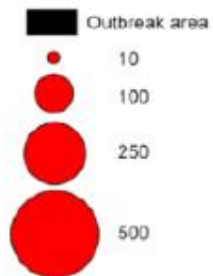# Big Data for Migration:
Uses, opportunities and challenges

United Nations Expert Group Meeting
Improving Migration Data in the Context of the 2030 Agenda

UN Headquarters, New York, 20—22 June 2017

Average daily numbers of sims that moved out from the communal sections surrounding Saint-Marc, Oct 15 to Oct 23, 9:00 am, 2010.
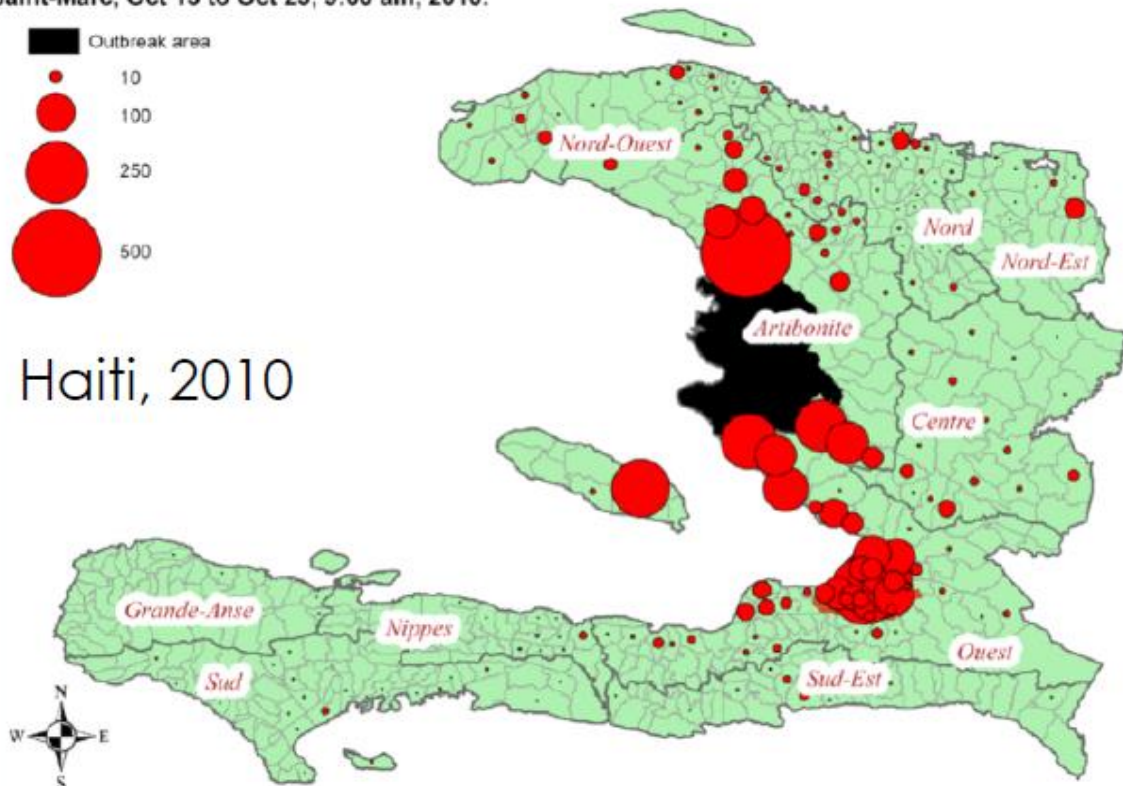
Haiti, 2010

Figure 5. Average daily numbers of SIMs moving out of the cholera outbreak area. October 15 to October 23, 2010, divided per communal section of destination. The data were disseminated to relief agencies at the outset of the outbreak (October 24, 2010).
doi:10.1371/journal.pmed.1001083.g005

Outline

Source: Eazystock blog

1. **The rationale**
2. **The potential**
3. **The challenges/opportunities**
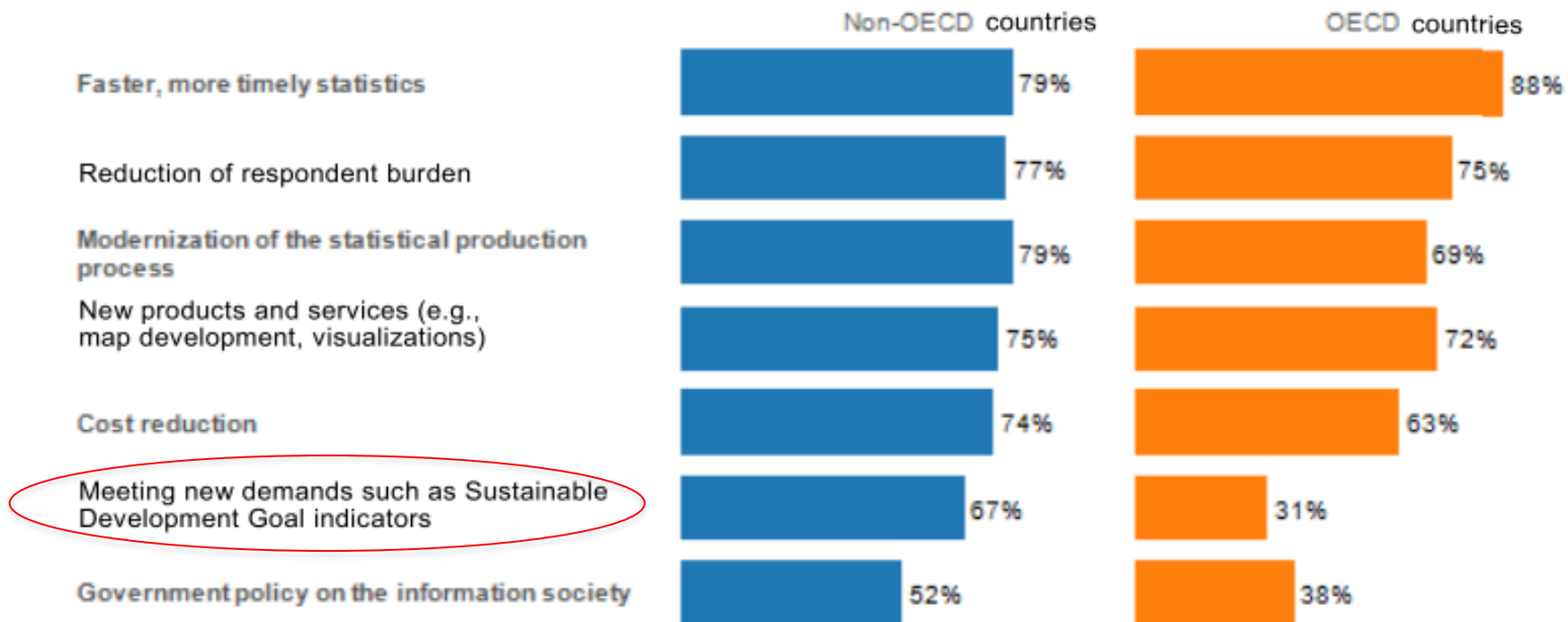
1. The rationale

a) The global context

Migration at the top of the global policy agenda → Renewed calls for enhanced migration data

- New York Declaration for Refugees and Migrants (Global Compact)

- Agenda 2030 for Sustainable Development – *leaving no one behind*

- Push for a "Development Data Revolution" to support SDG monitoring

- Creation of a Global Working Group on Big Data for Official Statistics

- No agreement on a global action plan:
    - How to make progress ?
    - What should be the priorities ?
    - How to utilize **new data sources** ?

Main reasons for and business benefits of using big data

based on responses from 93 countries (32 OECD, 61 non-OECD)

	Non-OECD countries	OECD countries
Faster, more timely statistics	79%	88%
Reduction of respondent burden	77%	75%
Modernization of the statistical production process	79%	69%
New products and services (e.g., map development, visualizations)	75%	72%
Cost reduction	74%	63%
Meeting new demands such as Sustainable Development Goal indicators	67%	31%
Government policy on the information society	52%	38%

b) The (persistent) issues with migration data

- **Lack of reliable, comparable, timely and disaggregated data on migrant stocks and flows**

- **Limitations of traditional sources of migration data**

    - **Population censuses:** main source but infrequent
    - **Household surveys:** costly and methodologically problematic (sample size)
    - **Administrative records:** events, not people

- **(Some of the) knowledge gaps**

    - Migration flows
    - Public opinion on migration
    - Migration policies
    - Migration and the environment   ⟶ **SDGs**
    - Well-being of migrants
    - Irregular migration
    - Migration potential/forecasting

65 YEARS
International Organization for Migration

Global Migration
Data Analysis Centre
GMDAC
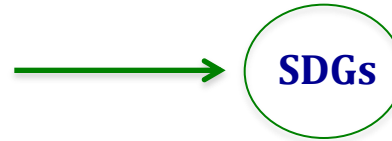
2. The potential

a) The global data revolution

90% of data that exists in the world today was created just over the past 2 years

TOTAL POPULATION	ACTIVE INTERNET USERS	ACTIVE SOCIAL MEDIA ACCOUNTS	UNIQUE MOBILE USERS	ACTIVE MOBILE SOCIAL ACCOUNTS
7.210 BILLION	3.010 BILLION	2.078 BILLION	3.649 BILLION	1.685 BILLION
URBANISATION: 53%	PENETRATION: 42%	PENETRATION: 29%	PENETRATION: 51%	PENETRATION: 23%
FIGURE REPRESENTS TOTAL GLOBAL POPULATION, INCLUDING CHILDREN	FIGURE INCLUDES ACCESS VIA FIXED AND MOBILE CONNECTIONS	FIGURE REPRESENTS ACTIVE USER ACCOUNTS, NOT UNIQUE USERS	FIGURE REPRESENTS UNIQUE MOBILE PHONE USERS	FIGURE REPRESENTS ACTIVE USER ACCOUNTS, NOT UNIQUE USERS

Source: *A World that Counts: Mobilising the Data Revolution for Sustainable Development* (2014)

b) (like) Big Data

What is big data?

*Big data represents the **Information** assets characterized by such a high **Volume**, **Velocity** and **Variety** to require specific **Technology** and **Analytical Methods** for its transformation into **Value*** (de Mauro et al., 2014)

From the 3Vs:
- **Volume**
- **Velocity**
- **Variety**

To the 3 Cs:
- **Communities**
- **Capacities**
- **Crumbs**

Letouzé (2016)

5 characteristics
- **Big data is produced anyway**
- **No random sampling: n = N**
- **Real-time accessibility**
- **Merging different sources ("data fusion")**
- **Full name: Big Data Analytics**

Hilbert (2014)

c) Alternative uses of big data in migration studies

Big data type	Examples of past uses	Further potential
Mobile phone Call Detail Records (CDRs)	Tracking post-disaster displacement (e.g. Haiti earthquake, 2010; Nepal earthquake, 2015; Cyclone Mahasen, Bangladesh, 2013) Estimating seasonal mobility (e.g. France and Portugal, 2014)	Early-warning systems, migration forecasting, internal and temporary migration patterns (e.g seasonal mobility)
Other types of mobile phone data (e.g. mobile money transfers)	Identifying modalities and determinants of mobile money transfers to share risk in post-disaster situation (e.g. Rwanda earthquake, 2008)	Patterns of remittance-sending activity; **remittance costs?**
IP addresses of website logins and sent e-mails	Estimating and predicting short and medium/term migration (e.g. from repeated logins to the same website during a certain period), and age- and gender-specific migration flows/rates (e.g. through sent e-mail)	Mobility tracking and forecasting, temporary and circular migration
Geo-located social media data **Online media content**	Inferring migration trends and comparing patterns of internal and international migration through Twitter activity, or migration of professionals through LinkedIn. Analysing migration-related public discourse (e.g. Oxford University Migration Observatory, 2011)	Mobility tracking and forecasting; internal and temporary/circular migration; early-warning for forced displacement; public attitudes; **migrants' perceptions?**
Google searches/Internet activity	E.g. Google search volume index for "Polski" used to estimate the number of foreign-born in the UK. More recently in asylum-seeker flows to Europe by sea (Pew Research Centre).	Migration potential/forecasting; (sudden changes in migratory patterns)

Ex.: Call Detail Records

FLOWMINDER.ORG

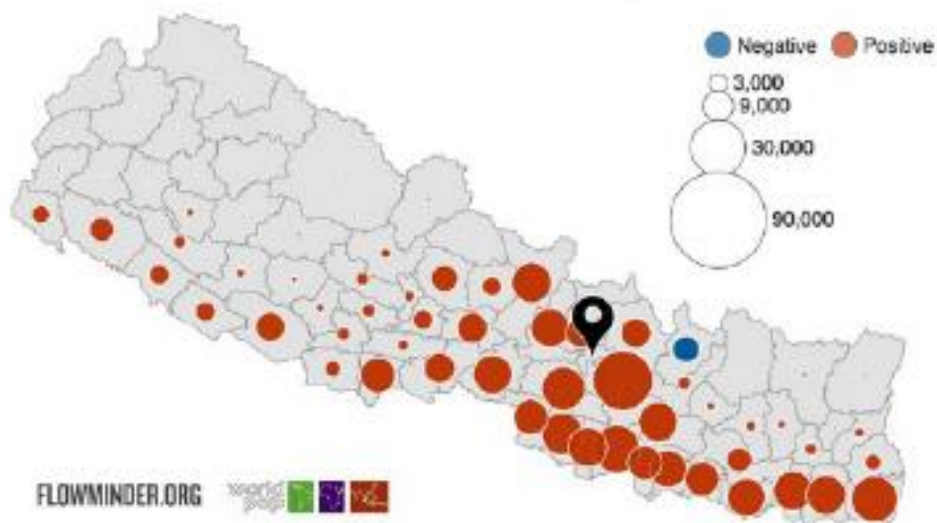**Nepal Population Estimates as of May 1, 2015**

2. Kathmandu Valley

Kathmandu Valley is here defined as the districts Kathmandu, Bhaktapur and Lalitpur. Kathmandu Valley is one of the most densely populated areas in Nepal and home to ca 2.8 m people [1].

Key findings:

→ An estimated 390,000 people more than normal had left the Kathmandu valley - comparing May 1 with the day before the earthquake April 24 (ratio to the population: 14%).

→ An estimated 247,000 persons less than normal had come into the area during the same period(ratio to the population: 8.8%)

→ People leaving Kathmandu Valley went to a large number of areas, notably the populous areas in the south and the Central and West Development Regions.

[1] www.worldpop.org

Ncell

Pre-earthquake population
2.8m

Population outflow (above normal)
+390,000
(246,000~540,000)

Population inflow (above normal)
−247,000
(−155,000−−339,000)

Above normal flows from Kathmandu Valley to other districts

Negative ● Positive
3,000 / 9,000 / 30,000 / 90,000

Above normal flows from Kathmandu Valley to other districts (comparing pre-earthquake April 24 and May 1).

Source: Bengtsson et al. (2015), Flowminder.org

IP addresses of sent e-mails
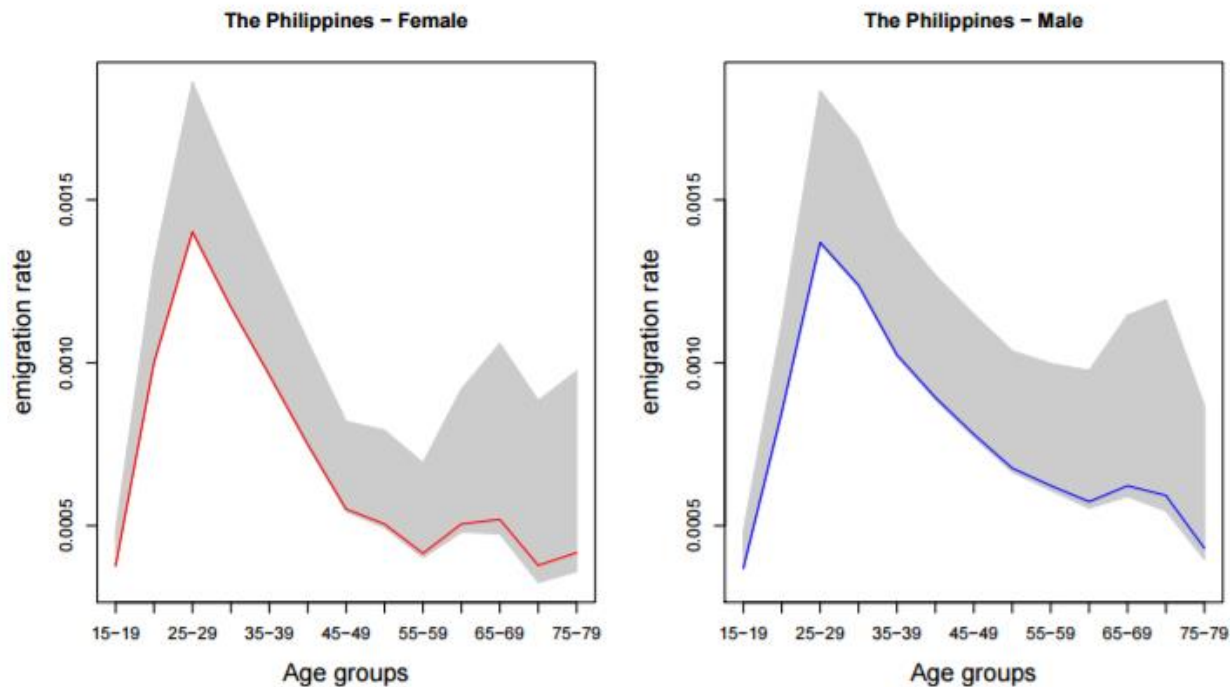

The Philippines − Female

The Philippines − Male

Figure 4. Estimates from e-mail data (2009-2011) of age and gender-specific emigration rates for the Philippines. The gray area represents the sensitivity of the estimates to the choice of the parameter for selection bias correction.

Source: Zagheni and Weber (2012)

IOM's Global Migration
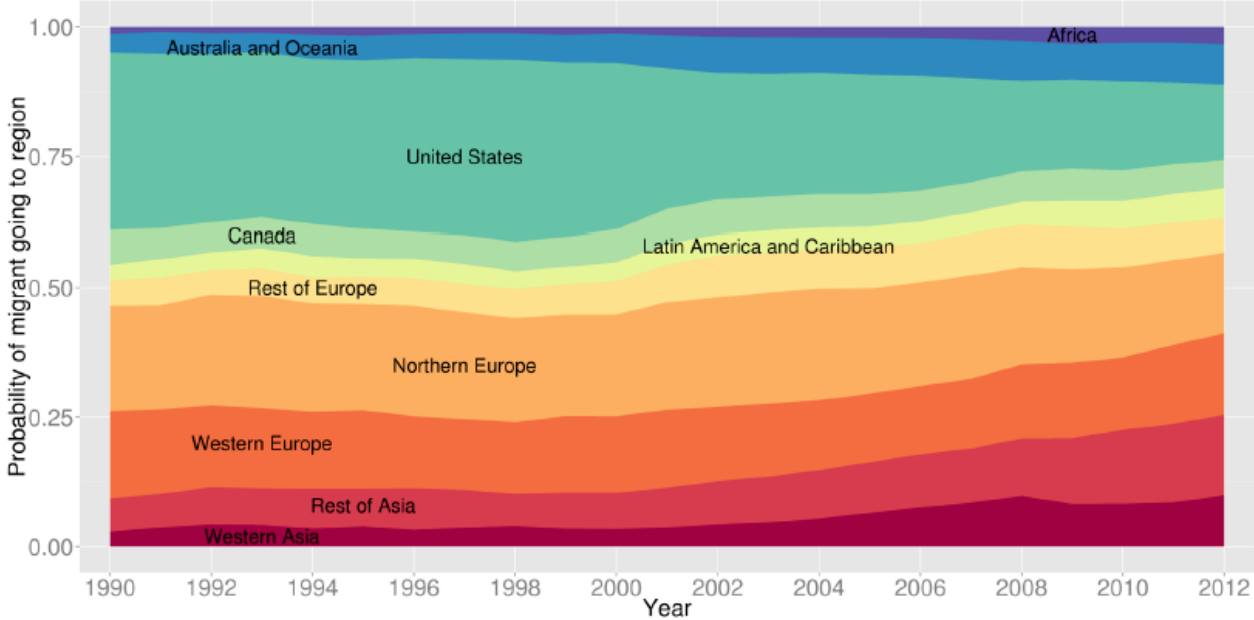Data Analysis Centre
GMDAC

Social media data: LinkedIn

Fig. 2. Distribution of Migration Flows, by year and region of destination, 1990-2012.

IOM's Global Migration
Data Analysis Centre
GMDAC

d) Big data and migration-related SDG targets (1/2)

- **10.c By 2030, reduce to less than 3% the transaction costs of migrant remittances and eliminate remittance corridors with costs higher than 5%**

    10.c.1 Remittance costs as a proportion of the amount remitted

    Example: **Blumenstock et. al. (2013)** (though only national)
   **But also** data from money transfer operators or mobile services?

- **16.2 End abuse, exploitation, trafficking and all forms of violence against and torture of children**

    16.2.2 Number of victims of human trafficking per 100,000 population, by sex, age and form of exploitation (Tier II)

    Example: **Polaris project**

Human trafficking: Polaris National Human Trafficking Hotline and BeFree textline

IOM's Global Migration
Data Analysis Centre
GMDAC
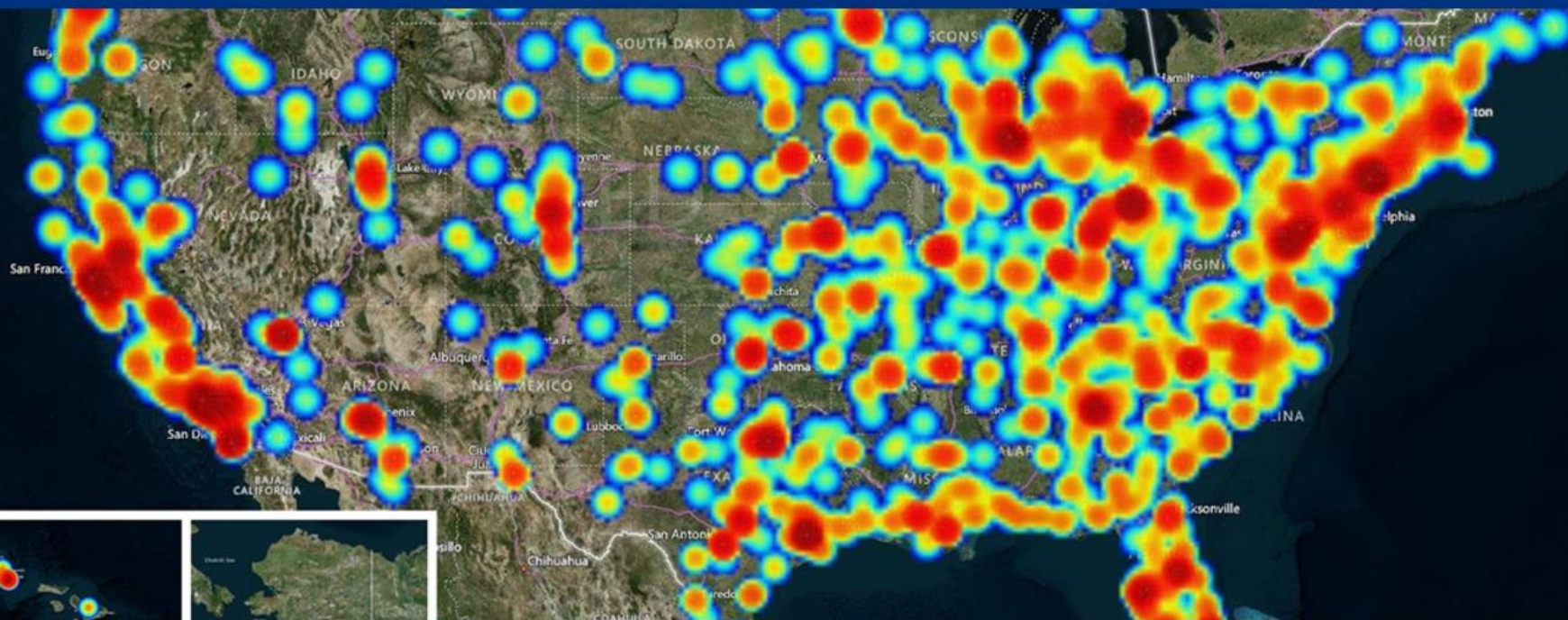
d) Big data and migration-related SDG targets (2/2)

- **10.3 Ensure equal opportunity and reduce inequalities of outcome, including by eliminating discriminatory laws, policies and practices and promoting appropriate legislation, policies and action in this regard**

    10.3.1 Proportion of population reporting having personally felt discriminated against or harassed in the previous 12 months on the basis of a ground of discrimination prohibited under international human rights law

    Example: **Global Pulse mobile-based survey of well-being**
    But also, social media data on migrant perceptions?

- Others: **irregular migration flows** (10.7)?

3. The challenges/opportunities

Big data and migration

Opportunities

- Wide coverage, real-time data
- Lower costs (?)
- Potential to understand:
  - Drivers of migration
  - Migration potential/forecasting
  - Attitudes towards migration
  - Migrants' perceptions
  - Mobility patterns
  - Spending/remittances patterns
- **SDG monitoring** (with other sources)

Challenges

- Privacy and ethical issues
- Civil liberty/security concerns (surveillance?)
- Regulatory/legislative frameworks
- Technical and analytical issues
- Data sharing: new infrastructure/security systems
- Sample bias
- Widening of the "digital divide"
- Mobility, not migration?
- Political/economic barriers

To conclude

- Exploring the **complementarity** between innovative and traditional sources

- How to make big data **"actionable"** (Letouzé, 2016) **for migration**:

    1) Regulatory and legislative framework
    2) Dialogue between regulators, users and providers
    3) Infrastructure and security systems
    4) Private-public partnerships for data exchange
    5) More migration research based on big data sources

- **On migration-related SDGs:**

    - Identification of trends, not hard data on progress

    - Big-data-based indicators?

Marzia Rango

Research Officer
IOM's Data Analysis Centre (GMDAC)
Taubenstr. 20-22
10117 Berlin, Germany

mrango@iom.int
gmdac@iom.int
Twitter: @MarziaRango