The European Commission’s science and knowledge service

Joint Research Centre

Big data and migration
Buzz-phrase or policy relevant applications?
Summary (3 examples from JRC work)

- Local perspectives on integration (Census, Human Settlement Layer and CDR)
- Migration vs. mobility (Air passengers data)*
- Public perceptions on migration (Twitter)*

* work in progress please don’t quote
Local perspectives on integration
(research questions)

- How many migrants at local level?
- Comparative analyses of concentration, diversity, segregation across cities, nationalities
- Drivers of segregation
- Relations with education, income segregation, electoral outcomes
Local perspectives on integration (data)

Census data

+ Human Settlement Layer & land cover ->

Concentration
Diversity
Residential segregation

+ CDR -> Activity based segregation

Concentration of migrants of Chinese origin in Florence (ITA) (2011)

Concentration of migrants of Chinese origin in Milan (ITA) (2011) and density of mobile phone activity towards China. Own elaboration on the basis of Census (ISTAT) and Telecom Italia data

Geometries of Census data in 8 EU MS
Local perspectives on integration (further research)

- School demographics -> school segregation
- House prices/micro census -> income segregation
- Electoral data -> migration and politics
- Register based census -> segregation in time

https://bluehub.jrc.ec.europa.eu/datachallenge/
Migration vs mobility
(research questions)

• Data on migration flows within Africa
• Capture more ‘fluid’ forms of migration (circular migration, temporary migration) not included in the one year definition of official statistics
Migration vs mobility (data)

Outbound flows from Dakar (SEN) to Italian airports (2016). Own elaboration of SABRE data.

Outbound flows from Dakar (SEN) to African airports (2016). Own elaboration of SABRE data.
Migration vs mobility (preliminary results)

Relation between emigration rate and intensities and air passengers (2015). Own elaboration on the basis UNDESA, Abel 2016 and SABRE data.

Outbound (green) and inbound (brown) passengers flows between Italy and Senegal (2010-2017). Own elaboration on the basis SABRE data.

Purpose of visits to UK (Q1 2017) Own elaboration on the basis UK-IPS data.
Public perceptions on migration (research question)

“The Daily Me is not an utopian dream, and it would create - is creating - serious problems from the democratic point of view”

“Echo chambers can lead people to believe in falsehood”

Is polarization taking place in public discourses on migration in Twitter?
Public perceptions on migration (data)

Number of Tweets related to migration, migration & Trump, migration & Brexit collected from the public Twitter API

Key indicators of influence and migration specificity for a selection of Twitter accounts

46 M Tweets (99.3 K/day)

5 M accounts
Public perceptions on migration (preliminary results)

1. Network based on mentions between users
2. Community detection on network
3. Information (url) propagation between communities
4. Characterize communities based on accounts descriptions

US Conservatives

International organizations and researchers

Information propagation between two communities of Twitter accounts
Lesson learnt

Limitations
• Proprietary (see CDR)
• Noisy
• Sampling bias
• Not targeted to migrants
• Case studies
• From models to machine learning

Opportunities
• Larger sample size in respect of classical surveys
• Not bound to statistical definitions
• Space-time granularity
• Timeliness
• Covering areas of the world with limited migration statistics
Conclusions

• ‘Alternative’ often better than ‘Big’
• CDR great potential for human mobility but proprietary and no supranational coverage
• A lot of data on migration is already available in administrative sources and micro census (it is just difficult to assemble it)
• Participatory approaches add value to data
• Applications at supra-national level possible for Twitter for discourses about migrants

• LinkedIn potential for high skilled migration and temporary migration
Europe Media Monitor (EMM)

- 6,000 News Sites
- 300,000 articles per day
- 70 Languages
- 1000 Categories/Topics
- 30,000 keywords
- Runs 24/7
- Visitors 25,000

Domains:
- Border Security, Cybercrime & Fighting Crime
- Socio Political Event Monitoring
- Public Health Threats & Food Safety
- Conflict Early Warning
- Science and the Citizen
EMM: JRC Research into Fake News

- Frequent sources of disinformation
- Clustering articles
- Monitoring narratives
- Text similarity to known bad articles
- Propagation through Twitter
- Examining tonality of known fake news
EMM and Fake News: partnerships

**Existing**
- EEAS East Stratcom Task Force (http://euvvsdisinfo.eu)

**Looking for**
- Domains: migration, public health threats.
- Content: narratives, sources, examples of fake news.
- Activities: assess results produced by our tools.
Any questions?
You can contact us at JRC-EMM-SUPPORT@ec.europa.eu