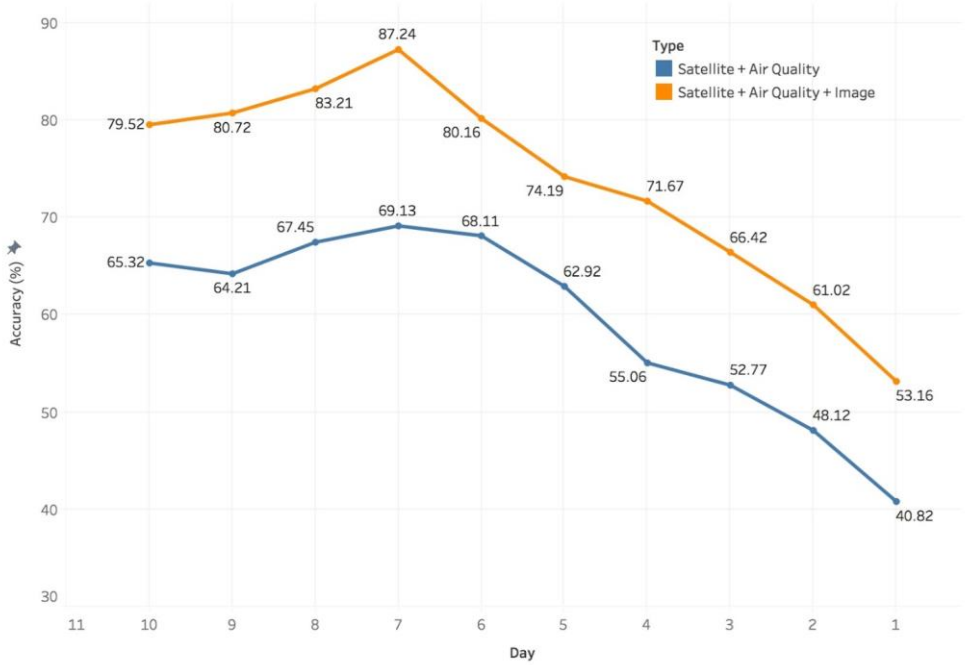
The research is using mobile phone data from an operator in Vanuatu to extract proxies for four types of statistical indicators namely:

education, household characteristics, household expenditure, and household income.

To validate the accuracy, the results are being compared to data from official statistics provided by the National Statistics Office in Vanuatu.

NOWCASTING AIR QUALITY USING SOCIAL MEDIA

1. Air Quality Index — from the Indonesian authorities' ground sensors,
2. Fire Hotspot Data — from NASA's satellites,
3. Air Temperature — from US National Oceanic and Atmospheric Administration
4. Social Media Photos — from Twitter and Instagram users.



Open Questions

Can we identify organizations working in Myanmar under each category ?

1. **Social computing groups**
2. **Groups initiated by for-profit companies**
3. **Academic humanitarian initiatives**
4. **Non-profits**

Can we identify the problems we want to solve using Big Data in Myanmar ?

1. **List of Crisis events/ Non Crisis events**
2. **Prioritize the events based on vulnerability and impact**

Can we build a IoT consortium of organization from different categories who can take a lead to work on humanitarian issues ?

http://bit.ly/2v8uicZ

Thank you
Q & A Session