

The European Commission's science and knowledge service

Joint Research Centre



Big Data for Migration: promises and challenges of non-traditional data for migration statistics

UNECE-Eurostat Work Session on Migration Statistics
24 - 26 October 2018
Geneva

Michele Vespe

European Commission, Joint Research Centre, Knowledge Centre for Migration and Demography (KCMD)

Marzia Rango

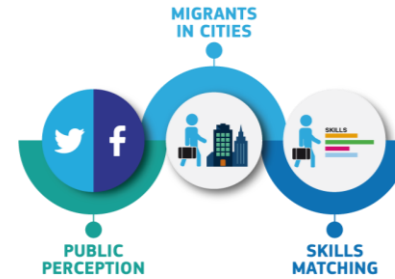
International Organization for Migration, Global Migration Data Analysis Centre

Knowledge Centre on Migration and Demography

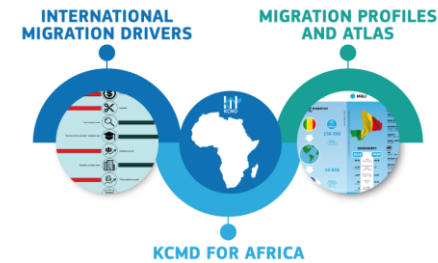


**KNOWLEDGE CENTRE
ON MIGRATION
AND DEMOGRAPHY**

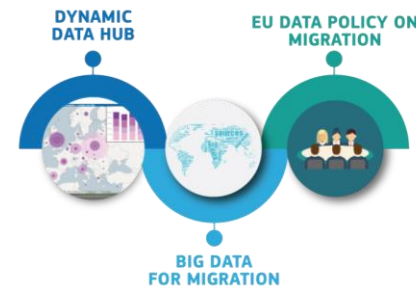
INTERNAL DIMENSION



EXTERNAL DIMENSION



DATA FOR MIGRATION

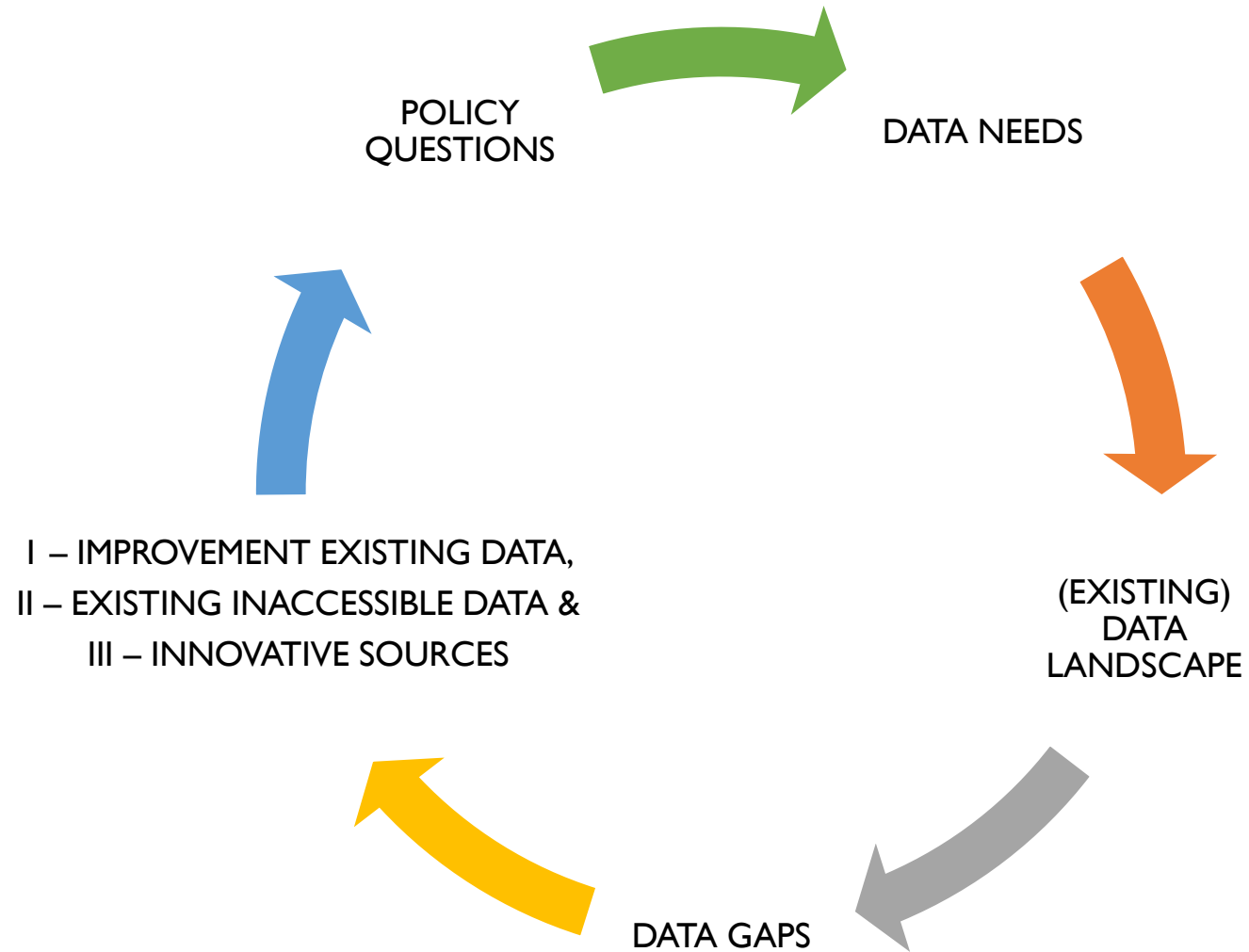


Why (better) migration data?

To (better):

- balance the public debate on migration;
- understand migration;
- evidence-inform migration policies

Improving Migration Data



Landscapes of existing data

- Population Census
- Surveys
- Population registers
- Administrative data
- Operational data
- Alternative data /big data

Reliability, frequency, attributes, sample size/frame, granularity, timeliness, costs...

KCMD Initiatives

Pre-processed data access & visual analytics



To help policy makers, researchers and stakeholders to *analyse* data

<https://bluehub.jrc.ec.europa.eu/migration/app/index.html>

~20 datasets

Discovered, filtered & quality checked catalogue of data sources

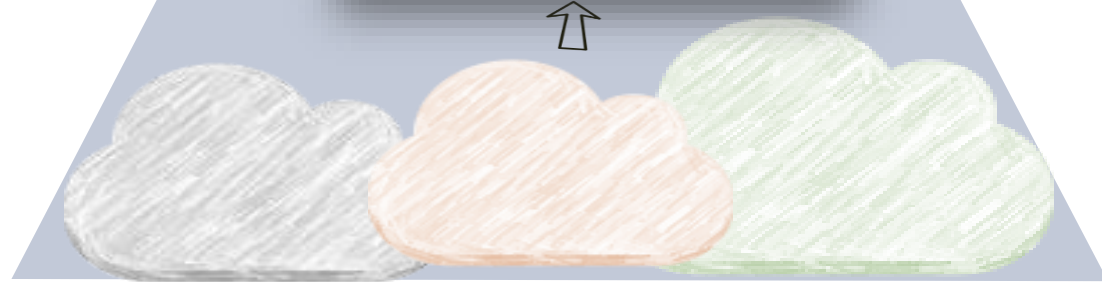


To help policy makers, researchers and stakeholders to *discover* and *use* data

<https://bluehub.jrc.ec.europa.eu/catalogues/data>

~140 datasets

Fragmented & scattered data



Big Data for Migration (BD4M) Alliance

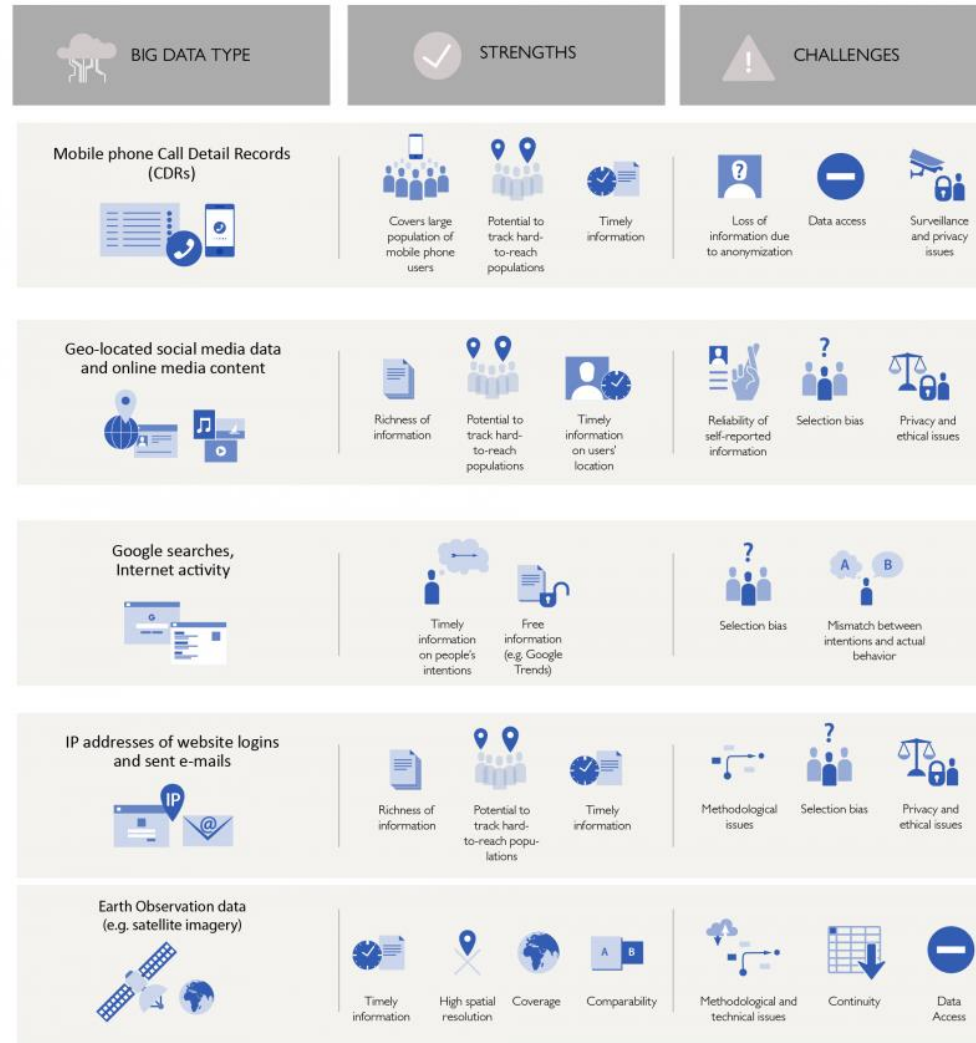


The European Commission and the International Organization for Migration launch the **Big Data for Migration Alliance (BD4M)**, a global initiative to unlock the potential of big data sources and provide valuable insights related to migration



The potential of big data and innovate data sources

- Data access
- Confidentiality
- Security/ethical issues
- Methodology (bias)
- Fragmentation
- Definitions



Big data and alternative data sources on migration: from case studies to policy support

Most promising application:

Social media advertising platforms

Opportunities

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

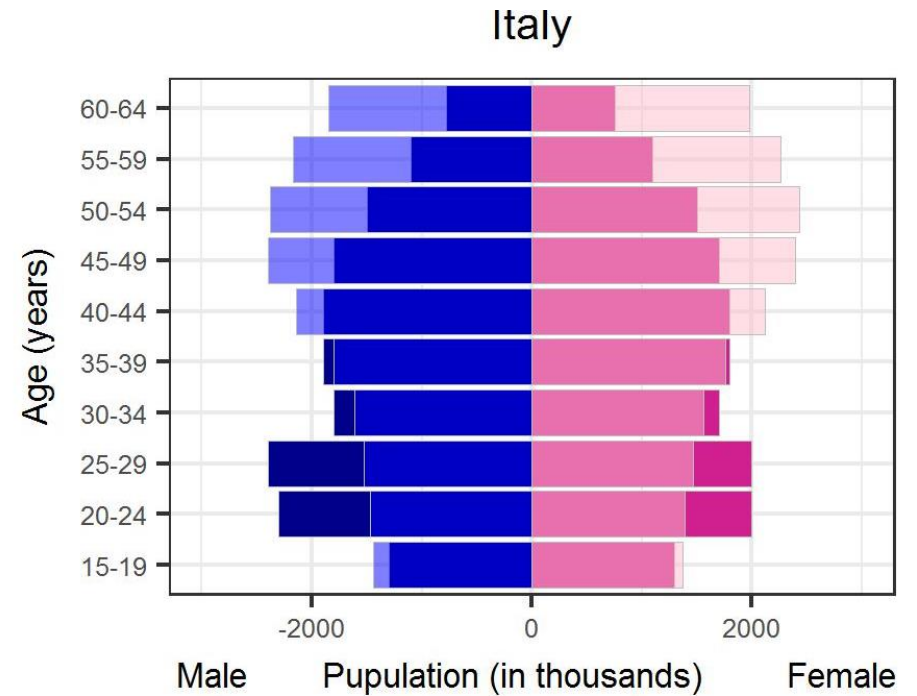
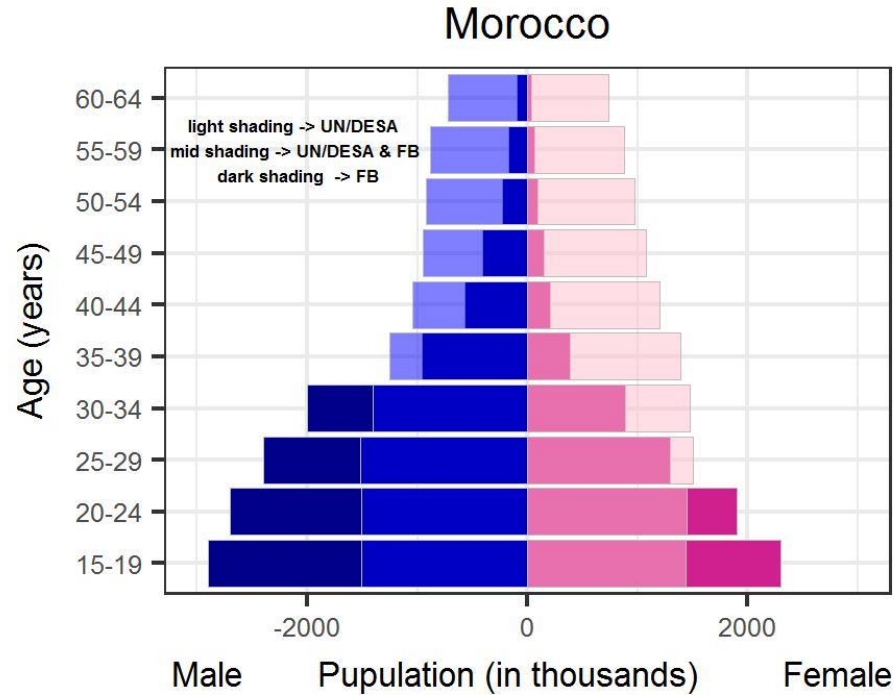
Challenges

- 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 shows the Facebook Ads Manager interface for a user named Michele Vespe. The main content area is titled "Audience" and shows targeting options. A text box with a red border and a red arrow pointing to the "United Kingdom" location selection says "FB controls over rounding (used to be 20, now 1000)". Another text box with a red border and a red arrow pointing to the "United Kingdom" location selection says "Drill down to 1km". The "Audience size" section shows a gauge chart and the text "Your audience selection is fairly broad." and "Estimated Audience Size: 62,000 people". The "Estimated daily results" section shows "Reach: 3,800-11,000" and "Link clicks: 33-150". The "Detailed targeting" section shows "Behaviours > Ex-pats" and "Ex-pats (Italy)".

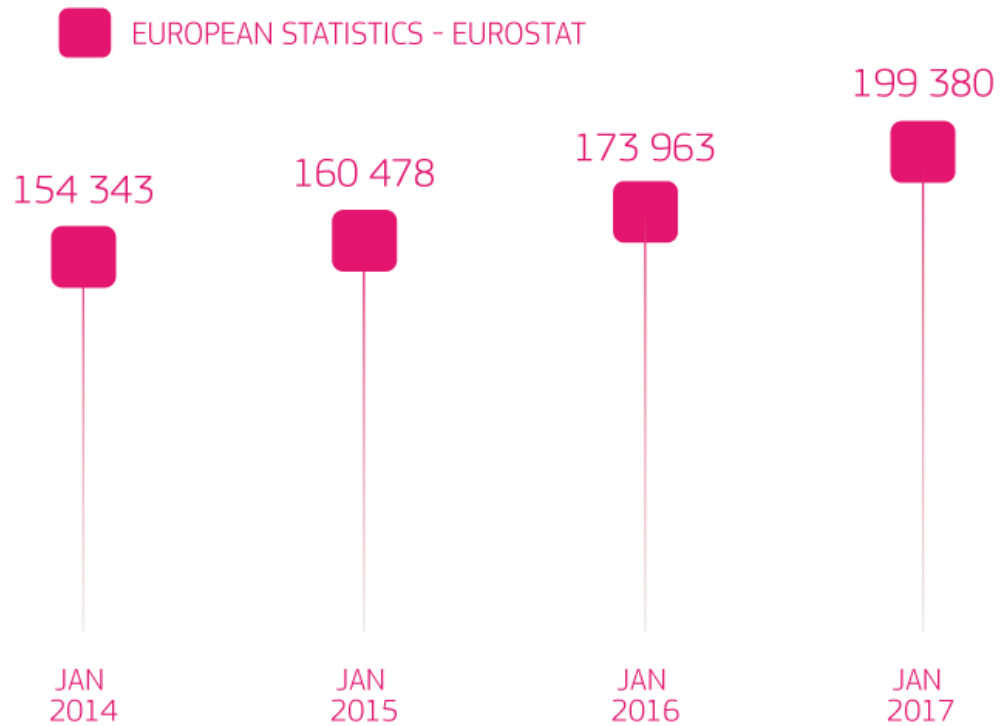
Understanding the bias

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

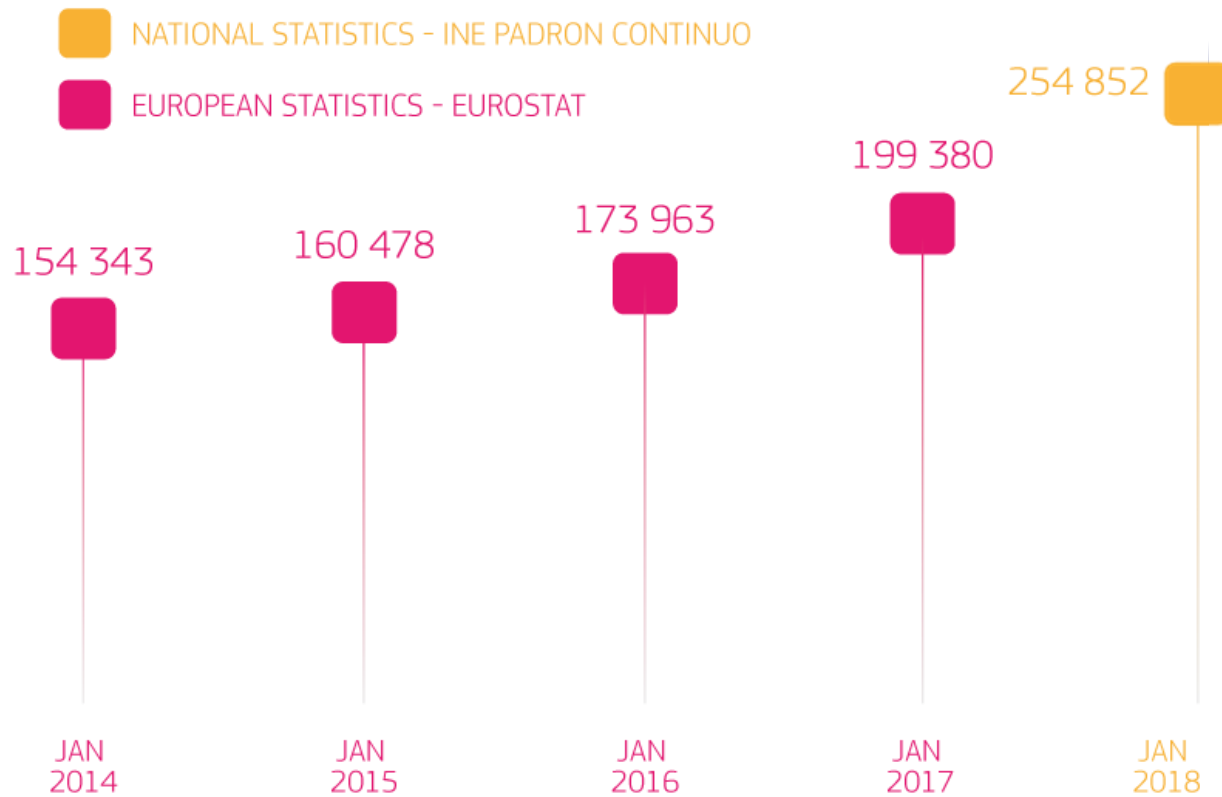


Social Media Potential - Facebook

Latest EU official statistics: 1-Jan-17,

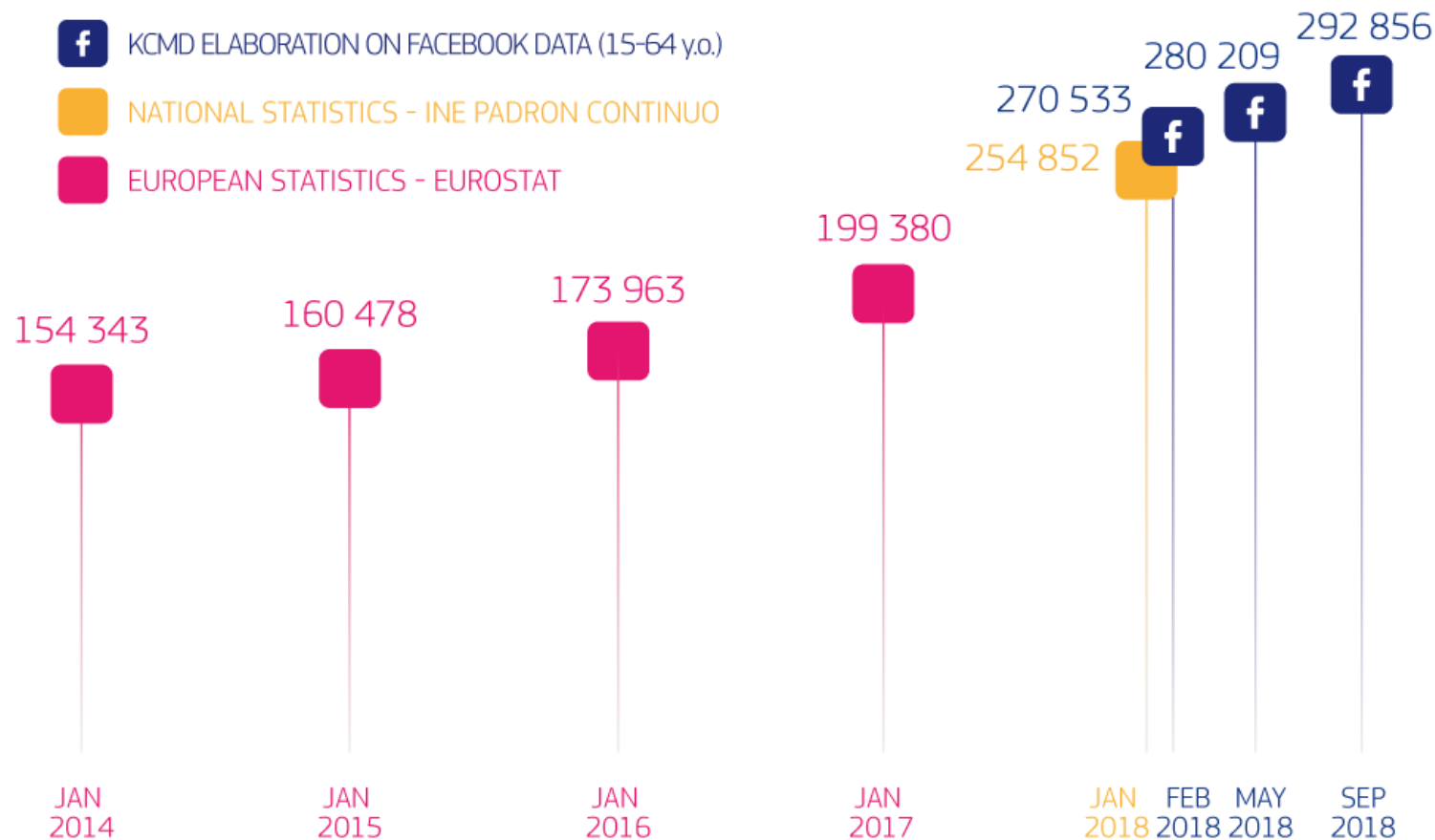


Social Media Potential - Facebook



Latest EU official statistics: 1-Jan-17,
Spanish Statistical Office: 1-Jan-18

Social Media Potential - Facebook









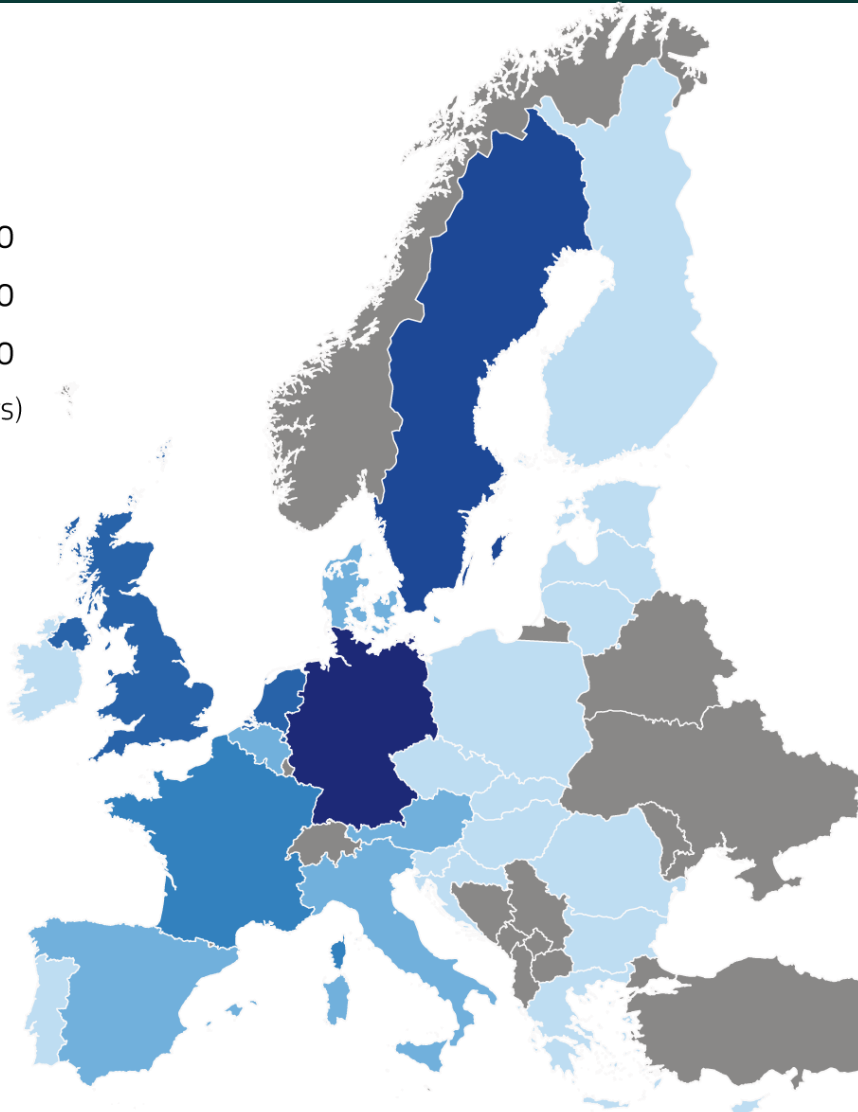
Latest EU official statistics: 1-Jan-17,
Spanish Statistical Office: 1-Jan-18

Facebook Advertising Platform data:
high refresh rate – real time census

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

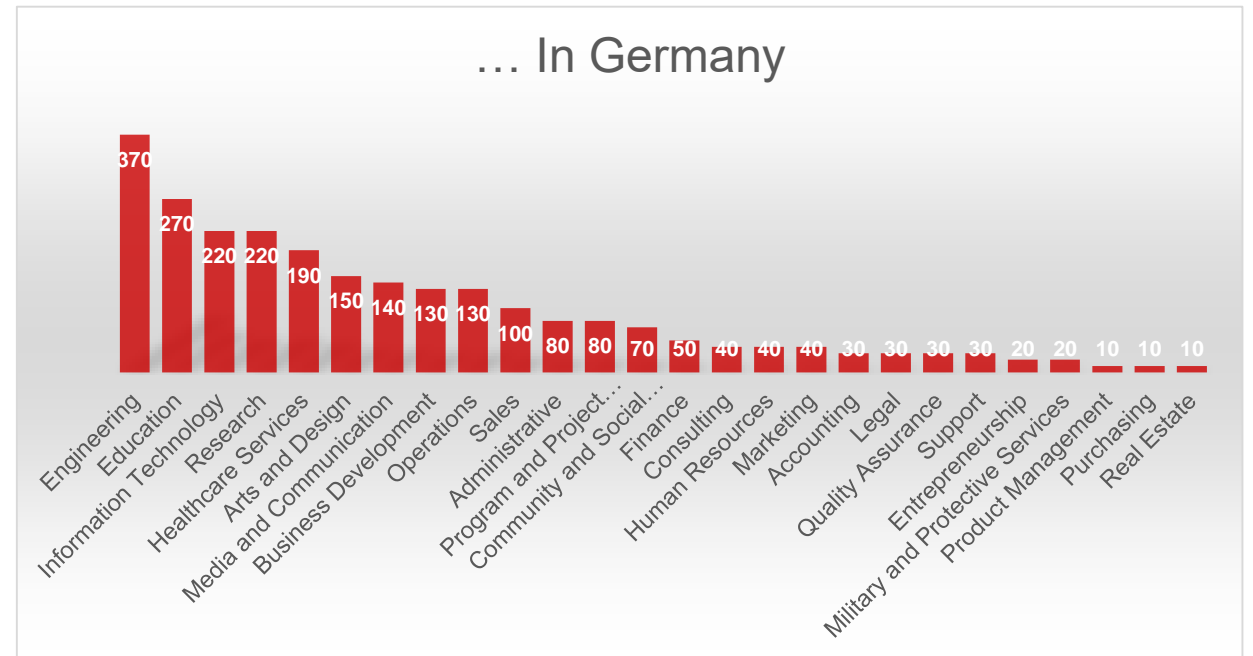
Migration data innovation potential II: Characteristics

-  < 200
 -  200 - 500
 -  500 - 1000
 -  1000 - 2000
 -  2000 - 3000
 -  3000 - 4000
- (Absolute numbers)

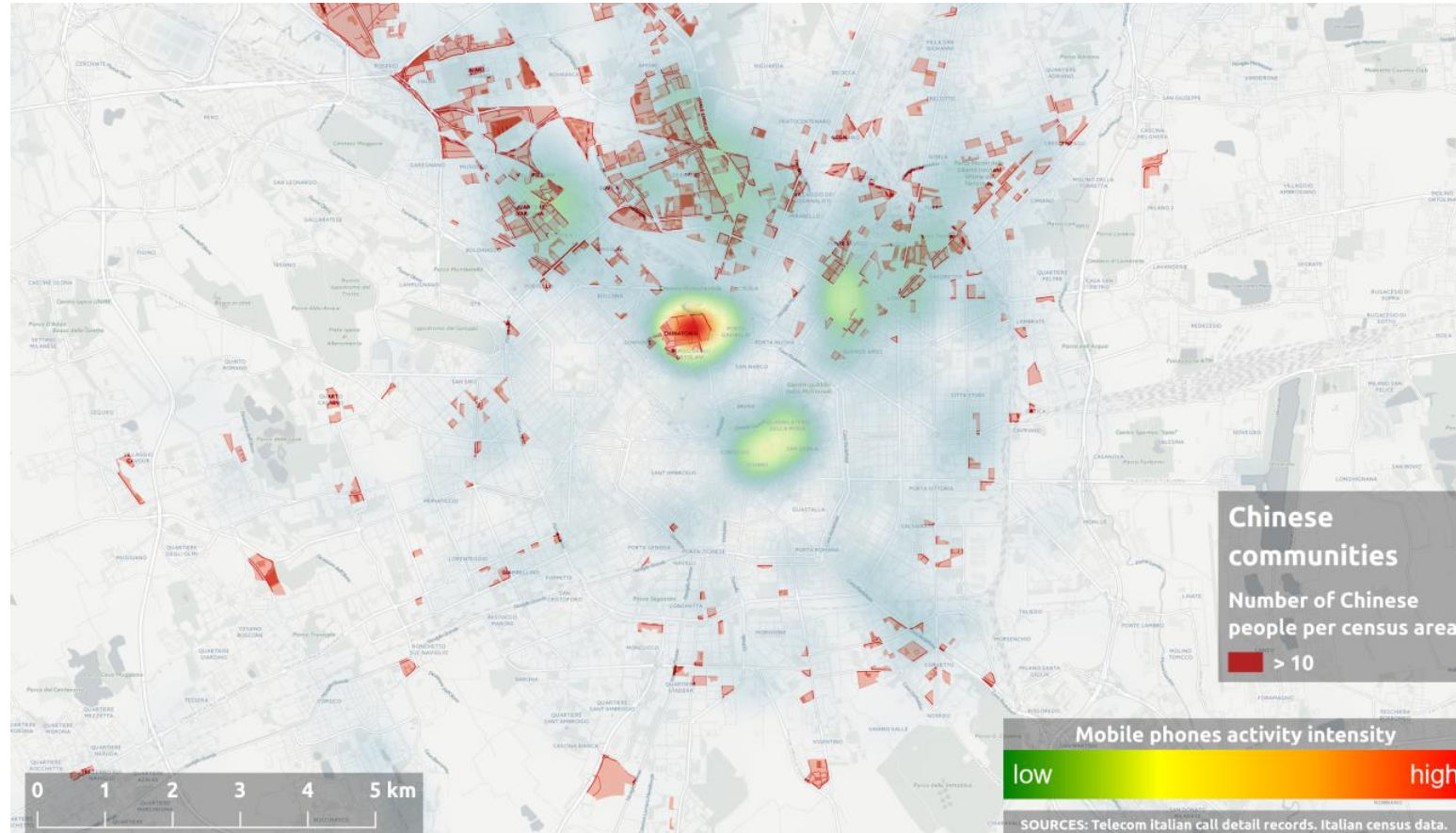


Example of highly-skilled migrants using LinkedIn advertising platform

... In Germany



Migration data innovation potential IV: Granularity (space)



Density of mobile phone traffic with China and areas of highest concentration of migrants from 2011 Census

Sources:

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

Conclusions

- Data can help balancing the public debate around an often misperceived theme such as migration
- Policy advice and knowledge management for migration require full awareness of data opportunities and limitations
- Data timeliness, disaggregation by time, space and attributes can be addressed through research and alternative methods



Any questions?

You can find us at [@M_Vespe](#) & michele.vespe@ec.europa.eu
[@MarziaRango](#) & MRango@iom.int