

Incorporating Big Data in the Revision of the 1998 UN Recommendations on Statistics of International Migration

UN Expert Group Meeting on Migration Statistics

United Nations

New York, 1—3 July 2019

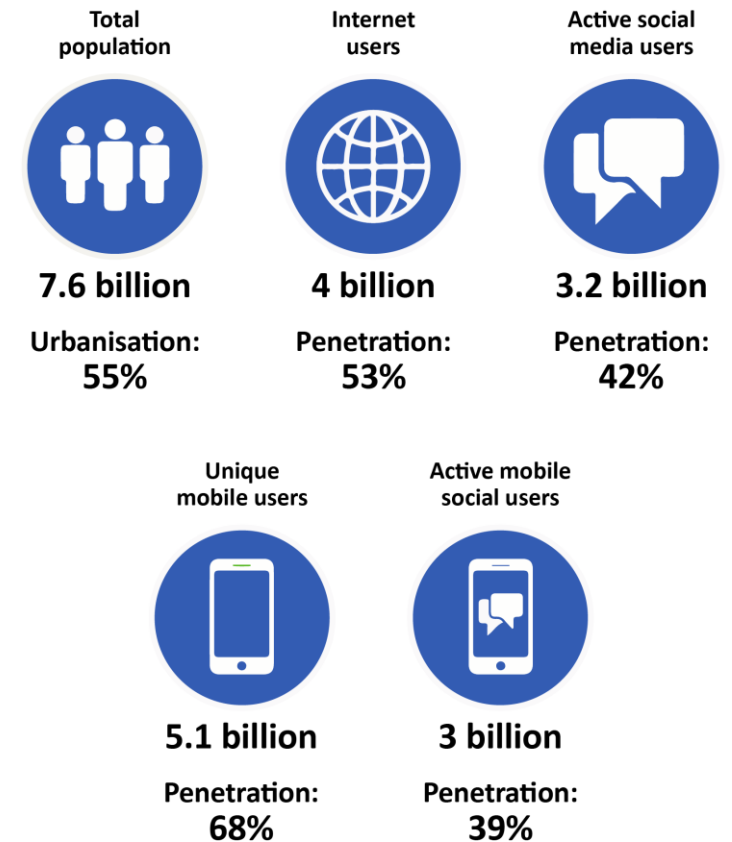


A data revolution for migration?

- Renewed calls (needs) for evidence on migration: the SDGs and the Global Compact for Safe, Orderly and Regular Migration (GCM)
- The (persistent) paucity of migration data from traditional sources globally
- Push for data innovation and calls for a 'data revolution' for sustainable development

Digital around the world in 2018

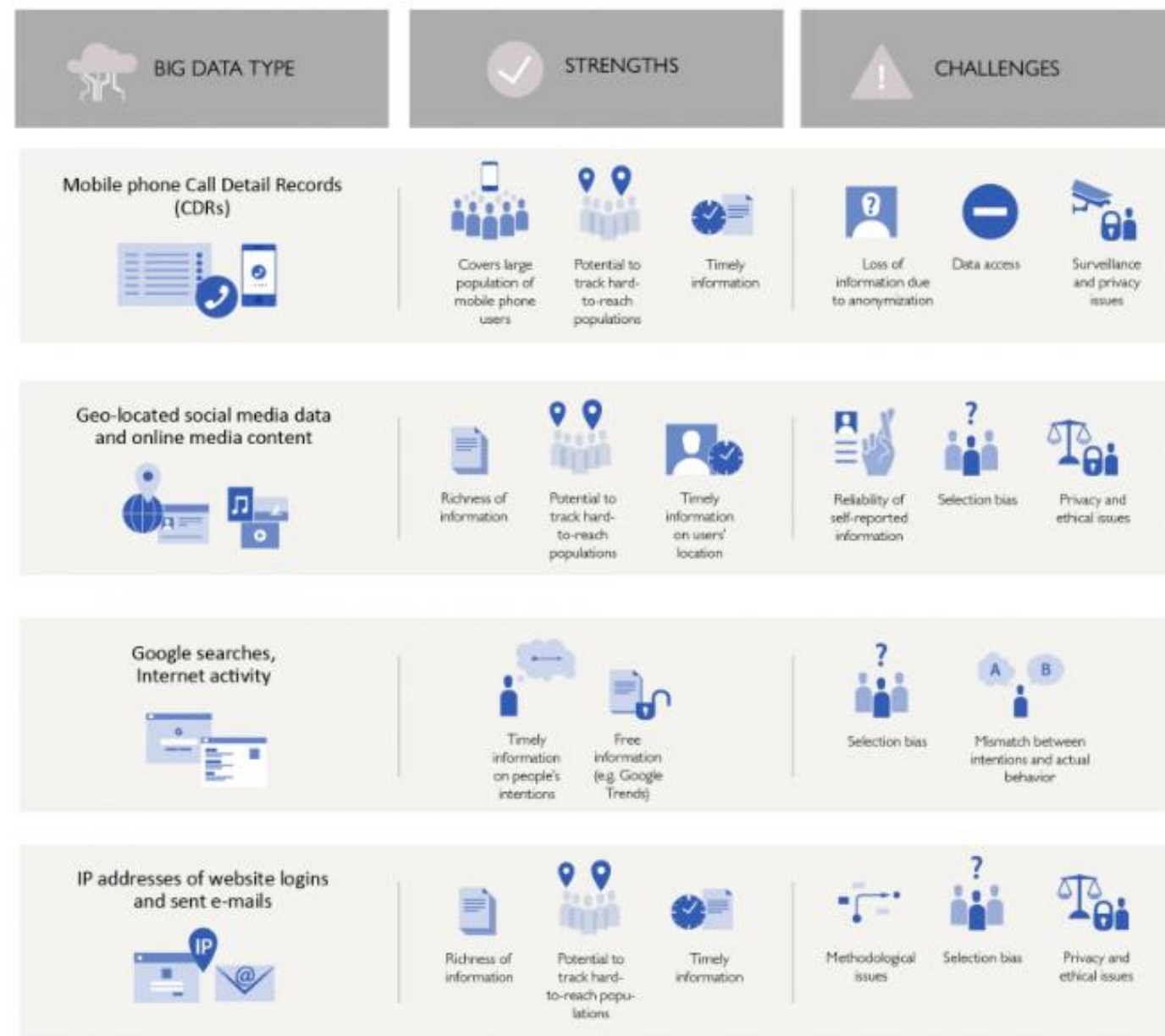
Key statistical indicators for the world's internet, mobile and social media users



Infographic based on Hootsuite and We Are Social, 2018. For sources and further information please see original at: <https://wearesocial.com/uk/blog/2018/01/global-digital-report-2018>

The potential

- High spatial resolution
- High frequency of update
- Timeliness (virtually real-time)
- Wide coverage (hard-to-reach populations)
- Larger sample size compared to surveys
- Richness of information
- Relatively low cost (depending on...)



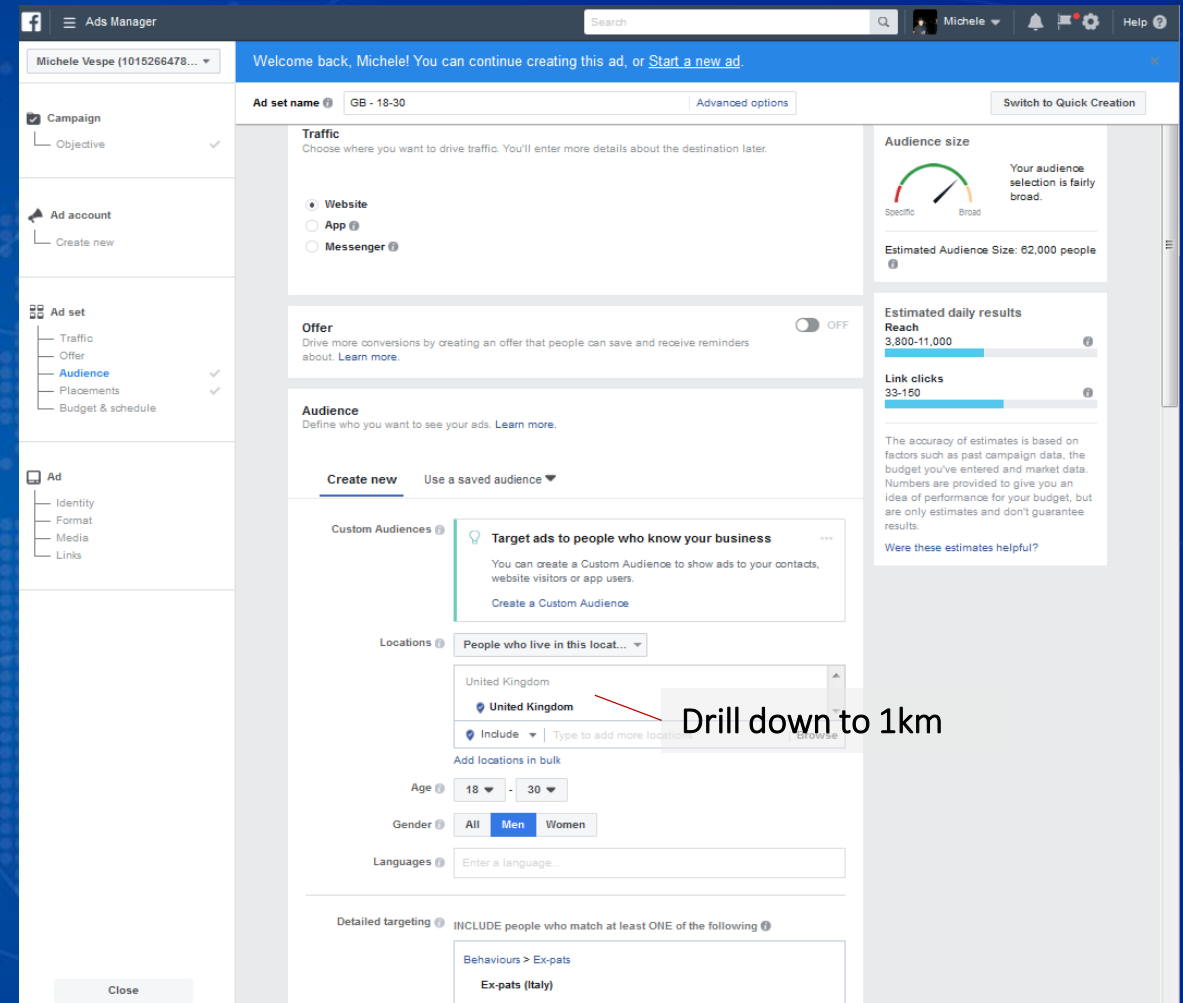
I: Social Media advertising platforms

Opportunities

- Real time census
- “inexpensive”
- Attributes like country of origin, education, age, sex (self-reported) and interests (likes)

Challenges

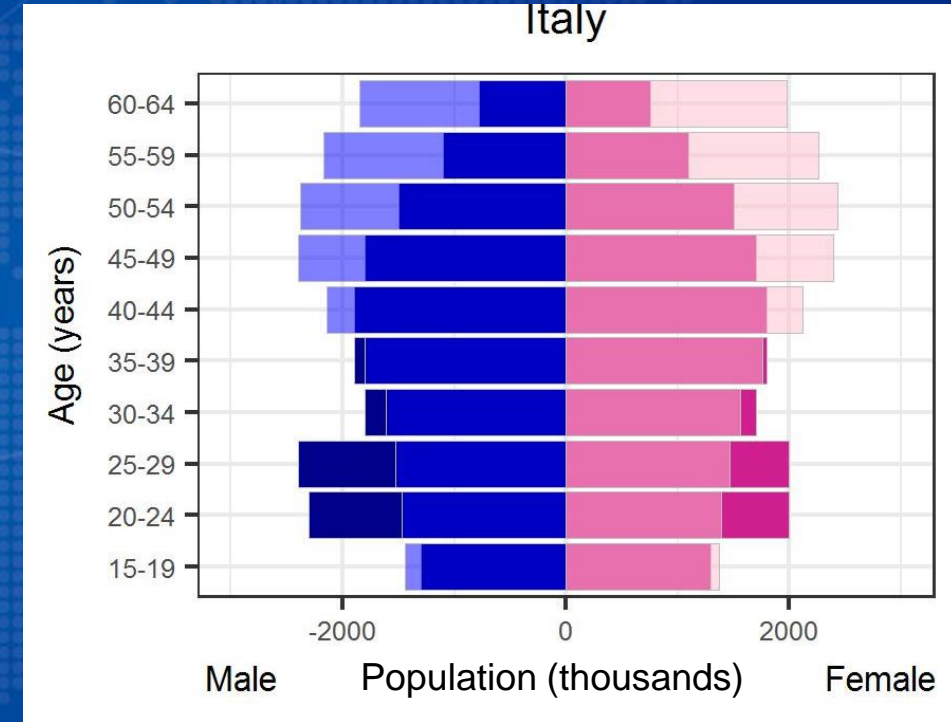
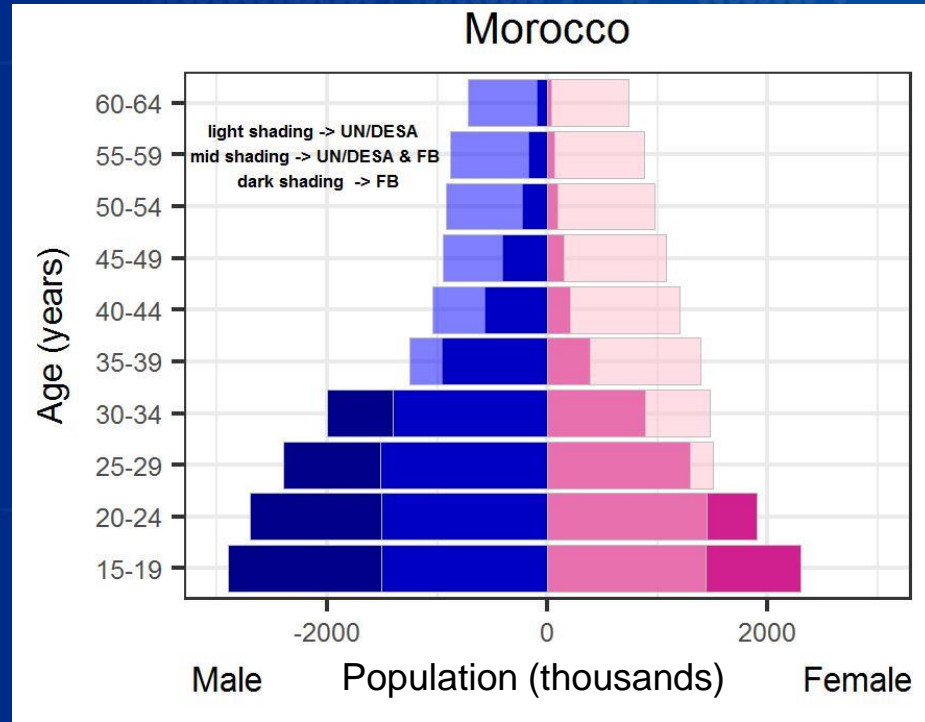
- Information reliability (often self-reported by user)
- Definitions (proprietary)
- Aggregations/rounding (proprietary)
- Penetration rate / selection bias (country, sex, age, education, sector, urban/rural...)
- Fake or double accounts
- Assimilation: expats destination or origin penetration rate?



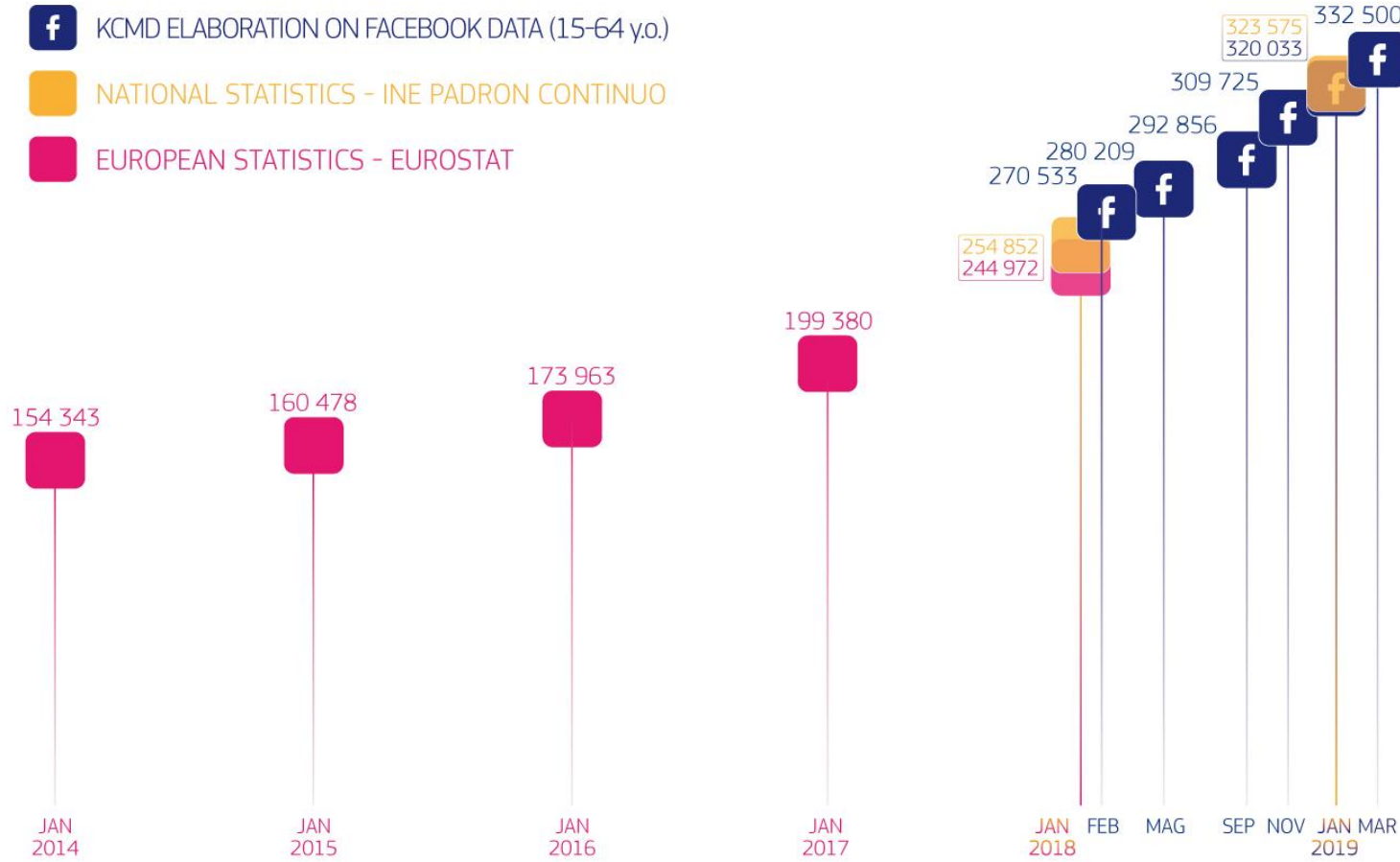
The screenshot displays the Facebook Ads Manager interface for a user named Michele Vespe. The main section is titled 'Ad set name' and shows the ad set 'GB - 18-30'. The 'Traffic' section is set to 'Website'. The 'Offer' section is currently turned off. The 'Audience' section is the primary focus, showing 'Custom Audiences' with a highlighted option: 'Target ads to people who know your business'. Below this, the 'Locations' section is set to 'People who live in this location...' and is currently set to 'United Kingdom'. A red arrow points to the 'United Kingdom' selection with the text 'Drill down to 1km'. The 'Age' range is set to 18-30, and 'Gender' is set to 'All'. The 'Languages' section is empty. The 'Detailed targeting' section is set to 'INCLUDE people who match at least ONE of the following' and includes 'Behaviours > Ex-pats' and 'Ex-pats (Italy)'. On the right side, there are several summary cards: 'Audience size' showing an estimated size of 62,000 people, 'Estimated daily results' showing a reach of 3,800-11,000 and link clicks of 33-160, and a note about the accuracy of estimates.

Understanding the bias

Correcting the bias: a question of penetration rate (popularity and internet access) but also gender gap



'Migrant' stocks from social media

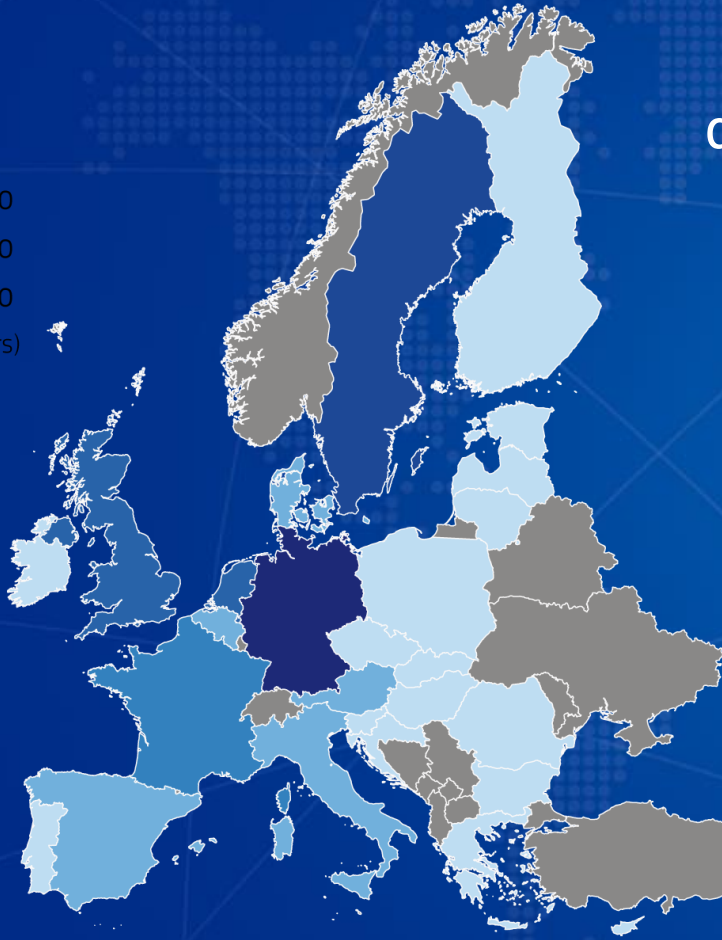


Venezuelans in Spain

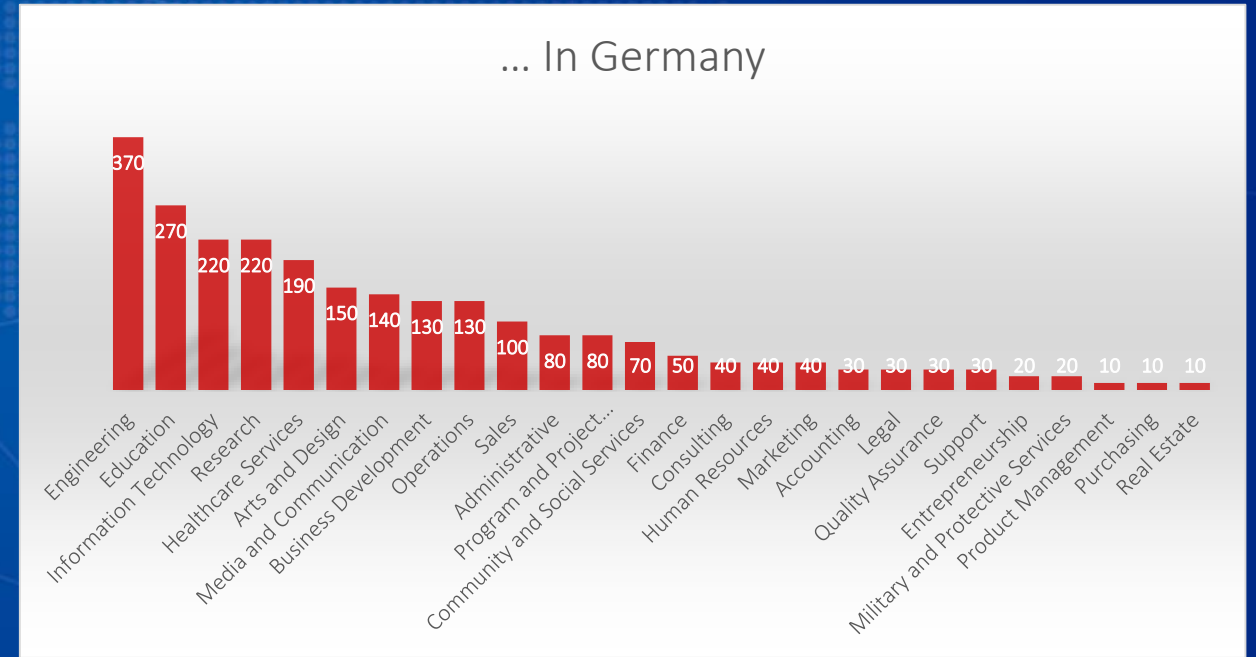
Facebook Advertising Platform data:
high refresh rate & real time

Ref: "Migration Data using Social Media" Spyratos, S., M. Vespe, F. Natale, I. Weber, E. Zagheni and M. Rango, doi:10.2760/964282, 2018

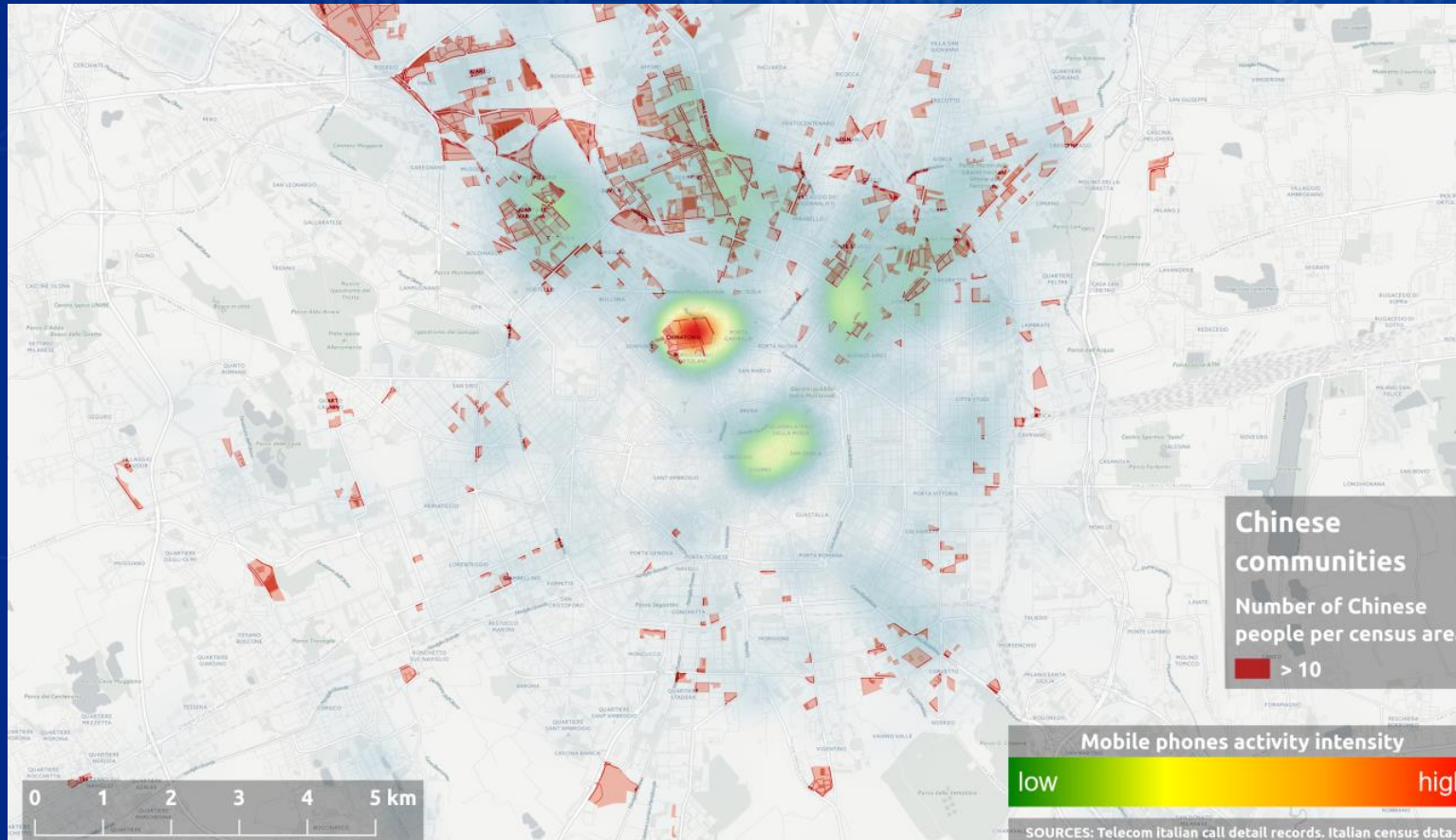
Migrant characteristics from social media



Syrian users of LinkedIn in Europe:
distribution (left) and sector of occupation (below)



Urban distribution from mobile phone data



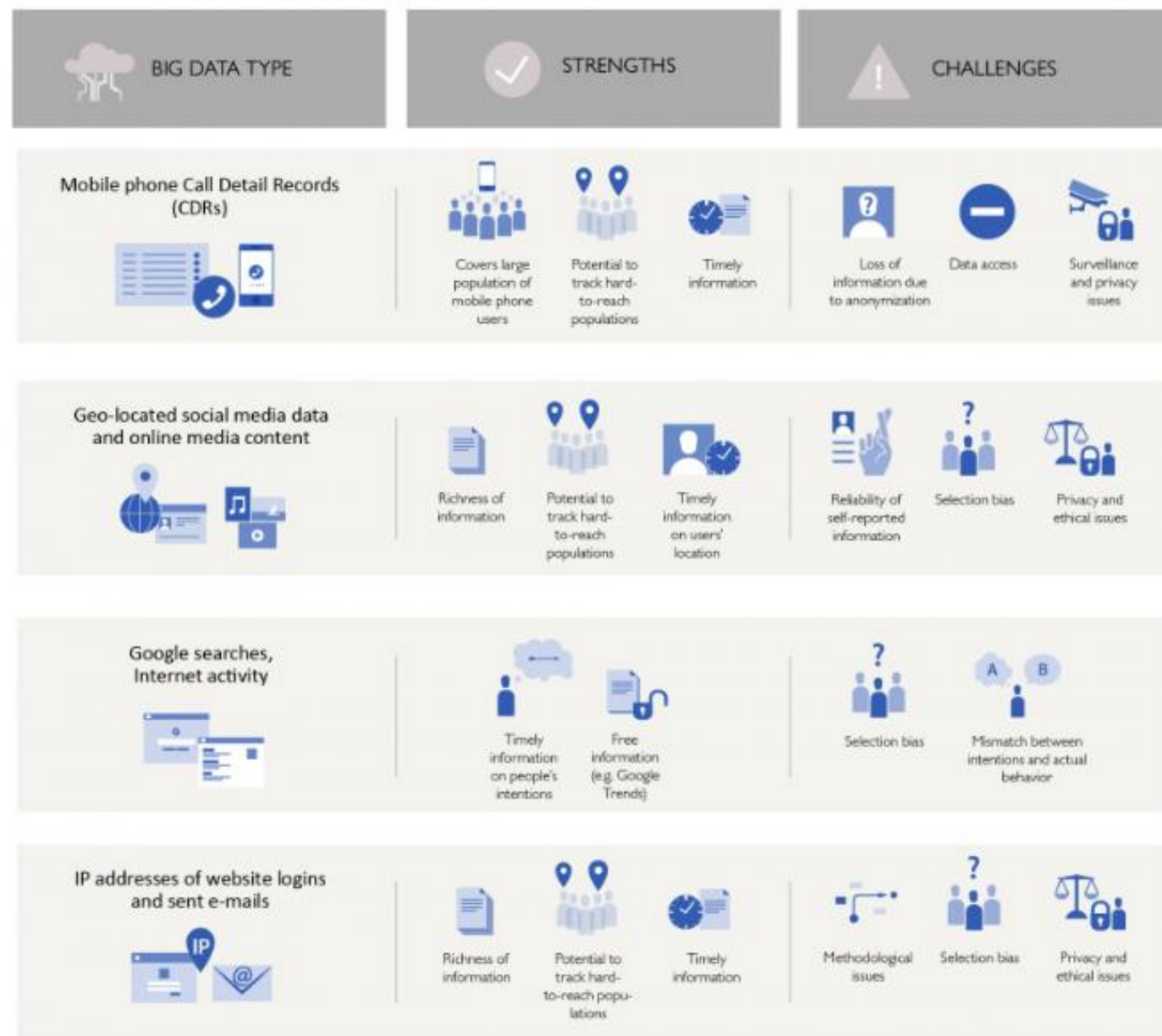
Density of mobile phone traffic with China and areas of highest concentration of Chinese immigrants (2011 Census)

Sources:

- Italian Census Data &
- Call Detail Records, Telecom Italia

Challenges

- Data access/Continuity
- Confidentiality
- Security/Ethical issues
- Methodology (selection bias)
- Reliability
- Definitions
- Fragmentation



Harnessing data innovation for migration: The Big Data for Migration Alliance (BD4M)

- Launched on 25 June 2018 in Brussels
- Convened by the EC KCMD and IOM's GMDAC
- Aim: Foster data innovation for migration analysis & policymaking through *partnerships*



Launch of the Big Data for Migration Alliance, Joint Research Centre, Brussels, 25 June 2018
© European Commission 2018

Big Data for Migration Alliance (BD4M)

Main areas of work:

a) Awareness-raising and knowledge-sharing

- The ‘10 questions’ that matter on migration
 - with NYU GovLab
- Repository of big data & migration projects
- Creation of a network of ‘data stewards’

b) Capacity-building

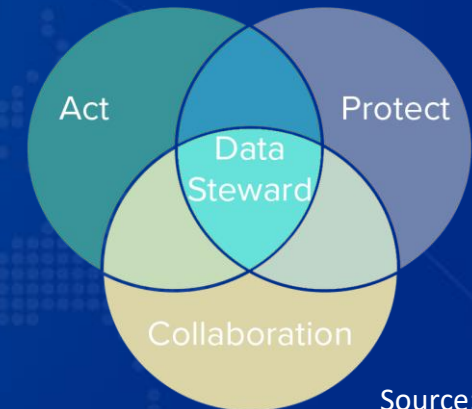
- Preparation of guidance and training materials
- ‘Study visits’ and exchange of experiences across countries

c) Policy-oriented analysis

- Exploring EU/global frameworks to sponsor applied research (e.g. issuing data challenges)
- Working paper series



Launch of the Big Data for Migration Alliance, Joint Research Centre, Brussels, 25 June 2018. © European Commission 2018



Source: GovLab

The “100 Questions” initiative

- Identifying the most important societal questions whose answers can be found in data and data science
 - 10 key questions on migration – in partnership with GMDAC and KCMD, within the BD4M framework
- Building a community of ‘bilinguals’
- Creating new data collaboratives

“Organizations are far more likely to share data in response to a compelling use case. By identifying important questions that data science and machine learning can help us answer, we can spark collaborations between organizations that have critical data sets, teams that can derive insights from the data, and individuals and organizations that can make better decisions based on those insights.”

- Tom Kalil

Chief Innovation Officer, Schmidt Futures

How to incorporate Big Data in the Revision of the 1998 Recommendations on Statistics of International Migration?



Discussion

- Any other examples of good practice that you would like to mention? What are the issues encountered when exploring big data sources and how were they overcome?
- What are the policy needs? What are the key questions that are currently hard to answer in a timely fashion using traditional data sources? Can new data sources help?
- Definitions: is it possible to harmonize definitions across traditional and new data sources or should we not even attempt it?
- Should the Revision include guidance on how to explore big data sources and what would this look like (e.g. elements to consider? Examples? How to address methodological and other challenges?)

Contact

Marzia Rango

IOM's Global Migration Data
Analysis Centre (GMDAC)

mrango@iom.int
gmdac.iom.int
migrationdataportal.iom.int

Michele Vespe

European Commission's
Knowledge Centre on Migration and
Demography (KCMD)

michele.vespe@ec.europa.eu
<http://ec.europa.eu/jrc/en/migration-and-demography>

Unlocking the potential of private data for public good: 'Data collaboratives'



The screenshot shows a Medium article page. At the top left is the Medium logo. The author's name 'Stefaan G. Verhulst' is displayed with a profile picture and a 'Follow' button. The article title is 'Data Collaboratives: Exchanging Data to Improve People's Lives'. Below the title is the byline 'By Stefaan Verhulst and David Sangokoya, The GovLab'. The main text begins with 'In late July 2014, a sick passenger from Liberia traveled to Nigeria and brought the Ebola virus to Lagos, Africa's largest city, with a population of 21 million. In response, government agencies, universities and hospitals collaborated with private telecommunications companies and healthcare organizations to collect and share data on infected patients and trace those who had come into contact with them. State government health officials also initiated emergency steps to share information on a daily basis among actors involved in stemming the crisis. After two months, the virus was contained in Nigeria and the country declared Ebola-free.' The text continues to discuss the broader implications of data for public health and society. At the bottom of the article is a word cloud with 'data' as the largest word, and other words like 'private value', 'help', 'Data', 'analysis', 'tasks', 'help', 'data', 'value', 'private', 'value', 'data', 'value', 'private', 'value'.

Stefaan G. Verhulst
Apr 22, 2015 · 8 min read

Data Collaboratives: Exchanging Data to Improve People's Lives

By Stefaan Verhulst and David Sangokoya, The GovLab

In late July 2014, a sick passenger from Liberia traveled to Nigeria and brought the Ebola virus to Lagos, Africa's largest city, with a population of 21 million. In response, government agencies, universities and hospitals collaborated with private telecommunications companies and healthcare organizations to collect and share data on infected patients and trace those who had come into contact with them. State government health officials also initiated emergency steps to share information on a daily basis among actors involved in stemming the crisis. After two months, the virus was contained in Nigeria and the country declared Ebola-free.

Several of society's greatest challenges—from addressing climate change to public health to job creation—require greater access to data, more collaboration between public- and private-sector entities, and an increased ability to analyze datasets. This relationship between data and public benefits was vividly demonstrated in case study after case study at the recently concluded [Cartagena Data Festival](#).

Yet for all the potential, a limiting factor is that much of the data valuable for solving public problems actually resides within the private sector—for example, in the form of click histories, online purchases, sensor data, and, as in the case of the above example, call data records. Amid the proliferation of apps, platforms and sensors, data on how people and societies behave is increasingly privately owned. We believe that if we truly want to leverage the potential of data to improve people's lives, then we need to accelerate the creation and use of "data collaboratives."

The term *data collaborative* refers to a new form of collaboration, beyond the public-private partnership model, in which participants from different sectors—including private companies, research institutions, and government agencies—can exchange data to help solve public problems. In the coming months and years, data collaboratives will be essential vehicles for harnessing the vast stores of privately held data toward the public good.

SHARE
♡
👍
🔖
@

data
private value
Data
analysis
tasks
help
data
value
private
value
data
value
private
value

"The term data collaborative refers to a new form of collaboration, beyond the public-private partnership model, in which participants from different sectors—including private companies, research institutions, and government agencies—can exchange data to help solve public problems.

In the coming months and years, data collaboratives will be essential vehicles for harnessing the vast stores of privately held data toward the public good."

Stefaan G. Verhulst - April 22, 2015 - Medium

medium.com/@sverhulst/data-collaboratives-exchanging-data-to-improve-people-s-lives-d0fcfc1bdd9a#.f1ib5frfh



Six types of data collaboratives

(Source: GovLab)



Problem: Lack of data diversity, leading to unrepresentative interventions.

Data Cooperatives or Pooling

Corporations and other important dataholders group together to create “data pools” with shared data resources.

E.g. Counter-Trafficking Data Collaborative



Problem: Lack of external actors to apply data analysis skills within public sector.

Prizes & Challenges

Corporations make data available to qualified applicants who compete to develop new apps or discover innovative uses for the data.

E.g. Telecom Italia Big Data Challenge



Problem: Limited information and data for academic researchers, stymying their progress.

Research Partnerships

Corporations share data with universities and other academic organizations giving researchers access to consumer datasets and other sources of data to analyze social trends.

E.g. Flowminder/WorldPop in Namibia

E.g. Flowminder/NCell



Problem: Inability or lack of resources to create data-driven products to solve a public problem.

Intelligence Products

Shared (often aggregated) corporate data is used to build a tool, dashboard, report, app or another technical device to support a public or humanitarian objective.

E.g. Facebook API data



Problem: Inability to access useful data continuously from particular companies, like social networks.

Application Programming Interfaces (APIs)

APIs allow developers and others to access data for testing, product development, and data analytics.

E.g. The HDX



Problem: Lack of expertise to analyze or use private sector data, even when given access.

Trusted Intermediary

Corporations share data with a limited number of known partners. Companies generally share data with these entities for data analysis and modelling, as well as other value chain activities.