



# BIG DATA SOLUTIONS IN FORCED MIGRATION

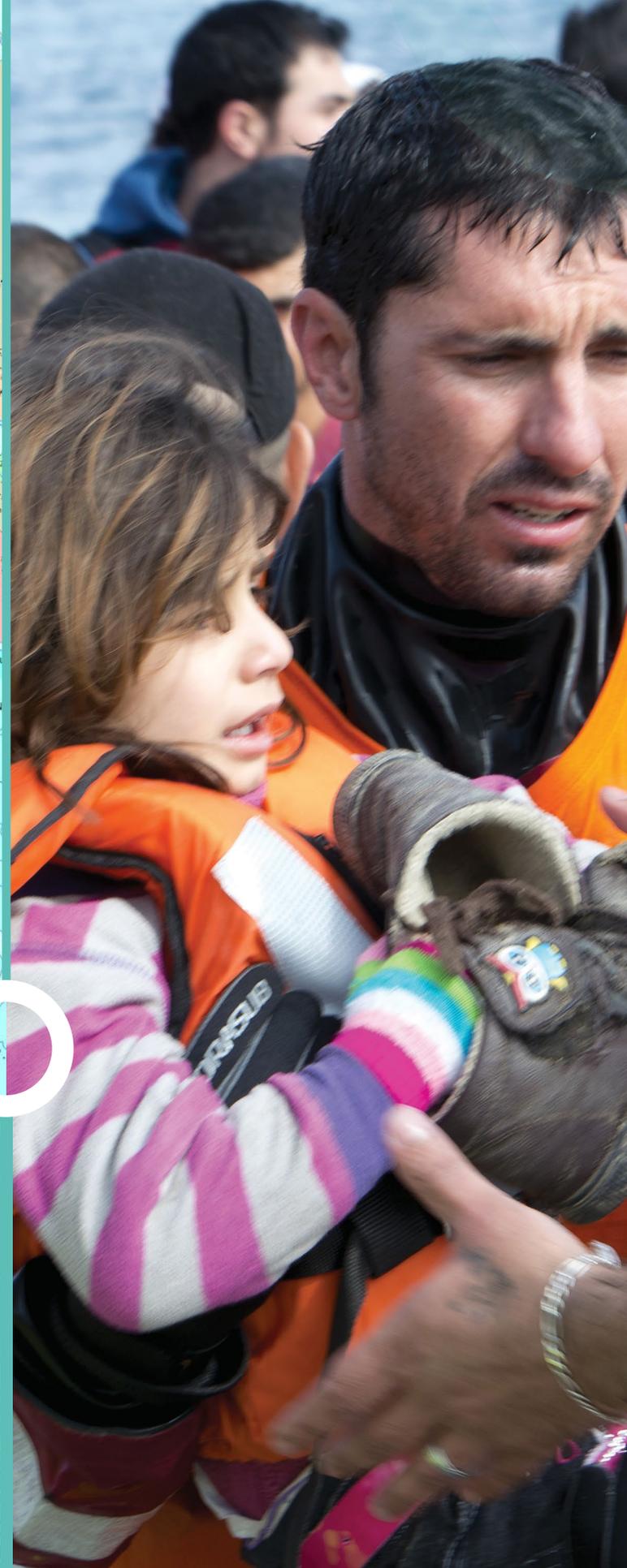
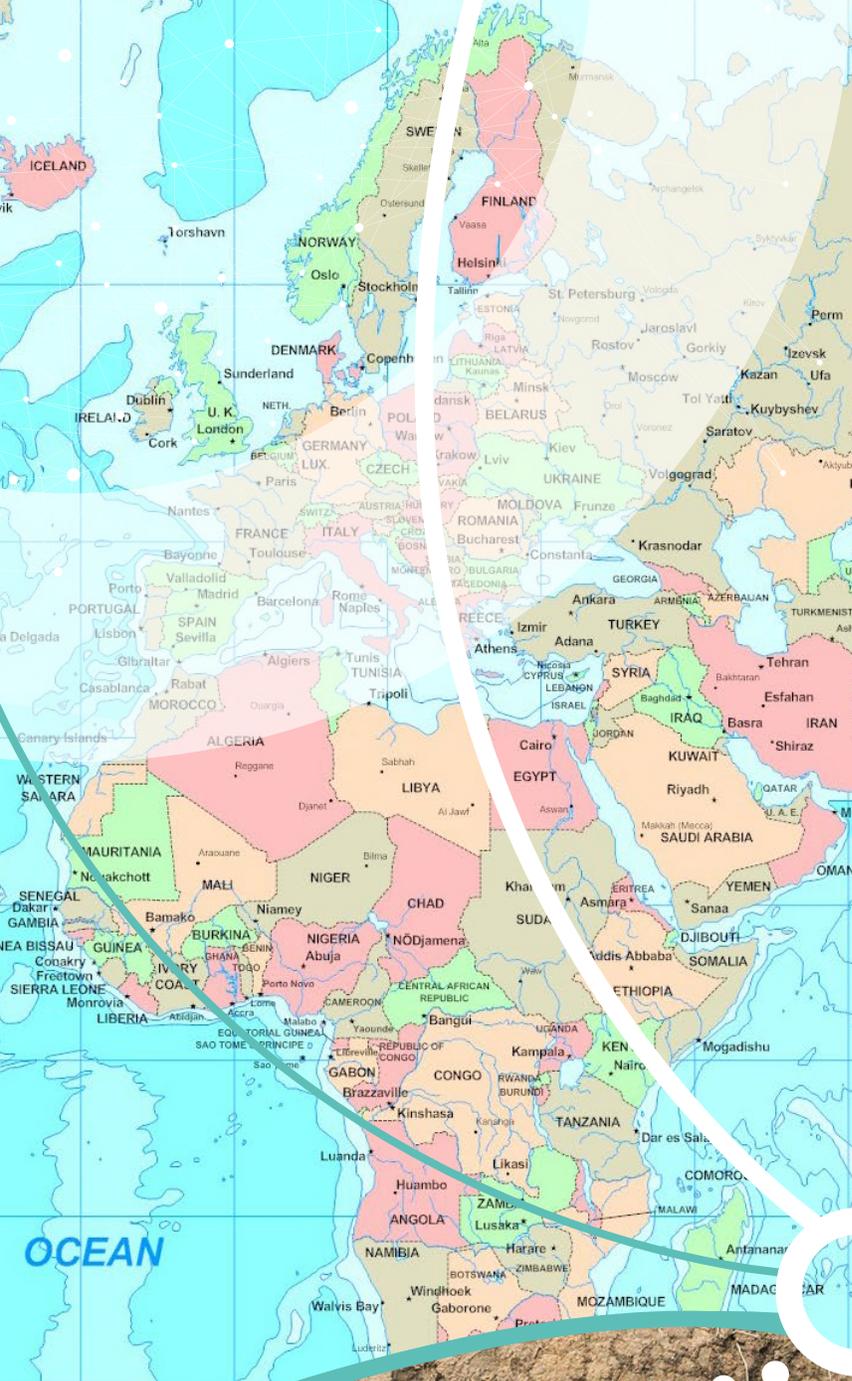
*Innovations in analytics to promote humane, sustainable responses to forced migration*



INNOVATIONS IN  
**BIG DATA**  
ANALYTICS



**WORLD BANK GROUP**  
Global Themes





# Contents

The Migration Data Gap .....	2
What is Big Data in a Development Context? .....	4
Big Data’s Potential in Migration Management.....	5
Definitions.....	8
Profiling Big Data Approaches.....	9
1. Monitoring Hate Speech – South Sudan and Myanmar.....	10
2. OSRx – Monitoring Risk Indicators for Future Conflict.....	14
3. Predicting Movements of People from Weather Data .....	15
4. Predicting Outbreaks of Mass Atrocities.....	16
5. Mapping Syria’s Conflict – the Carter Center .....	17
6. Crowdsourcing for Situational Analysis .....	18
7. The Bigger Picture – Profiling Needs through Satellite Imagery .....	19
8. USAID Famine Early Warning Systems Network .....	20
9. Migration Tracking from Population-based Models .....	21
10. Geotagging Internet Data to Follow Movements.....	25
11. The World Bank’s Research Insight Tool .....	26
12. Analyzing Social Media for Responses to Migration.....	27
Conclusion: Big Data for Migration Management of the Future .....	29
References .....	32
Acknowledgements.....	34

# Acronyms

<b>ACLED</b>	Armed Conflict Location and Event Data platform
<b>CDRs</b>	Call Detail Records
<b>FEWS NET</b>	Famine Early Warning Systems Network
<b>GDELT</b>	Global Data on Events, Location and Tone project
<b>GFDRR</b>	Global Facility for Disaster Reduction and Recovery
<b>GIS</b>	geographic information system
<b>GTSIM</b>	gravity-type spatial interaction models
<b>IDP</b>	internally displaced person(s)
<b>NASA</b>	National Aeronautics and Space Administration
<b>NGO</b>	non-governmental organization
<b>OCHA</b>	UN Office for the Coordination of Humanitarian Affairs
<b>OSRx</b>	Open Situation Room Exchange
<b>PET</b>	Potential Evapotranspiration
<b>RIT</b>	the World Bank's Research Insight Tool
<b>UAVs</b>	unmanned aerial vehicles
<b>UNHCR</b>	United Nations High Commission for Refugees
<b>USAID</b>	United States Agency for International Development
<b>YSPS</b>	Yangon School of Political Science



The numbers of people forcibly displaced from their homes each year is rising rapidly, with devastating impact at all levels, from individual lives through to the geopolitical. According to the UN High Commission for Refugees (UNHCR), there were an estimated 59.5 million forcibly displaced people in 2014, up from 37.5 million just 10 years before.<sup>1</sup> In 2015 alone, 65.3 million people were forcibly displaced by conflict, violence or human rights violations. Of these, internally displaced people (IDPs), who fled but remained in their own country, accounted for around 60 percent. Refugees, who had crossed an international border, made up 30 percent and asylum seekers the remainder. The UNHCR reports that one in every 122 people is now a refugee, internally displaced or seeking asylum.<sup>2</sup>

Forced migration describes this movement under coercion of people away from their home area. Its rapid rise poses numerous logistical and policy issues for governments, international organizations and humanitarian agencies. Challenges around needs assessment, service provision and infrastructure threaten to prevent attainment of the Sustainable Development Goals and the World Bank's twin goals of reducing poverty and sharing prosperity. Such challenges are accentuated by the lack of resources in many destinations that receive forced migration. Developing countries host 86 percent of the forcibly displaced population, with the least-developed group of countries hosting nearly 20 percent, according to the UNHCR in 2016. Middle-income countries also host

large numbers of refugees. Turkey hosts 7 million, for example, while a fifth of Lebanon's inhabitants are refugees from Syria. Increasing numbers of countries are beginning to accept that complex migration flows are now a "structural characteristic of all societies."<sup>3</sup>

Traditionally, humanitarian responses have focused on addressing the protection and immediate needs of forcibly displaced people. However, consensus is growing that these must be complemented by longer-term development interventions to help forced migrants assimilate to their new context. These interventions include those focusing on loss of assets, trauma, lack of legal rights and inadequate access to economic opportunities. Unlike economic migrants, people fleeing from

## THE MIGRATION DATA GAP

violence and persecution may find themselves in foreign places where job opportunities are limited or they are unable to use their skills. Host communities, especially in low- and middle-income countries, often face their own development challenges. The arrival of large numbers of people can dramatically alter the environment in which poverty reduction efforts are being implemented. Longer-term responses to large-scale forced migration aim to help host communities and countries continue to progress in their own poverty reduction strategies in a transformed context.

Accurate figures and information are essential for understanding and predicting refugee flows and formulating policies in response. However, an overview of data collection methodologies suggests potential for over- or underestimation, owing to variations between countries in definitions and methodologies, as well as limited data availability on returnees (especially IDPs). These inaccuracies have serious implications for people forcibly displaced, their host communities and policymakers in governments and international agencies. Accurate understanding of the scale and nature of forced migration is essential for the optimal deployment of humanitarian aid, protection of migrants, preparing host communities, managing social change, and long-term budgeting and investment.

The need to measure and understand the drivers of forced migration flows is clear. Governments, NGOs and humanitarian organizations require more information on their nature and magnitude to effectively aid those affected and address core causes. Decision-makers planning interventions or focusing resources to help forcibly displaced people need reliable, detailed information, both to determine the areas from which migrants are originating (so underlying causes can be understood) and to identify destination areas receiving large numbers of migrants

Yet specific figures for those forcibly displaced are only partially available. According to the International Institute of Sustainable Development, there could be up to a tenfold increase in the number of internally displaced people by 2050, although there is no general consensus on this estimate.<sup>4</sup> The 2012 World Disasters Report claimed that most forcibly displaced migrants are displaced long-term or are permanently dispossessed, costing the international community an estimated \$8 billion annually.<sup>5</sup>

Information on forced migration is traditionally gathered by countries and development agencies drawing on official data sources, such as population registers, visa records, residence permits, border statistics, national censuses and household surveys. However, such approaches are slow, resource intensive and may be inaccurate – for example, if migrants do not wish to report that they



are leaving, or a strong border enforcement infrastructure is lacking. National censuses can be unreliable or unavailable. Of the 49 Sub-Saharan African countries, only 12 have conducted a census in the last decade.<sup>6</sup> Uncertainty from traditional data-collection methods undermines efforts to accurately project migration flows and formulate effective policies to manage refugees and IDPs and allocate resources to where they are needed most, to ensure food security, adequate infrastructure and public health.<sup>7</sup>

Without this crucial information, efforts by governments and NGOs to provide aid that reaches those most in need are undermined, as many forcibly displaced migrants go undetected. Urban areas are particularly challenging, as IDPs may easily blend in with a city's urban poor population. One study found that IDPs in urban areas are poorer and "at a greater disadvantage and experiencing more insecurity than their non-IDP neighbors."<sup>8</sup> This information gap is not one governments and the international development community can ignore. Some estimates suggest IDPs in urban areas constitute half of all IDPs (of whom 26 million are displaced by conflict and 36 million by natural disasters).<sup>9</sup>

The 2030 Sustainable Development Goals stress the importance of improving data collection on migration. Innovative and affordable approaches are needed to bring previously inaccessible information about displaced populations to governments, NGOs and humanitarian organizations seeking to aid victims and tackle root causes. Effective relief efforts for forced migration require sophisticated and reliable data on the magnitude, demographics and behaviors of the affected population – which provides a crucial role for big data.

Emerging technologies based on big data analytics offer potential for vastly enhanced data on forced migration flows, which can provide detailed and accurate information on the needs and motives of forced migrants. Big data approaches offer policymakers and international organizations the potential to accurately measure – and even predict – forced migration flows, as well as instability and violent conflict that occur as a result. Traditional means of collecting data from refugees and IDPs remain useful, but need to be complemented by insights from new analytical capabilities.

## WHAT IS BIG DATA IN A DEVELOPMENT CONTEXT?

Big data is an umbrella term used to describe the constantly increasing flows of data emitted from connected individuals and things, as well as a new generation of approaches being used to deliver insight and value from these data flows. It is said that more data has been generated in the past two years alone than in all previous years combined. Big Data can be defined as high-volume, high-velocity and high-variety datasets that can be analyzed to identify and understand previously unknown patterns, trends and associations.

While most of the attention given to big data has focused on high-income countries, the rapid diffusion of technologies such as the internet, cell phones, ground sensors, drones and satellites – to name a few – is also driving big data innovation in low- and middle-income countries. And while data flows in the developing world are typically smaller and less diverse than in the developed world, they still present incredible opportunities for data scientists, economists and statisticians to use big data to enhance or supplement traditional analytical approaches.

Big datasets in the developing world still maintain many of the unique characteristics that make big data different from traditional datasets. They include the comparatively large volume of data, its varied and unconventional sources, and the relative speed with which it accumulates. Such characteristics enable analysts to access radically new insights and understand phenomena that traditional

data collection systems cannot offer. Unlike traditional sources of development data, such as household surveys, which address specific research questions, big data is usually produced in the course of some other activity (such as making a cell phone call). This, along with the size and complexity of some datasets, requires different research methods. Big data analytics is the emerging set of tools and methods to manage and analyze this explosive growth of digital information. It includes data science methods such as machine learning, predictive analytics and visualization. These methods offer significant opportunities to draw on real-time information to address development challenges.

The spread of analytics expertise, open-source software and low-cost analytics packages means that big data will become increasingly indispensable in helping low- and middle-income countries analyze trends and develop policy. The potential of big data in developing nations will only grow as they continue to digitize fast.

## BIG DATA'S POTENTIAL IN MIGRATION MANAGEMENT

Governments and organizations mandated with managing forced migration can draw on many sources of big data, and innovations in analytical techniques, to create demand-driven tools for understanding, predicting and managing migration flows. Advances in data collection and analysis for improving situational awareness and planning responses to forced migration are based on a variety of sources. Many high-potential applications draw on social media and the internet. Unlike most other methods – both traditional and next-generation – analysis of internet data allows for concurrent assessment of both domestic and international migration. Most approaches are unable to capture both. The use of cell phone data, for example, is likely to feature a downward bias when applied to international analyses for mapping migration, as migrants may have to switch cell phone network when they move to a new country. Geo-located use of applications such as Twitter or email are reliable indicators of movement, as users can access these anywhere. Within countries, however, analysis of Call Detail Records (CDRs) can be highly informative for research, policy formulation and implementation of responses to forced migration.

Crowdsourcing, through which information and ideas are sourced by soliciting contributions from large groups of people, can generate high-frequency data on local trends in areas where traditional data collection methods such as surveys are difficult or prohibitively

expensive. Crowdsourcing participants are usually self-selected volunteers, often part of the affected communities themselves, who are on the ground and can send relevant information to crisis responders via online platforms, SMS, Twitter or email. A variant of this is “crowdseeding,” where participants are paid for their information and ideas. This allows for a smaller but more informed crowd that can be trained in specific data collection techniques. In dynamic crises such as political violence or civil war, where people are forcibly displaced, crowdsourcing and crowdseeding approaches can quickly provide critical information to inform decision making by those mandated to respond. It can map crisis hotspots, help determine the immediate cause and magnitude of violence, and identify affected groups and their relief needs. Big data analytics also enables the application of remote sensing – such as satellite imagery and unmanned aerial vehicles (UAVs) – to gather data on migration flows and drivers such as environmental change or conflict.

These results can be displayed through continuously updated or “live” interactive data visualization dashboards, which serve as decision support tools to inform stakeholders’ responses. Armed with tools for accurate analysis of forced displacement and its effects, policymakers can begin to address key questions, such as which factors are forcing people from their homes, and how governments, development organizations and businesses can best

respond. Monitoring forced migration and understanding its drivers are essential to formulating effective policies to increase resilience to fragility, conflict and violence.

Gathering data through these new means would enable governments and international organizations to gain insights into several key policy areas:

- ***Forced migration flows and dynamics***  
Understanding migration in low- to middle-income countries through innovative data collection methods would provide development institutions with increasingly more effective tools to identify, understand and locate migrants and make evidence-based interventions. Integration of new and emerging data sources with traditional information would yield powerful insights to inform policymaking and fund allocation. By “mashing up” different sources of data – including on conflict events, environmental stressors or local governance concerns – policymakers could develop a more comprehensive view of the drivers of forced migration, and potentially design pre-emptive initiatives.
- ***The effects of violent conflict***  
In recent years, the refugee crisis in Europe has raised forced migration to the top of many countries’ agendas. Between March 2011 and May 2016, the Syrian War alone displaced almost 12 million people – a fifth of the world’s total displaced persons. Such massive flows of refugees can mean resource-strapped host countries face new social, political and economic tensions.

At worst, they can become destabilized themselves.<sup>10</sup> Big data analytics could deliver insights from far greater numbers of affected people, enabling response organizations to allocate their resources more effectively.

- ***Environmental stressors driving displacement***

Increasingly, forced migration occurs as a result of natural disasters and environmental stressors, such as drought and food shortages. In 2013, three times more people lost their homes to natural disasters than war or conflict.<sup>11</sup> The number of weather-related disasters has increased dramatically from 100 per year in the mid-1970s to 350 per year by 2011. Such disasters disproportionately affect people living in developing countries.<sup>12</sup> According to the Internal Displacement Monitoring Centre, an estimated 26.4 million people have been displaced annually by natural disasters since 2008.<sup>13</sup>

Globalization and ease of mobility mean that the effects of climate change in low- and middle-income countries have a potentially worldwide impact on security, politics and economic development. The Migration Policy Institute notes that in 2015, extreme droughts in Sub-Saharan Africa exacerbated food and income insecurity in the region, motivating migration to urban areas or internationally.<sup>14</sup> In Central America, drought is prompting increased migration regionally and to the US.<sup>15</sup> Climate change in Somalia has been

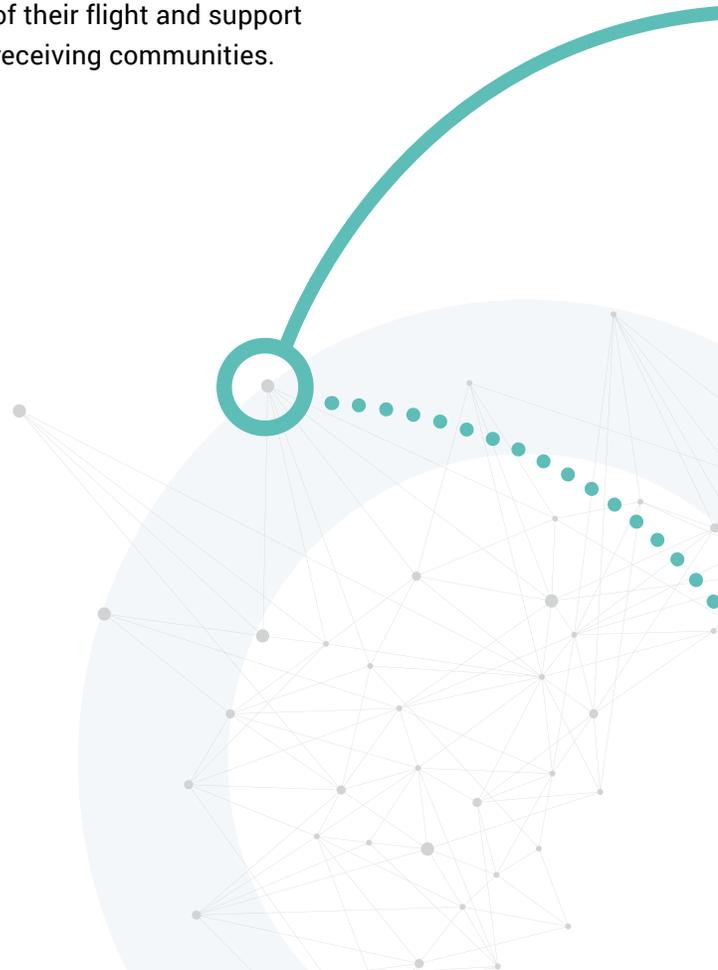
a “threat multiplier,” as consecutive years of drought and famine have exacerbated protracted armed conflict and insecurity.<sup>16</sup>

There are currently no reliable estimates for migration that results from environmental factors. The most widely cited figure is that 200 million people will migrate by 2050<sup>17</sup>. It can be complicated to identify migratory movements that occur as a result of environmental stressors alone, as such movements are often prompted by a combination of factors including conflict and natural disasters.<sup>18</sup> Traditional methods of collecting data on environmental stressors are logistically difficult and costly, and do not guarantee accuracy or a comprehensive understanding of how these stressors compel migration. Such information is paramount to supporting migrants and increasing the resilience of populations at risk.

- ***Political, social and economic dimensions of forced migration***

If the dynamics and root causes of forced migration are understood, effective responses may be more easily formulated and evaluated. When socio-economic factors force people to migrate, the migration is likely to be internal and from rural to urban areas in particular.<sup>19</sup> A study in Uganda cited the primary reasons for migration into urban centers as loss of livestock, hunger and death of a family member.<sup>20</sup> These factors were often closely linked, as many people lost animals or

relatives in violent raids and conflict, leading to food insecurity as the family lost their means of sustenance or earning an income. Urban centers offered the promise of employment and better security, especially for those who lost property or relatives to violence.<sup>21</sup> Mass migration to developing cities can exacerbate the strain on resources that such cities already experience, often forcing displaced people to settle into crowded, unregulated slums.<sup>22</sup> Cities’ inability to absorb influxes of migrants can intensify instability, resource shortages, poverty and the risk of public health disasters, such as outbreaks of infectious disease. Innovative methods of gathering information could help inform comprehensive policies which provide relief for migrants, address the root causes of their flight and support receiving communities.



# Definitions

**Forced migration** refers to the movement under coercion of a person or people away from their home or home area.

**A refugee** is a person “who is outside his or her country of nationality or habitual residence; has a well-founded fear of being persecuted because of his or her race, religion, nationality, membership of a particular social group or political opinion; and is unable or unwilling to avail him- or herself of the protection of that country, or to return there, for fear of persecution” (United Nations Convention Relating to the Status of Refugees). Asylum seekers are people in the process of having their refugee status determined.

**Internally Displaced Persons (IDPs)** are “persons or groups of persons who have been forced or obliged to flee or to leave their homes or places of habitual residence, in particular as a result of or in order to avoid the effects of armed conflict, situations of generalized violence, violations of human rights or natural or human-made disasters, and who have not crossed an internationally recognized State border” (UN Guiding Principles on Internal Displacement).

Refugees and IDPs are often referred to as **forced migrants**. By implication, other migrants – in particular economic migrants – are often called voluntary migrants. A distinguishing feature of forced migration is that the migrants may not have sufficient time and choice to determine when and how to leave and where to go. The term vulnerable migrants is also used to refer to unaccompanied child migrants, smuggled persons, victims of trafficking, and migrants who become stranded in another country.





# Profiling Big Data Approaches

**In an era of experimentation, exciting big data technologies and techniques are currently being developed for migration management, and could provide significant capability over the coming decade. Despite the potential hurdles to big data approaches in developing nations, numerous initiatives for managing forced migration have been tested in low- and middle-income countries. Big data approaches have so far been deployed in migration crises primarily for situational analysis, but as digital progress continues, many more such approaches will be harnessed for migration management in lower-income countries, at a faster rate than previously.**

To help development practitioners within and beyond the World Bank take advantage of these trends, this brief profiles key areas of experimentation to highlight the potential for big data analytics to improve responses to forced migration. These emergent approaches showcase valuable new ways of understanding unfolding events and trends that are difficult to measure, illustrating the potential of analytics in understanding and managing forced migration. The brief's conclusion outlines key challenges, lessons and potential around big data analytics in migration management.

All the big data approaches profiled are applicable in a wide variety of countries and contexts. They could achieve improvements as diverse as real-time tracking of mass migration, prediction of migration scale and patterns, accurate assessment of migrants' needs, and optimum allocation of resources to support forcibly displaced people and their host communities. Together these approaches highlight new possibilities for innovation in the quest for forced migration management that protects and enhances the lives of everyone affected.

## 1. MONITORING HATE SPEECH – SOUTH SUDAN AND MYANMAR

Big data can play an effective role in helping policymakers' understanding of conflict – a primary driver of forced displacement – and the core issues behind it. For example, hate speech is often a leading indicator and driver of instability, which in turn forces people to migrate. Profiling the transmission and dispersion of inflammatory terms can provide a better understanding of the dynamics unfolding, and can suggest interventions more likely to address underlying tensions driving the hate speech. The use of big data in this way can help in up-front program planning and in longer-term evaluation of impact.

In December 2013, two years after South Sudan achieved independence, civil war broke out in the country. By May 2017, tens of thousands of people had been killed and 3.5 million displaced from their homes, with one million more at risk of widespread famine.<sup>1</sup> Organizations responding to and trying to prevent the violence have widely recognized that hate speech and fake news propagated through social media have played a major role in inciting conflict on the ground. Facebook pages have become sources of and breeding grounds for hate speech and fake news. In particular, members of the country's diaspora have used social media channels to distribute fake news and propagate hate speech.

While these issues have affected countries worldwide, in South Sudan the stakes are

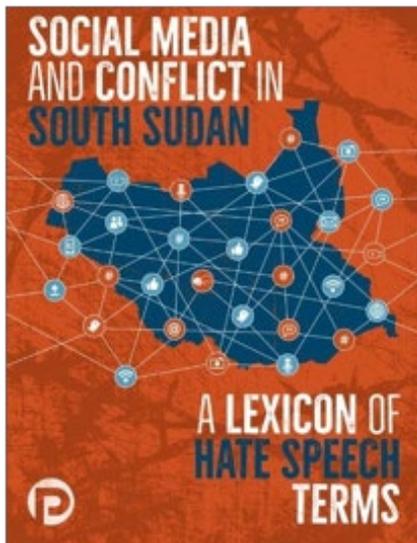
especially high. Hate speech and fake news can lead to mass atrocities. In response, two peace messaging initiatives – #DefyHateNow and #AnaTaban – have emerged, seeking to engage with South Sudanese people in the country and the diaspora, calling for peace and rejecting those who seek to inflame existing divisions. Despite such efforts and recognition of the issue, organizations engaged in South Sudan had little, if any, information on what the hate speech surrounding the conflict looked like, where it occurred, who drove it and how to stop it. This information gap prevented meaningful efforts to monitor and counter the trend.

To address the gap, PeaceTech Lab – a non-governmental organization at the intersection of peacebuilding and technology, based at the United States Institute of Peace – developed an initiative to better understand the links between hate speech, fake news and conflict dynamics in the country. The Lab developed a methodology that could combine knowledge and insights from South Sudanese people themselves with big data analytics to provide useful and actionable information for groups seeking to counter this trend and prevent future atrocities.

The project began with a survey of South Sudanese people to identify hate speech terms and corresponding attributes – the term or phrase, translations or similar uses,

<sup>1</sup> <http://www.aljazeera.com/news/2017/05/2m-children-displaced-south-sudan-conflict-170508033845325.html>

the group or groups usually targeted by it, the reason it is inflammatory, the platform or platforms it is used in, and any examples or links they could share. This information formed the basis for a “lexicon report” which explained the nuances of hate speech usage in the South Sudan conflict. Using social media monitoring and analytics, the Lab was



then able to identify hate speech content relevant to the conflict, and to share trends and examples with key stakeholders.

The social media monitoring and analytics quickly became central to the project. For example, one term identified by respondents was “MTN”, the name of the mobile phone company with the largest network in South Sudan. The term has been used as a derogatory name for Dinka people, with MTN’s tagline – “everywhere you go” – used to signify that the Dinka are everywhere in South Sudan. Given the massive amount of content that related only to the mobile network, an important task in identifying hate speech was to filter out irrelevant content. Using a social media analytics platform called Crimson Hexagon, which enables users to create “monitors” using keyword queries, the Lab created a monitor for this project using the lexicon as a reference, and trained an algorithm to recognize patterns of text

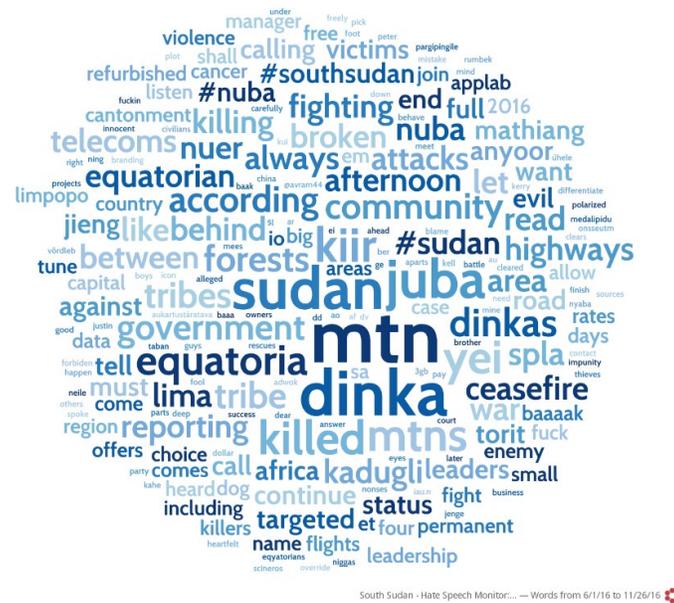
## 5. MTN

**Definition:** According to some respondents, Equatorians use this relatively new term to describe Dinkas; others indicate that it’s used widely to create fear about Dinkas’ encroachment on other communities’ traditional lands and annihilation of those communities. It’s based on the slogan for the MTN mobile service provider: MTN is “everywhere you go.” According to one person, it’s “used to target Dinkas who are found all over the country, like MTN service. It targets Dinkas who have abandoned their lands and scattered all over other lands—and especially against Dinkas when they’re traveling. Vehicles are stopped, and drivers are asked whether MTN are in the cars.” This has reportedly happened to public transport on the Juba-Yei road. In the more recent conflict in 2016, the term has evolved to mean the coordination of operations against the Dinkas.

**Why it’s offensive/inflammatory:** It stirs fear by exaggerating the number and location of Dinkas within South Sudan, suggesting an increasing presence and pervasive (negative) influence throughout the country, specifically in competition for land, access to water, government services, and jobs. It’s a coded, action-oriented word: An *MTN* with “no service available” may mean a Dinka who’s unarmed and therefore may be attacked.

designated as hate speech. This enabled the Lab to identify and analyze trends surrounding these hate speech terms. The Lab produced biweekly reports that provided context of recent events, visualizations from social media monitoring, and sample posts. These reports were shared with a range of organizations working in and on South Sudan, providing better understanding how online hate speech connected with violence on the ground.

By providing early warnings of potential threats, as well as information from social media to contextualize those threats, this approach shows the potential of big data analytics for atrocity prevention, which in turn helps prevent forced migration. In a context such as South Sudan, where organizations have extremely limited access to information, big data can provide information critical to understanding relevant trends in near real-time and to predict where violence is likely to occur.



 **James Bitoro**  
November 20 · WordPress · 

EQUATORIA IS ASKING [DINKA](#) TO GO BACK TO BAHR AL GHAZAL. The people of Great Equatoria are asking all Dinkas "MTN" to go back to their region and, only return when there is peace in the country. All the insecurity and chaos is the result of Dinkas looting properties and grabbing lands which do not belong to them. This is done by the support of [Dinka government](#) under the overall command of Paul Malong. The [Dinka president](#) should lead by example and, leave LURI area where his cattle are kept and return to his homeland. There will never be peace in South Sudan until the Dinkas start to act like human beings.

### **Partnership for a Myanmar Dashboard**

A similar approach is being taken in Myanmar, where big data specialists are combining human and machine learning to identify hate speech online. The country's recent explosion in access to cheap connectivity and mobile handsets has coincided with an explosion in the use of hate speech, primarily found on Facebook pages. In 2014, the US Institute of Peace in Yangon sponsored a Peacetech Exchange workshop to develop solutions to hate speech. Participants from a new advocacy and research organization, the Yangon School of Political Science (YSPS), partnered with Bindez, an artificial intelligence and big data company, to develop a system for monitoring hate speech in the country. Starting with workshops of stakeholders in civil society and journalism, the participants identified a lexicon of terms that regularly inflame tensions online. These terms were used to "train" Bindez's machine learning software to collect, classify and analyze candidate terms. The outputs were provided to workshop participants for validation, which in turn improved the machine learning result-set.

Data visualization software provides a dashboard view that displays real-time trends in sentiment analysis (monitoring people's feelings) alongside potential instances of hate speech online. This partnership between YSPS and Bindez shows how humans and machine learning systems can combine to provide new datasets to address tensions in society which may drive conflict and forced migration. The human analysts can detect subtleties of meaning that a computer might miss, while

the computers rapidly process and manipulate volumes of data that humans cannot. YSPS has shared the results of the monitoring dataset with civil society organizations in both Yangon and Mandalay interested in countering hate speech online through dialogue, engagement and in some cases, reporting and requests for content removal.

### **Monitoring program impact**

For organizations seeking to help prevent and mitigate violent conflict, the use of big data to track hate speech within a region over time can give an indication of how instability is fluctuating. Interventions planned to reduce the impact of hate speech, such as conflict-sensitive journalism training to find alternative terms, or inter-ethnic dialogues, can be measured by whether the tone of conversation and the use of inflammatory terms have changed over time in social media, blogs and news media. The tracking of hate speech within a country that is struggling with refugee integration can be an effective method of tracking overall sentiment toward a refugee population.

Up-front investment in big data monitoring during program design can also contribute to long-term program evaluation. By tracking the tone of conversation and the use of hate speech, organizations can better understand the dynamics and size of refugee integration problems. Over time, such tracking can show whether interventions have had an impact.

## 2. OSRx – MONITORING RISK INDICATORS FOR FUTURE CONFLICT

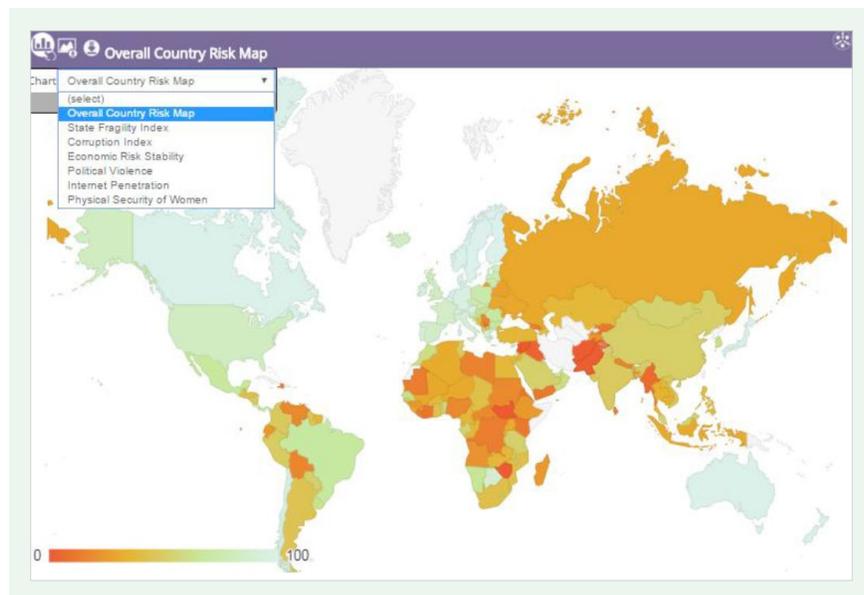
PeaceTech Lab researchers have also developed an analytics-based tool that aggregates and visualizes two large datasets capturing geospatial information on global protests, violent events and conflict.<sup>23</sup> The Open Situation Room Exchange (OSRx) is publicly available and provides a situational awareness platform to support peacebuilders and policymakers in contexts of rising tension and violent armed conflicts. The platform draws on large datasets from two sources:

- The Global Data on Events, Location and Tone (GDELT) platform, a comprehensive open database of human society, with over a quarter-billion records and news reports. GDELT monitors news media from every country in print, broadcast and web formats in over 100 languages, stretching from 1979 through present day, with daily updates, to

explore the frequency and concentration of violent events and protests.

- The Armed Conflict Location and Event Data (ACLED) project, which collects data on violence in developing states, covering dates and locations of political violence, the type of event (such as battles, killings, riots or protests), groups involved, fatalities and changes in territorial control.

OSRx produces open, easily navigable datasets that can inform policy decisions, with dashboards providing real-time news and social media analytics, peace and conflict indicators and risk forecasts of violence and political instability – as shown in the fragility map below. These can form the basis for predictions of forced migration and for pre-emptive policy responses.



Source: PeaceTech

### 3. PREDICTING MOVEMENTS OF PEOPLE FROM WEATHER DATA

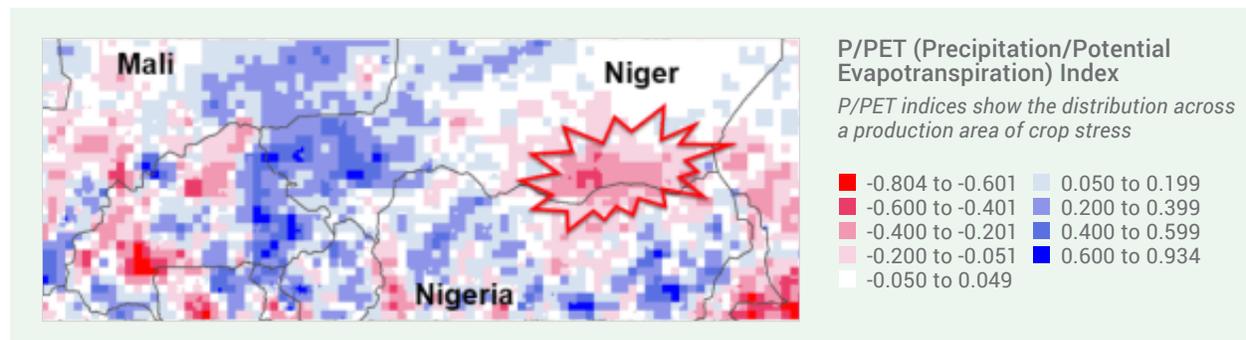
Big data analytics can also offer indicators of potential unrest from unexpected sources such as weather data. Very high-resolution agronomic weather data can identify “pocket droughts,” or areas where relatively small parcels of land experience season-long drought with significant agronomic impact. These compact yet intense droughts can lead to competition over scarce resources and can change migration patterns. In some cases, this may leave populations more vulnerable to extremist ideas and conflict, or contribute to population growth in megacities already experiencing significant unemployment.

Software company aWhere Inc harnesses agriculture analytics to drive improvements in farming. Its global-scale agronomic modeling system collects and creates over a billion points of data every day to construct daily “agro-meteorological” datasets for every “grid” of around nine kilometers in size. The system calculates current rainfall (P for precipitation) as a ratio of the evaporative demand of the environment (termed “Potential Evapotranspiration” or PET), then compares

this with the last five to 10-year average of the same P/PET ratio. This allows aWhere to provide real-time maps identifying droughts – including pocket droughts that may have implications for conflict.

The approach makes it possible to examine the historical dataset for a region beset by violent extremism, such as the Niger-Nigeria border, to see how weather could have predicted the potential for increased migration based on early indications of a pocket drought. In the map below, covering Eastern Mali, Burkina Faso, southern Niger and northern Nigeria, the gradient between the blue (wetter and lower PET) and red (drier and higher PET) depicts P/PET for June 1 to August 14, 2015, compared to the previous five years of ‘normal’ conditions.

The highlighted area, centered around Zinder, Niger, normally has an agro-climate type that can produce sorghum, but this was not possible in 2015. Even though the larger region was receiving rainfall, this area received far less rainfall than normal. Temperatures were



Source: PeaceTech

also higher and humidity lower than normal, compounding the impact of too little rainfall on local agricultural production. Displaced people were observed throughout the area south of the highlighted red area. This drought led to more food insecurity, with likely negative impacts including increased activity by terrorist group Boko Haram in the region.

Economic concerns along with poor governance, corruption and authoritarian rule have all been identified as drivers of violent extremism. Recent studies have also shown that **droughts lead to more extremist activity**, including civil conflict. By developing regional models based on weather data, host governments and international relief organizations can generate early warning signals when pocket droughts are forming. This allows them to take preventative measures for drought preparedness and agricultural assistance to stave off increased civil conflict.

Remote sensing data can also play a critical role in anticipating a population's movements, for example, in response to environmental stressors. By drawing on data that tracks climate change using geospatial imagery, and investigating the relationship between climatological events and migration patterns, policymakers can plan pre-emptive responses to natural disasters and climate change. For example, researchers at the University of Colorado Boulder have shown that Mexican rainfall estimates and other climate data from satellite imagery were predictive of domestic and international migration, particularly in regions where agriculture is the largest sector of the local economy.<sup>24</sup>

## 4. PREDICTING OUTBREAKS OF MASS ATROCITIES

Mass atrocities have, until recently, been viewed as extreme and unpredictable events, without pattern in day-to-day life. However, big data is making advances towards predicting violent events, potentially giving policymakers the opportunity to address pre-emptively the resulting forced migration. In 2013, in partnership with Humanity United, USAID held an innovation contest, the Tech Challenge for Atrocity Prevention, that invited data science and technology experts to develop big data technologies that could predict when and where mass atrocities would happen at a sub-national level. The challenge sought innovative solutions to prevent and mitigate mass atrocities.<sup>25</sup>

A joint Harvard and NASA Tournament Lab submission developed a set of algorithms to assess the risk of future atrocity events taking place on a month-to-month basis. The entry used data from past atrocities and the Global Data on Events, Location and Tone (GDELT) platform – a comprehensive open database that monitors news media from every country in print, broadcast and web formats translated from over 60 languages, from 1979 through present day. Researchers recently evaluated their algorithms by using them to assess whether they could have predicted the 2014 kidnapping of Nigerian schoolgirls in Borno State, Nigeria. They found that during the period from 2009 to 2012 covered by the model, Borno State was one of four regions in Nigeria at high risk of mass atrocity.<sup>26</sup> Although the data was not contemporary to the actual event, patterns of turbulence

in the region had identified the risk. The researchers claimed that if real-time data were available, the algorithm would be increasingly effective in providing policymakers with early warning of the risk of mass atrocities – and therefore of potential forced migration. These warning signals could prompt interventions before mass atrocities take place.

## 5. MAPPING SYRIA'S CONFLICT – THE CARTER CENTER

Other big data approaches also offer insights into conflict dynamics and resulting migration. Since 2014 the Carter Center's Syria Conflict Mapping Project has analyzed open-source data in order to profile the Syrian conflict. Run in collaboration with analytics software company Palantir, the project provides humanitarian organizations with insights on armed groups and power structures in the region. By mining social media for posts and videos from the conflict, project analysts have identified attributes for over 5,600 armed groups with some 100,000 fighters – including when and where events occurred, weapons possessed, the networks of militant groups, militant symbolism and uniforms, and evidence of mass atrocities.<sup>27</sup> Using Palantir Gotham, the company's leading data integration and analytics tool, researchers are able to find meaningful insights from vast sources of big data. These help decision-makers provide aid to people who need it most – including those displaced – while increasing the security of their workers through better situational awareness.



## 6. CROWDSOURCING FOR SITUATIONAL ANALYSIS

Crowdsourcing is also emerging as an important tool for profiling crisis situations as they develop. The open-source Ushahidi crowdsourcing platform, for example, was initially developed to map reports of post-election violence in Kenya in 2008, using information submitted via the web and mobile phones. It has since been repurposed for data collection and monitoring many crisis events, such as water shortages, earthquakes and natural disasters, as well as routine monitoring of high-crime areas or during elections.

The Ushahidi platform was deployed in Syria in 2012 to compliment an open-source web and social media tracking platform that mines thousands of online sources for evidence of human rights violations, killings, torture and detainment. The crowdsourcing tool, called Syria Tracker, coupled the open-source data with crowd-sourced human intelligence, such as field-based eye-witness reports shared via webform, email, Twitter, Facebook, YouTube and voicemail.<sup>28</sup> Using this approach, the Syria Tracker team and its relatively small group of volunteers have verified almost 90 percent of the documented killings mapped on the platform, thanks to video or photographic evidence. They have also named around 88 percent of those reported killed by Syrian forces since the uprising began. Depending on the levels of violence, the turnaround time for a report to be mapped on Syria Tracker is one to three days. The team produces weekly situation reports based on the data collected, along with

detailed graphical analysis. Files providing a more precisely geo-tagged tally of deaths per location are made available regularly and can be uploaded and viewed using Google Earth. This approach could easily be applied to other contexts and issues, including any situation of mass forced migration.



## 7. THE BIGGER PICTURE – PROFILING NEEDS THROUGH SATELLITE IMAGERY

In 2015, Nigeria's government asked the Global Facility for Disaster Reduction and Recovery (GFDRR) for assistance in assessing the needs of people affected by the conflict between the country's security forces and terrorist group Boko Haram. The conflict has affected nearly 15 million people since 2009. In 2014, the violence intensified, and by 2017 had caused an estimated 20,000 deaths and displaced nearly 2 million people. The widespread destruction and massive displacement have resulted in acute service delivery and infrastructure shortages, as well as an urgent need for peacebuilding initiatives.

A grant-funding mechanism managed by the World Bank, GFDRR is a global partnership that provides knowledge, funding and technical assistance to disaster risk management projects worldwide. The facility works on the ground with numerous local, national, regional, and international partners. In Nigeria, it needed to survey six states in order to provide an overarching framework for stability, peacebuilding and recovery. This raised challenges in assessing the needs of so many people over large areas with sparse and damaged infrastructure.

In response, GFDRR provided expert assistance in the use of assessment technologies based on satellite imagery – one of the first times such imagery was used extensively in a large-scale assessment. The approach enabled gap-filling and validation of baseline and damage assessment

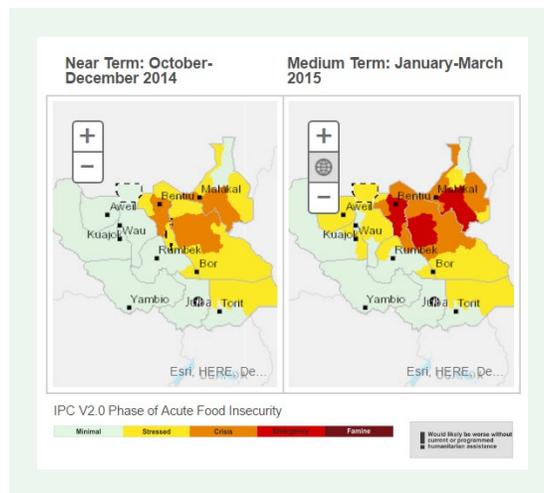
data, and provided the foundation for the estimation of need across numerous sectors, including housing, the environment, health and education. To interpret the data, the GFDRR coordinated sector experts from the EU, the UN and the World Bank, as well as Nigerian officials, and provided technical advice to help characterize and estimate the impact of the crisis.

Together they produced a Recovery and Peace Building Assessment, which profiled in detail the impact of conflict and gave a framework for support to assist people affected. It identified and prioritized needs and made recommendations for peacebuilding, economic recovery and restoring critical infrastructure and services. Drawing on sources such as the Landsat remote sensing project operated by the US Geological Survey and NASA, this satellite-based approach could be used to track forced migration and its causes (such as conflict or reduced water sources) and consequences (such as camps and urban sprawl). Detailed images taken from satellites and unmanned aerial vehicles are being used to track the growth or contraction of refugee camps in Jordan and Syria, for example.<sup>29</sup> The UN Operational Satellite Applications Program (UNOSAT) captures and analyzes images, generating data to help decision-makers allocate resources and deliver services for maximum impact in helping forced migrants.

## 8. USAID FAMINE EARLY WARNING SYSTEMS NETWORK

Big data approaches are also enabling projection of forced migrants' needs, through techniques such as crowdsourcing and satellite imagery. The Famine Early Warning Systems Network (FEWS NET), for example, provides early warning indicators and analysis of food insecurity to inform decision-makers responding to humanitarian crises. Created by the US Agency for International Development (USAID), the system has been used to assess forced migration in response to food insecurity and conflict in places such as Unity State, South Sudan, a zone of conflict with neighboring Sudan. In partnership with DigitalGlobe, a US commercial satellite network operator, FEWS NET assessed existing food insecurity outcomes for Sudanese refugees and South Sudanese IDPs in Unity, and projected future outcomes. The project was based on analysis of big data from satellite imagery and crowdsourced imagery from over 25,000

volunteers using DigitalGlobe's crowdsourcing platform Tomnod. This harnesses gamification and community-building techniques to build a large base of volunteers who enjoy examining satellite imagery. In just two weeks of analysis, FEWS NET identified 46,000 permanent shelters, temporary shelters and herds of cattle over 14,000 square kilometers.<sup>30</sup> In combination with existing data on food production in the region, FEWS NET was able to project demand for food, assessing that acute food insecurity in South Sudan would intensify to an "Emergency" classification due to escalating conflict and the resulting limited access to food and income sources by people forcibly displaced (as shown in the map).<sup>31</sup> This approach can be applied to other contexts, and has been used to pinpoint areas of concern and high concern in countries including Ethiopia, Nigeria, Yemen, Somalia and Afghanistan.



## 9. MIGRATION TRACKING FROM POPULATION-BASED MODELS

An essential component of effective decision-making to target resources to help displaced people is accurate, timely data on population movements and distributions in low- to middle-income countries. To help source such data, a consortium of researchers has collaborated with the Flowminder Foundation and its open-source demographic analytics resource WorldPop to aggregate publicly available population and spatial data, and model migration patterns in Sub-Saharan Africa. Enabled by new big data sources such as ongoing demographic surveys, satellite imagery, urban data and geographic information system (GIS) statistical modeling, the consortium has created powerful tools to understand and predict migration in regions where data is traditionally sparse, offering a powerful input to international development planning and execution.

The approach combines big data from spatial datasets on 10 countries in Sub-

Saharan Africa with local census data to develop a comprehensive view of migration. Although traditional national-level censuses can record migratory movement, complete census data is often difficult to obtain for low-income countries, and privacy concerns in such countries often inhibit deep analysis of this data. Census “microdata,” however, are becoming increasingly available for public use and include anonymized samples that provide demographic and migration information about the population concerned. Using this data from the 10 countries and pairing it with previously identified push-pull factors for migration in low-income countries (for example, geographic, sociodemographic, economic, climatic, environmental), the researchers created a series of gravity-type spatial interaction models (GTSIMs), the researchers created a series of gravity-type spatial interaction models (GTSIMs), which predict the amount or likelihood of people, goods, services or information moving



between two locations. This allows analysts to understand the migration dynamics and key drivers for the populations examined.<sup>32</sup> Through the models, Flowminder explored two questions: where have the populations migrated to, and why? By analyzing the underlying variables, researchers were able to explain the key drivers for migration in Sub-Saharan Africa, including the effects of climate change and changing social, demographic, economic and environmental contexts based on conflict events, political instability or commercial interests. They also assessed the predictive capabilities of their models and found them able to explain up to 87 percent of internal migration, predict future internal migration with correlations of up to 0.91, and predict migration to other countries with correlations of up to 0.72.<sup>33</sup> By showing the similarity between actual five-year census data on movement between the provinces of South Africa (map A) and the model's projections for the same period (map B), the researchers demonstrate this potential to predict

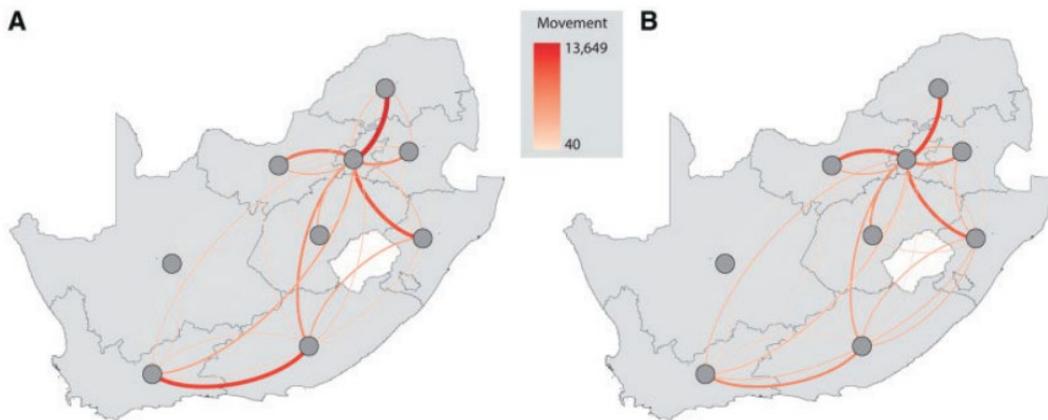
migration flows. This gives policymakers a warning signal to develop earlier responses.

### Tracking migration to save lives in Nepal

In a real-time crisis situation, Flowminder's migration tracking played an integral role after the devastating earthquake which struck Nepal in April 2015. Large population movements occur after most major disasters, but very limited information has been available to understand where affected people are located once these movements have begun. Flowminder researchers pioneered the use of anonymous mobile network operator data to assess population displacement after the 2010 Haiti earthquake and cholera outbreak. These data enable understanding of where affected people are located, essential for effective humanitarian response operations.

In light of Nepal's long-recognized risk of severe earthquakes, in December 2014 Flowminder and mobile network operator Ncell agreed a collaboration to respond to

Actual Five-Year Internal Migration in South Africa Compared to Researcher's GTSIM



Source: Flowminder Foundation et al.



future earthquakes and to underpin long-term development objectives in Nepal. With support from the [Rockefeller Foundation](#), the project was initiated and a rapid response capacity set up by the Flowminder team in Kathmandu just one week before the earthquake occurred. With a moment magnitude of 7.8 and more than 300 aftershocks, it killed more than 9,000 people and injured 23,000, destroying or severely damaging an estimated 500,000 buildings.

In the aftermath, mass population movements were immediately apparent as people fled affected areas.

Using their pre-earthquake preparations, Flowminder could quickly start analysing Ncell's data showing people's movements as they travelled between individual cell phone transmitter towers. Following the initial earthquake, the WorldPop project

mapping team rapidly produced updated static population density maps for Nepal, with a spatial resolution of 100x100 meters, including gender and age distributions. These data were used by the UN Office for the Coordination of Humanitarian Affairs (OCHA) and other key relief agencies in estimations of the number of people affected. The Flowminder and WorldPop mobile analysis team then used these population data in combination with the Ncell anonymised data from 12 million mobile phones in Nepal to quantify the impact of the earthquake on population movements.

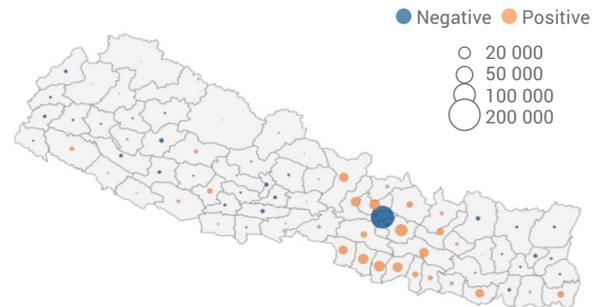
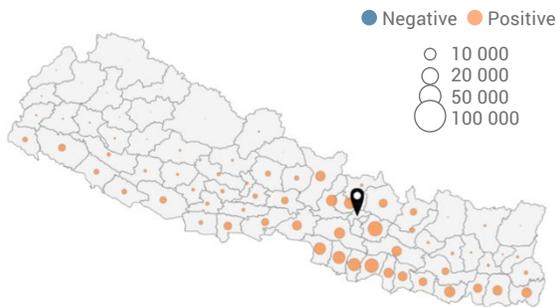
Despite pre-earthquake preparations not being completed, the team successfully published the first comprehensive analyses of population displacement less than two weeks after the earthquake. These results were released as a report to the UN and a range of relief agencies. After adjusting for normal population

movement patterns which would have occurred in the absence of the earthquake, the analysis showed that an estimated additional 500,000 people had left Kathmandu two weeks after the earthquake. The majority of these went to the districts surrounding Kathmandu and to the Terai areas in the south and south-east of Nepal – information completely

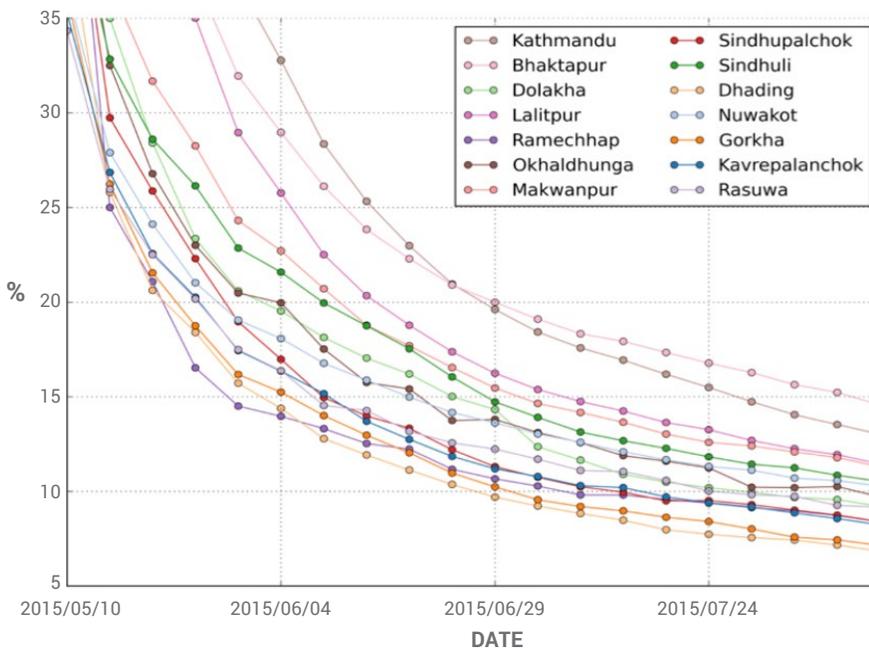
unknown previously. An estimated 1.8 million people above normal levels had left their home districts. After the initial report, a number of further reports were released, published either by Flowminder directly or in collaboration with OCHA, and disseminated through official OCHA information channels for response efforts.

Among people with homes in Kathmandu Valley, an estimated 500,000 more than expected had left the valley.

Date 5th May 2015; estimated above-normal number of people inside each district, who have homes in other districts and moved from their home district sometime after the earthquake. Inflows to a district are composed of people leaving their home district to come into the new district and Nepalese relief workers (relief workers coming into Nepal after the earthquake are excluded).



Proportion of those who left their home district within the first two weeks of the earthquake and who had not yet returned at consecutive times after the earthquake.



Flowminder continues to analyze post-earthquake population movements in Nepal and is now working with key organizations to ensure that the capacity to respond with similar analyses in future earthquakes and disasters can be maintained. It is also planning for the implementation of several analysis products to support development and public health in Nepal.<sup>34</sup>

### **Predicting migration, testing responses**

Other types of model to predict migration flows can also be built from analysis of cell phone data. Researchers at the Harvard School of Public Health noticed that callers near a tower in Kericho, Kenya were leaving the region with abnormal frequency. They were also likely to travel to a specific area near Lake Victoria that had been noted by public health officials as a hotbed for malaria. This analysis prompted the officials to use satellite imagery to explore the potential cause of the disease, which they identified as a tea plantation with many migrants working in close proximity.<sup>35</sup> At the University of Birmingham, researchers used data derived from call detail records (CDRs) from the Ivory Coast to develop a mobility model to predict a hypothetical epidemic spread of disease. Drawing on data on the population's mobility and socio-geographic ties, they investigated the effectiveness of two potential containment strategies: geographic quarantines and peer-to-peer information campaigns.<sup>36</sup> These experimental approaches show that by aggregating CDRs to measure the mobility of a population and understand its behavioural patterns, decision makers can obtain unprecedented information on migration flows and test potential policy responses.

## **10. GEOTAGGING INTERNET DATA TO FOLLOW MOVEMENTS**

By revealing people's mobility patterns during peace and stability, big data is helping policymakers understand where they will seek refuge when facing conflict or under duress. This supports effective allocation of relief efforts for migrants. In 2013, researchers at Queens College CUNY, the Qatar Computing Research Institute and Stanford University used geo-tagged login information for Yahoo! users to track their international mobility.<sup>37</sup> They analyzed a sample of 100 million anonymized uses of Yahoo! web services from July 2011 to July 2012. From user login records, researchers identified "migrants" by selecting users who spent at least 90 days in two countries over the year, as opposed to "tourists," who spent more than three months in one country and spells of less than one month in another. The analysis yielded conditional probabilities of "migration" and "tourism" between countries. These propensities were mapped for each region, capturing conditional migration propensity and the relative frequency of migration. The researchers then applied these probabilities against several external data elements, including whether countries had colonial ties, a shared language and large volumes of bilateral trade.

Their findings showed that two countries' locations in the same geographic region has a relatively minimal effect on migration, but a much greater effect on tourism, suggesting that long-term migrants tend to move to new regions entirely. International connectivity

was greatly facilitated by common language, and the higher a destination country's GDP, the higher the odds of migration to that country. The study shows how analysis of internet data could help policymakers predict destination countries for refugee flows, based on their countries of origin.

Researchers at Georgetown University's Institute for the Study of International Migration have also mined the internet, to create a vast database that seeks to preemptively identify threats likely to lead to population dispersion. With over 600 million publicly available news articles, stored in an expandable open-source database (formerly called the Raptor dataset), the tool contextualizes past sources of "dread threat" (for example, the 2006 Somali civil war). In the near future, tools such as this may allow governments to develop more accurate predictions to identify the probability of violent conflict and act against potentially catastrophic violent events before they occur<sup>38</sup>. The Institute has received funding to explore the possible incorporation of social media data, mobile data and geo-tagged images to enrich its data further.

## 11. THE WORLD BANK'S RESEARCH INSIGHT TOOL

To facilitate post-crisis collaboration and insight-driven responses, the World Bank Group has created the Research Insight Tool (RIT). This knowledge management and discovery platform aggregates big data from news articles, social media and sectoral sources such as the UN's Humanitarian Data Exchange; data collection sites such as the Global Data on Events, Location and Tone (GDELT) platform and the Armed Conflict Location and Event Data (ACLED) Project; the Geneva Declaration's Armed Violence Map; the UN Global Disaster Alert and Coordination System, and crisis-specific sources such as the Carter Center's Syria Conflict Mapping Project. It analyzes text and curates the data by tagging topic areas and key information, enabling a full text search capability. Users can visualize the data from their selected sources, enabling accessible analysis, and can easily share insights with others.

RIT has helped team members from the World Bank's Global Practices in numerous crisis situations, providing pertinent insights to inform decision-making. The tool helped underpin the Bank's efforts to monitor the roles of women and gender in the Syrian Civil War, by analyzing recurring themes in news media relating to female fighters and activists, restricted mobility, early marriage and the demographics of IDPs and refugees. In Iraq, RIT helped users identify and profile cities where government forces had retaken control from ISIS. By assessing demographics and infrastructural damage in these cities, RIT enabled the Bank to make policy recommendations on where to focus post-conflict needs assessments and to understand how local demographics could affect relief efforts.

## 12. ANALYZING SOCIAL MEDIA FOR RESPONSES TO MIGRATION

In 2017, the UN Global Pulse and the UN High Commission for Refugees (UNHCR) published analysis of how aspects of the European refugee crisis were conveyed on Twitter.<sup>39</sup>The project team initially set out to explore the value of social media for monitoring both migrants' interactions with service providers and public views towards migrants, with levels of nuance difficult to attain through traditional tools such as surveys. However, based on inconclusive initial results and anticipation of more negative public views towards migrants following the 2015-16 terrorist attacks in Europe, the project refocused on analysis of host communities' sentiment towards migrants in cities that had experienced terrorist attacks.

Between 2015 and 2017, the project carried out 10 "micro-studies", creating a unique monitor for each study using the Crimson Hexagon social media analysis tool. The first six monitored tweets within Greece, covering two subject areas: interactions between migrants and service providers in Europe, and xenophobia. Interactions were monitored in Arabic and Farsi, and xenophobia in Arabic, Farsi, Greek and English. After difficulties emerged, including separating migrant tweets from postings by diaspora members or people simply tweeting in various languages, the researchers refocused on host communities' sentiment towards migrants. The study analyzed posts relating to the attacks in Nice, Munich, Saint-Etienne and Berlin. In total, they covered more than 400 million Twitter posts, over a 22-month period.

The findings revealed that within local active Twitter communities, only a small number of people connected migrants and the terrorist attacks. Most importantly, the project showed that being able to assess peoples' views in real-time provides a unique opportunity for UNHCR to counter negative public online behavior. It also allows the agency to better understand generalized perceptions with a view to formulating longer-term solutions to help forced migrants. Understanding migrants' and host communities' mutual sentiments reveals how both groups view and react to asylum conditions and protection, which can enhance program design and planning strategies.

However, the project also demonstrated that social media analysis alone can seldom provide a comprehensive overview of needs and opinions. Tweets are generally not representative of socio-economic diversity and age, and even advanced machine learning and geo-referencing capabilities may not always be entirely accurate. For example, the project's geo-based queries may have retrieved tweets posted outside its defined geographic boundary (false positives), while omitting others posted inside the boundary (false negatives).

The project highlighted that working with social media requires a dynamic mindset. The researchers needed to adapt and iterate rapidly. They set out to use social media posts to build a better, more nuanced understanding of complex aspects of the European refugee

crisis. However, as circumstances evolved, they adapted to focus on using social media to detect unexpected signs of opinion or ongoing events that could put migrants at risk. The streaming nature of social media posts enables the detection of such signals in near-real-time, potentially flagging up events that UNHCR may need to respond to quickly. Few data sources can facilitate such in-depth, rapid response mechanisms, and the UN is continuing to explore their potential.





# Conclusion: Big Data for Migration Management of the Future

Experiments in the use of big data for migration management are providing new sources of information and approaches for policymakers and humanitarian response teams. Informed and coordinated planning is needed to respond to forced migration in an effective, sustainable way. This must go beyond meeting migrants' immediate needs and consider longer-term solutions to mass, unplanned movements of people. To plan effectively, stakeholders need detailed descriptive and predictive information on both macro and micro levels – from large-scale movements such as refugee flows from Syria, to the needs of individual communities and citizens. Migration management of the future must be based on accurate projection models

and real-time information to enable optimal resource allocation, effective protection of those displaced, and active improvement of migrants' lives and those of host communities.

Developments in big data analytics are increasingly able to supply such accurate real-time information and predictive models. Future big data tools will leverage stronger remote sensing capabilities to develop deeper contextual analysis. This will enable policymakers to use detailed real-time information on factors such as population density or socioeconomic indicators when formulating responses to forced migration flows and planning pre-emptive strategies. Increasingly, policymakers' understanding of forced migration patterns and the needs of those displaced will be based on information sourced directly from forced migrants and members of their host communities. Platforms such as Syria Tracker are already redefining the role of community-sourced information in crisis response, and as the drivers of violent extremism and the resulting migration become better understood, big data methodologies will generate improved

predictive analytics for migration. Weather data layered onto geospatial data will also provide key decision support tools by identifying problems which can lead to forced migration, such as near-term droughts – especially in areas already beset by violent extremist activity.

### **Translating Big Data into Optimal Responses**

Use of big data has important implications for organizations seeking to help prevent and mitigate forced displacement. The use of on big data to track hate speech, for example, can give an indication of whether instability is increasing or decreasing. Up-front investment in big data monitoring during program design can also contribute to long-term program evaluation. Approaches such as sentiment analysis of social media can reveal the dynamics and scale of refugee integration problems, and can show over time whether interventions have had an impact.

The World Bank Group responds to forced displacement along several tracks. Among them is its work with global, regional and national partners to improve the data and evidence base for better policies and programming. The Bank collaborates with data producers and users to assess the best way to support countries in strengthening their migration data. This includes efforts to improve data quality and coverage, understand the impacts of migration on poverty and economic growth, and design policy recommendations for countries sending and receiving migrants. Big data offers increasing potential for demand-driven tools to help policymakers and response agencies achieve the best use of resources, and the best outcomes for migrants and their hosts. By

harnessing innovations such as those profiled in this brief, and using them alongside its wealth of data and policy-oriented information, the World Bank can provide cutting-edge, customized knowledge on migration to its clients.

### **A Rich Learning Process**

The potential for big data is huge, but it is not a magic bullet. Analytics alone cannot create better outcomes. It can inform the development and implementation of solutions, but government policies, investment, financing, local support and capable institutions are also essential. Many challenges still exist to the scalability of big data in low- and middle-income contexts. The most salient are barriers to accessing data, including individual privacy concerns around the use of big data and the ability to access proprietary data. These challenges must be overcome through the development of policies for responsible use of big data and the establishment of public-private partnerships.

Less developed contexts may lack the technological resources to collect data, including smartphones, sensors or access to cloud-based processing. Biases can be introduced when data is not representative – for example, the lower end of income distribution can be missed from data collected through cell phones. While not an important issue in higher-income countries, bias is likely to be a serious challenge in lower-income contexts. Big data approaches in such contexts cannot rely on a hardwired data collection infrastructure or high-tech tools such as GPS-enabled smartphones. Rather, they may be best applied to phenomena that can be tracked using

cheaper passive systems, such as satellite imaging, cell phones and crowdsourcing.

Innovative paths inevitably involve hurdles, reveal useful lessons and require perseverance. Rigorous validation or “ground-truthing” of new methodologies is essential. Big data demands that users capture, prepare and store data meticulously – and plan enough time to do so. To enable big data analytics, it is essential that authorities, policymakers, non-governmental organizations and other stakeholders promote the public availability of useful data from a wide range of sources (with policies to support its use).

Successful big data solutions often involve combining previously unrelated data sources or approaching existing situations from new angles – such as using remote sensing data, captured for other purposes, to assess factors such as the extent of refugee camps. Despite the central role of computational power, the human element also remains vital to the success of these projects – as clearly shown in the UN’s analysis of social media postings concerning migrants to Europe. Big data analysis can often be enhanced by traditional research techniques, such as censuses and household surveys. There is also need to invest in partnerships, or to combine human and computational analytical power for optimum results. Big data analytics is a team sport: Effective collaboration between data experts, technologists and business sector specialists is crucial.

Despite existing restrictions, the potential for big data to streamline policy and aid responses to forced migration remains significant. It can help bring about step changes, from local to pan-

national levels. Initiatives such as those profiled in this brief are just the beginning – both in terms of their own lifecycle and impact, and in terms of the variety of ways in which big data analytics will help improve migration management. In this era of experimentation, there are many more big data approaches on the horizon, which look set to enhance forced migration management and response planning in the coming decade.

### **Looking Forward: Mainstreaming Big Data for Managing Forced Migration**

Responsibly sourced and prudently used, big data is a public good when operationalized to help forcibly displaced people. It shows great promise in providing policymakers with new insights into the dynamics of forced migration as a result of fragility, conflict and violence. It can significantly transform the way interventions are designed and implemented, bringing great value to governments, organizations and – most importantly – refugees and IDPs. By convincing stakeholders that big data is a public good through responsible use, transparency and active partnerships, development organizations can advance the scalability of big data for development.

All these novel data approaches must be tested, validated and adapted for mainstream use – but the potential rewards of big data make such effort worthwhile. As the case studies in this brief show, big data analytics can improve understanding of forced migration, and the planning and delivery of responses. It can help initiatives within and beyond the World Bank achieve results through improved evidence, efficiency, awareness, understanding and forecasting. Ultimately, big data initiatives can be a powerful accelerator for ending poverty and boosting shared prosperity.

# References

- <sup>1</sup> United Nations High Commissioner for Refugees, "Worldwide displacement hits all-time high as war and persecution increase," last modified June 18, 2015, <http://www.unhcr.org/558193896.html>.
- <sup>2</sup> Ibid.
- <sup>3</sup> Dario Carminati, "Climate change and displacement: Protecting whom, protecting how?," last modified June 10, 2013, <https://phap.org/articles/climate-change-and-displacement-protecting-whom-protecting-how>.
- <sup>4</sup> Michael Werz and Laura Conley, "Climate Change, Migration, and Conflict: Addressing Complex Crisis Scenarios in the 21st Century," last modified January 3, 2012, <https://www.americanprogress.org/issues/security/report/2012/01/03/10857/climate-change-migration-and-conflict/>.
- <sup>5</sup> Mark Tran, "Forced migration in the 21st century: urbanized and unending," *The Guardian*, last modified October 16, 2012, <https://www.theguardian.com/global-development/2012/oct/16/forced-migration-21st-century-urbanised-unending>.
- <sup>6</sup> Marzia Rango, "How big data can help migrants," *World Economic Forum*, last modified October 5, 2015, <https://www.weforum.org/agenda/2015/10/how-big-data-can-help-migrants>.
- <sup>7</sup> National Research Council, *Beyond Six Billion: Forecasting the World's Population*, Panel on Population Projections, Committee on Population, Commission on Behavioral and Social Sciences and Education, Washington, DC (2010), <http://www.nap.edu/read/9828/>.
- <sup>8</sup> Anne Davies and Karen Jacobsen, "Profiling urban IDPs," *Forced Migration Review*, Issue 34 (2010): 13-15, <http://www.unhcr.org/4b97b0fa9.pdf>.
- <sup>9</sup> Ibid.
- <sup>10</sup> Jean-Marie Guehenno, "Conflict is Key to Understanding Migration," last modified May 13, 2016, <https://www.crisisgroup.org/en/regions/op-eds/2016/guehenno-conflict-is-key-to-understanding-migration.aspx>.
- <sup>11</sup> Ari Phillips, "Natural Disasters Displaced 22 Million People Last Year, 3 Times More Than War," last modified September 17, 2014, <http://thinkprogress.org/climate/2014/09/17/3568447/natural-disasters-displaced-22-million-people/>.
- <sup>12</sup> Giulia Bettin and Alberto Zazzaro, "The impact of natural disasters on remittance inflows to developing countries," last modified November 27, 2012, [https://editorialexpress.com/cgi-bin/conference/download.cgi?db\\_name=CSAE2013&paper\\_id=393](https://editorialexpress.com/cgi-bin/conference/download.cgi?db_name=CSAE2013&paper_id=393).
- <sup>13</sup> Internal Displacement Monitoring Centre, *Global Estimates 2015: People displaced by disasters*, last modified July 21, 2015, <http://www.internal-displacement.org/publications/2015/global-estimates-2015-people-displaced-by-disasters/>.
- <sup>14</sup> Liz Heimann, "Top 10 of 2015 – Issue #7: Climate Change and Natural Disasters Displace Millions, Affect Migration Flows," last modified December 10, 2015, <http://www.migrationpolicy.org/article/top-10-2015-issue-7-climate-change-and-natural-disasters-displace-millions-affect>.
- <sup>15</sup> Ibid.
- <sup>16</sup> Ibid.
- <sup>17</sup> Myers, Norman, "Environmental Refugees: An Emergent Security Issue", published by the Organization for Security and Co-operation in Europe, May 25, 2005
- <sup>18</sup> Michael Werz and Laura Conley, "Climate Change, Migration, and Conflict: Addressing Complex Crisis Scenarios in the 21<sup>st</sup> Century," last modified January 3, 2012, <https://www.americanprogress.org/issues/security/report/2012/01/03/10857/climate-change-migration-and-conflict/>.

- <sup>19</sup> Climate Change, Environment and Migration Alliance, "Climate Change, Environment and Migration: Frequently Asked Questions," last modified December 2010, [https://www.iom.int/jahia/webdav/shared/shared/mainsite/activities/env\\_degradation/CEEMA\\_top\\_10FAQs.pdf](https://www.iom.int/jahia/webdav/shared/shared/mainsite/activities/env_degradation/CEEMA_top_10FAQs.pdf).
- <sup>20</sup> Elizabeth Stites, John Burns, and Darlington Akabwai, "*It's Better to Sweat than to Die: Rural-to-Urban Migration, Northern Karamoja, Uganda*," Feinstein International Center (2014), [http://fic.tufts.edu/assets/TUFTS\\_1474\\_Migration\\_NKaramoja\\_Uganda\\_V5\\_online.pdf](http://fic.tufts.edu/assets/TUFTS_1474_Migration_NKaramoja_Uganda_V5_online.pdf).
- <sup>21</sup> Ibid.
- <sup>22</sup> Patricia Weiss Fagen, "Flight to the cities," *Forced Migration Review*, Issue 45 (2014): 14-17, <http://www.fmreview.org/sites/fmr/files/FMRdownloads/en/crisis.pdf>.
- <sup>23</sup> PeaceTech Lab, "About the Open Situation Room Exchange," <http://www.osrx.org/about>.
- <sup>24</sup> Dan Miller Runfola, Fernando Riosmena, et al., "Internal and International Mobility as Adaptation to Climatic Variability in Contemporary Mexico: Evidence from the Integration of Census Data with Satellite Imagery," Institute of Behavioral Sciences, University of Colorado Boulder, December 2013, <http://www.colorado.edu/ibs/pubs/pop/pop2013-0005.pdf>.
- <sup>25</sup> Tech Challenge to Prevent Mass Atrocities, "Meet the Tech Challenge Winners," [http://www.thetechchallenge.org/ttc\\_archive/winners.html](http://www.thetechchallenge.org/ttc_archive/winners.html)
- <sup>26</sup> Emily Calandrelli, "Could we predict the kidnapping of Nigerian girls?" last modified May 19, 2014, <https://emilycalandrelli.wordpress.com/tag/atrocity-prevention/>
- <sup>27</sup> Palantir, "Mapping the Syrian Crisis with the Carter Center," *The Palantir Blog*, last modified June 2014, <https://www.palantir.com/2014/06/mapping-the-syrian-crisis-with-the-carter-center/>.
- <sup>28</sup> Jim Giles, "Mapping the human cost of Syria's uprising," *New Scientist*, March 21, 2012, <https://www.newscientist.com/article/mg21328576-000-mapping-the-human-cost-of-syrias-uprising/>
- <sup>29</sup> Joanne Wheeler, "People Watching: Tracking and Monitoring Migration," last modified November 16, 2015, <http://interactive.satellitetoday.com/people-watching-tracking-and-monitoring-migration/>.
- <sup>30</sup> Rhiannan Price, *Crowdsourcing & Satellite Imagery: For an evidence-based food security assessment in South Sudan*, DigitalGlobe, presented in discussion with the authors.
- <sup>31</sup> FEWS NET, "Crisis persists in Greater Upper Nile, food security to deteriorate in early 2015," *Food Security Outlook*, <http://www.fews.net/east-africa/south-sudan/food-security-outlook/october-2014>.
- <sup>32</sup> Andres J. Garcia, Deepa K. Pindolia, Kenneth K. Lopiano, and Andrew J. Tatem, "Modeling internal migration flows in sub-Saharan Africa using census microdata," *Migration Studies*, August 4, 2014, <http://migration.oxfordjournals.org/content/early/2014/08/04/migration.mnu036.full.pdf>.
- <sup>33</sup> Ibid.
- <sup>34</sup> PLoS Currents: Disasters paper, <http://currents.plos.org/disasters/article/rapid-and-near-real-time-assessments-of-population-displacement-using-mobile-phone-data-following-disasters-the-2015-nepal-earthquake/>
- <sup>35</sup> David Talbot, "Big Data from Cheap Phones," *MIT Technology Review*, <https://www.technologyreview.com/s/513721/big-data-from-cheap-phones/>.
- <sup>36</sup> A. Lima, M. De Domenico, V. Pejovic, and M. Musolesi, "Exploiting Cellular Data for Disease Containment and Information Campaigns Strategies in Country-Wide Epidemics," School of Computer Science, University of Birmingham, <http://www.ucl.ac.uk/~ucfamus/papers/d4d.pdf>.

<sup>37</sup> Bogdan State, Ingmar Weber, and Emilio Zagheni, "Studying Inter-National Mobility through IP Geolocation," *International Conference on Web Search and Data Mining 2013*, February 4-8, 2013, <http://web.stanford.edu/~bstate/papers/wsdm46-state.pdf>.

<sup>38</sup> Institute for the Study of International Migration, "Forecasting the Break: Building Community and Capacity for Large-Scale Data Intensive Research in Forced Migration Studies," <https://isim.georgetown.edu/forecasting>.

<sup>39</sup> UN Global Pulse, UNHCR Innovation Service, 'Social Media and Forced Displacement: Big Data Analytics & Machine-Learning', 2017

## Acknowledgements

This solutions brief is one of several sector-focused big data knowledge products delivered through the "Innovations in Big Data Analytics" program which resides in the Global Operations and Knowledge Management Unit (GOKMU). It has been prepared jointly in cooperation with Deloitte Consulting and World Bank units working in Disaster Risk Management; Fragility, Conflict and Violence; and Forced Displacement. Stephanie Debere (WBG Consultant) is the lead author of the publication, Trevor Monroe and Dharana Rijal coordinated the publication, Greg Berger is the lead designer on the publication. A number of WBG staff provided valuable contributions to make this publication possible, including peer reviewers Kristen Schuettler, Raja Arshad, Caroline Bahnson, and Kristina Mikulova.

The *Innovations in Big Data Analytics* program works to accelerate the effective use of big data analytics across the World Bank, and to position the World Bank as a leader in the big data for development community.

For **additional information** about this brief or to find out more about the program, please visit <http://bigdata> (WBG intranet) or Trevor Monroe ([tmonroe@worldbank.org](mailto:tmonroe@worldbank.org)).



