



This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under grant agreement No 870661



HumMingBird

Using Big Data for Analysis of Migration and Mobility

Albert Ali Salah, Alina Sirbu, Rahman Momeni,
Bilgeçağ Aydoğdu, Carlos Arcila Calderon, Tuba
Bircan

Objectives

- to identify key uncertainties and reappraise the migration concepts
- projections based on the analyses of the patterns, motivations and modalities of migration
- to widen the EU's viewpoint on the policy migration nexus
- qualitative scenario building on the stories of migrants en route to complement quantitative scenarios
- to validate big data technologies to help estimate stock migration and migration flows
- to merge knowledge of the quantitative and qualitative approaches to provide a holistic view of migration

Premises

- Investigating how to harness new data sources for migration analysis and policy making
- Identifying the uncertainties and reappraising assumptions
- Exploring the reasons why migration predictions may not hold
- Conceptualising “Hubs of Migration”
- Listening to the migrants on the way to understand the root causes of migration
- Using big data as proxies for migration measures

Data sources: Satellite imaging

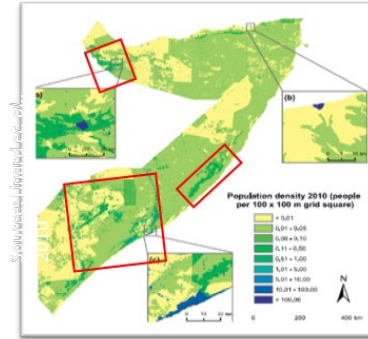
Aim: Demonstrate how satellite-based holistic analysis improve the understanding of environmental migration movements (2015 to 2019).

AOI: 13 districts (112,000 km²) in centre and South of Somalia where mostly affected by environmental migration.

Datasets used:

1. Sentinel-1 and 2
2. NASA DEM
3. Soil moisture index (SMI)
4. Standardised Precipitation-Evapotranspiration Index (SPEI)
5. Internal displacement data

(a). Somalia: population map 2010

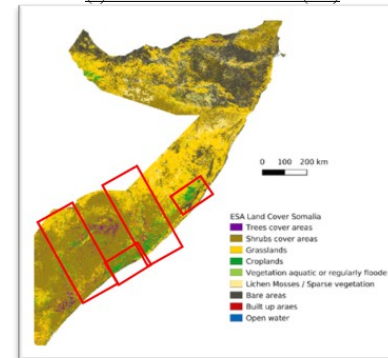


Source: <http://www.mappery.com/map-of/Somalia>

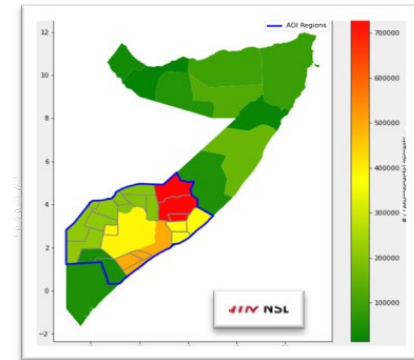
(b). Somalia: major rivers map



(c). Somalia: Land Cover (LC)



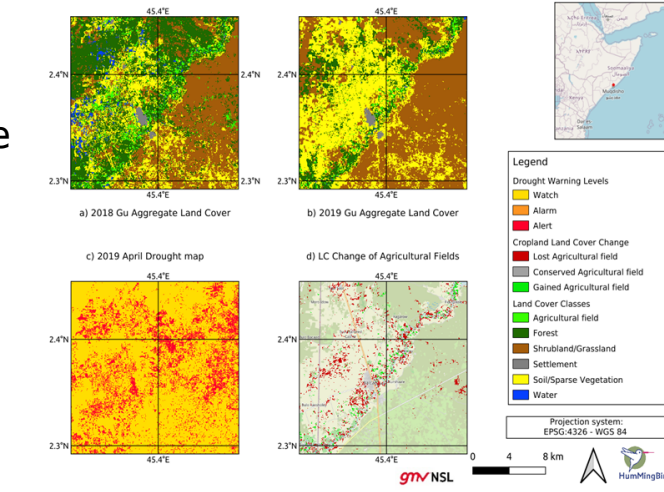
(d). Environmental migration



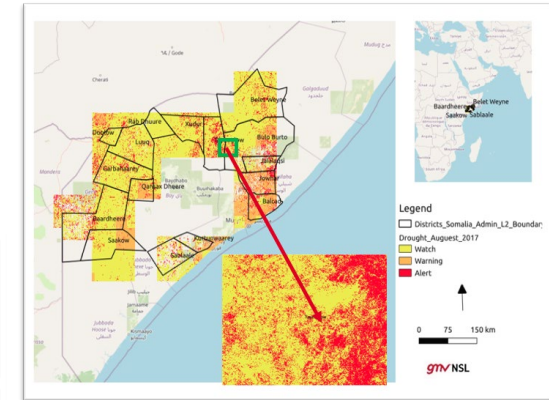
Data sources: Satellite imaging

Satellite derived products:

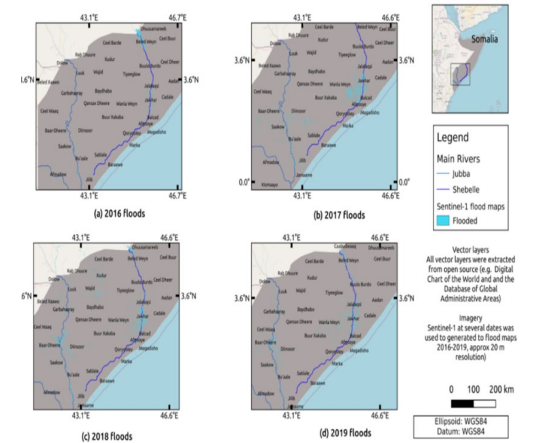
1. Vegetation spectral indice (NDVI, SAVI, and LAI)
2. Agricultural Drought Indicator (ADI)
3. Flood maps
4. Land Cover change maps



Drought LC change in the Gu Season of 2019. (c); affected drought areas and (d); matchup with areas of agricultural field loss



A mosaic of drought map in Somalia August 2017



Flood map time-series

Data sources: Satellite imaging

Key findings:

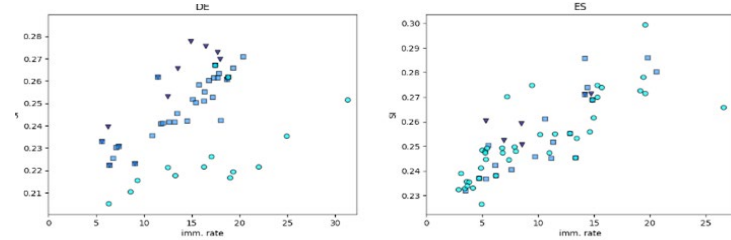
1. Slow-onset climate events (like drought) and rapid-onset (like flood) climate events have different impacts on society and the responses.
2. The impact of floods is mostly localised and limited to the districts suggesting that migrants from major floods try to stay close to their homes but outside the direct disaster zone.
3. Drought has a more progressive impact and people start to leave their homes with some delay after the drought started and it takes a much longer time to return home.
4. Compared to regular floods that are part of the agricultural climate and essential for the agricultural growing season, extreme floods are more likely destructive and lead to migration.

Data sources: Social media

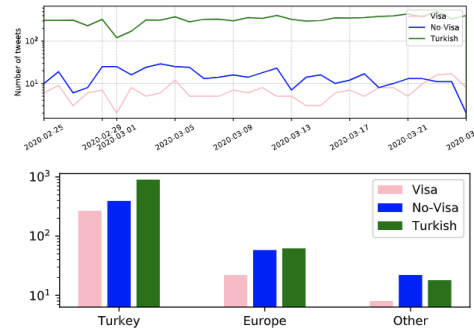
- Nowcasting stocks and flows
 - Twitter
 - Superdiversity index - indicator based on sentiment assigned to terms, high correlation with immigration rates (e.g. UK)
 - Dataset:
 - doi.org/10.5281/zenodo.6367083
 - Languages employed and geolocation-language as proxy for nationality (e.g. Turkey border rush)
 - Labelling migrants - nationality and residence based on geolocation of user and friends - stocks
 - Facebook - Social Connectedness Index
 - Important in nowcasting models

Country	SI	nTweet	twPc	nLang	langS	TTR
FR	0.77	0.32	-0.29	0.48	0.50	0.25
DE	0.48	0.30	0.05	0.48	0.50	0.15
NL	0.35	0.34	0.28	0.52	0.64	0.34
ES	0.81	0.31	0.47	0.05	0.53	-0.01
IE	0.53	0.22	0.24	0.32	-0.47	0.42

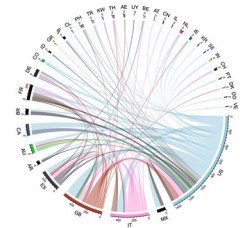
▼ NUTS1
 ■ NUTS2
 ● NUTS3



Scatter plot of Superdiversity and immigration by country



y	By citizenship (748 couples, 200+ countries)			
	ESTAT UMI 2019 cit, without sci 2020		ESTAT UMI 2019 cit, with sci 2020	
Model n.	1	20	1	18
Feature	coef, P> t	coef, P> t	coef, P> t	coef, P> t
Intercept	-0.0027		0.2080 ***	0.2072 ***
sci_2020			0.2024 ***	0.2024 ***
origin_PDI	-0.0056		-0.0009	
destination_PDI	0.0012		0.0036 *	
origin_IDV	-0.0053		-0.0011	
destination_IDV	-0.0021		-0.0051	
origin_UAI	-0.0033		-0.0017	
destination_UAI	-0.0033		-0.0049 *	-0.0023 *
origin_MAS	-0.0035	-0.0053	-0.0005	
destination_MAS	0.0022		-0.0002	
origin_area	0.0002		0.0004	
destination_area	-0.0017		-0.0040 *	-0.0020 *
origin_fb_users	-0.0012		-0.0012	
destination_fb_users	-0.0038	-0.0071 ***	0.0036	
origin_fb_users_perc	0.0054	0.0043	0.0039 *	0.0025
destination_fb_users_perc	0.0012		0.0010	
geodesic_distance_km	-0.0114 *	-0.0111 ***	-0.0025	-0.0038 **
origin_gdp_2018	0.0007		-0.0015	
destination_gdp_2018	0.0026		0.0026	
gdp_diff_2018	-0.0002		-0.0018	
neighbours	-0.0049		0.0042 **	0.0040 **
share_cont	0.0010		0.0017	
share_rei	0.0031	0.0033 **	0.0009	0.0013 *
share_lang	0.0084 ***	0.0077 ***	0.0010	0.0012
R2 (centered)	0.115	0.103	0.780	0.777
AIC	-3043.	-3065.	-4081.	-4100.
BIC	-2941.	-3037.	-3975.	-4059.



International Symposium on Intelligent Data Analysis
 IDA 2020: *Advances in Intelligent Data Analysis XVIII* pp 274-286 | [Cite as](#)

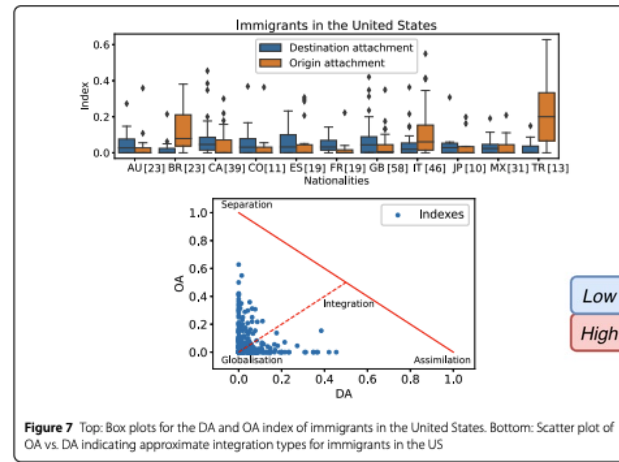
Digital Footprints of International Migration on Twitter

Authors: [Authors and affiliations](#)

Jisu Kim ✉, Alina Sirbu ✉, Fosca Giannotti ✉, Lorenzo Gabrielli ✉

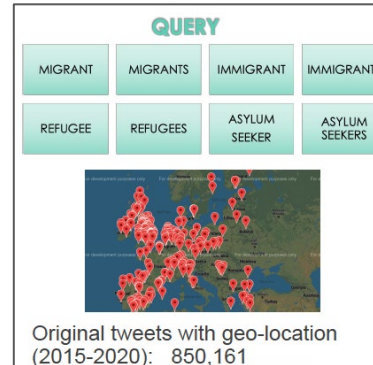
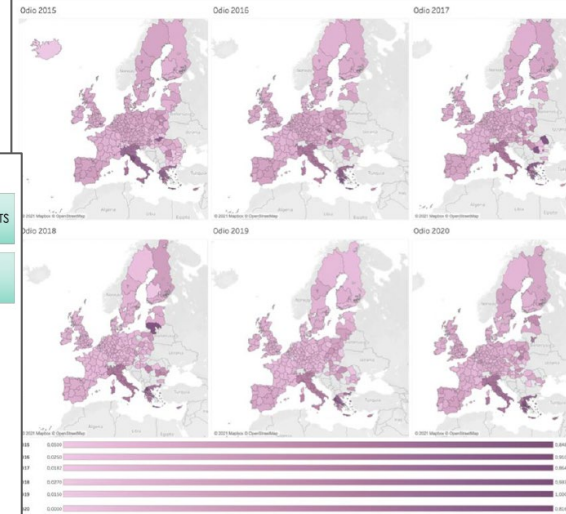
Data sources: Social media

- Integration - Twitter data
 - Origin and destination attachment
 - Attachment based on the topics discussed on Twitter
 - Dataset: doi.org/10.6084/m9.figshare.19348058.v1
 - Sentiment and hate speech towards migration
 - Indicators for the entire Europe
 - Dataset: doi.org/10.6084/m9.figshare.17430560.v1



	Low DA	High DA
Low HA	Marginalisation	Assimilation
High HA	Separation	Integration

Average level of hate speech per year from 2017 to 2020 on Twitter by European regions (NUTS 2)

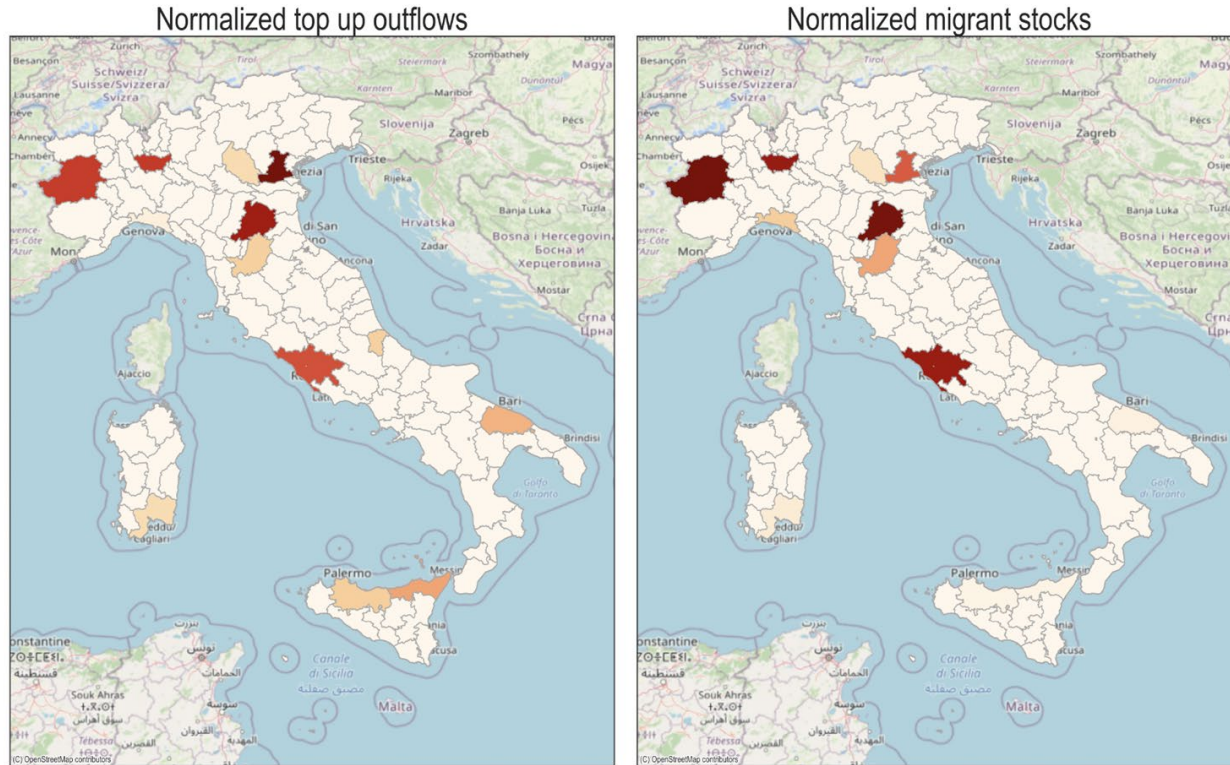


Data sources: Mobile phones

1. Call Detail Records (CDR)
2. Extended Detail Records (xDR)
3. Inbound Roaming (IR)
4. Outbound Roaming (OR)
5. Airtime Top-up Transactions (ATT)

Table: H(igh), M(edium), L(ow) values indicated. IR stands for Inbound Roaming, and OR for Outbound Roaming.

Indicators	Source	xDR	ATT	CDR	IR	OR
Internal migration flow		H	L	H	L	L
International migration flow		L	L	L	M	M
Migrant stocks		H	M	H	M	M
Migrant integration		M	M	H	L	L

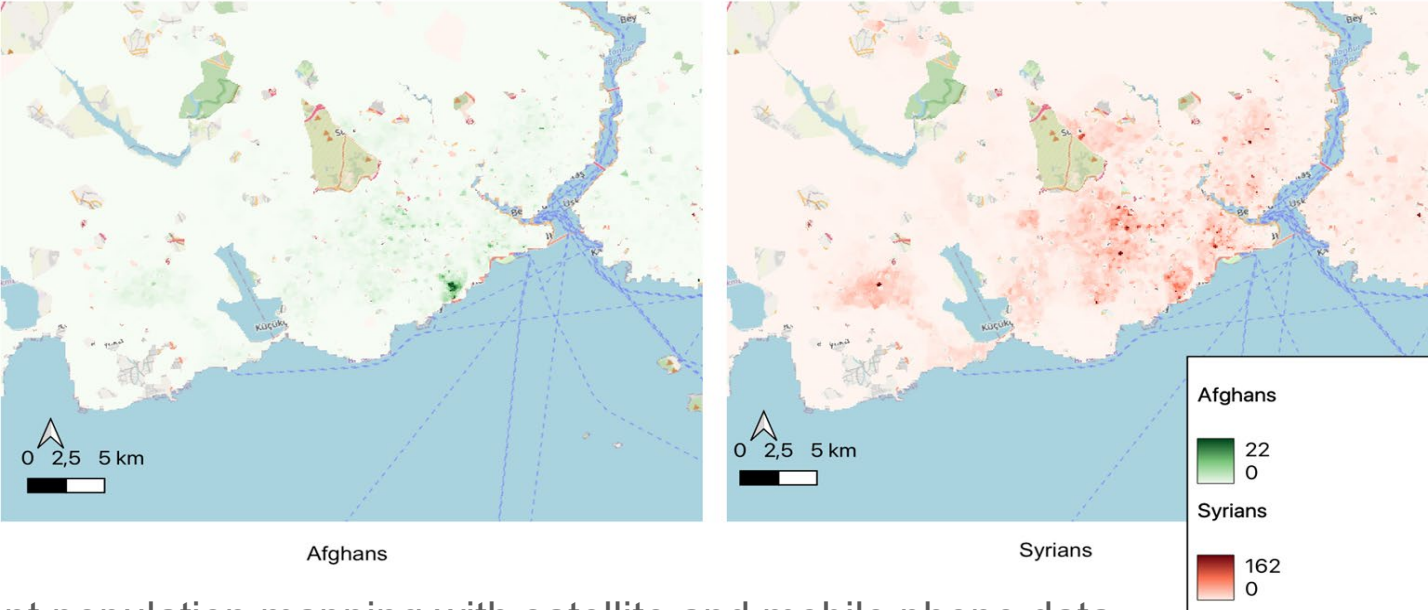


(a) The left figure shows the distribution of airtime top-up outflows to Cameroon, whereas the right figure shows the distribution of Cameroonian migrants living in Italy according to official statistics.

Aydoğdu, B., H. Samad, S. Bai, S. Abboud, I. Gorantis, A.A. Salah, "Analysing international airtime top-up transfers for migration and mobility," *Int. Journal of Data Science and Analytics*, to appear.

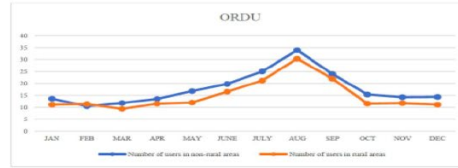
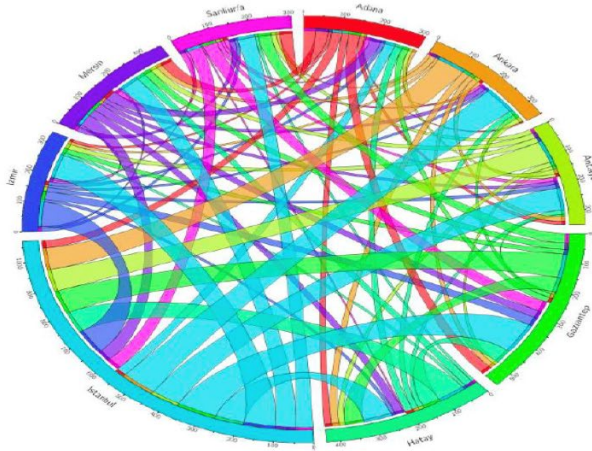
Data sources: Mobile phones

Population distribution of migrants in Istanbul



- Migrant population mapping with satellite and mobile phone data
- Spatial segregation indicators

Data sources: Mobile phones



- Migration detection algorithms (xDR vs. CDR)
- Seasonal migration
- Start and end dates
- Characteristics of migrants vs. non migrants

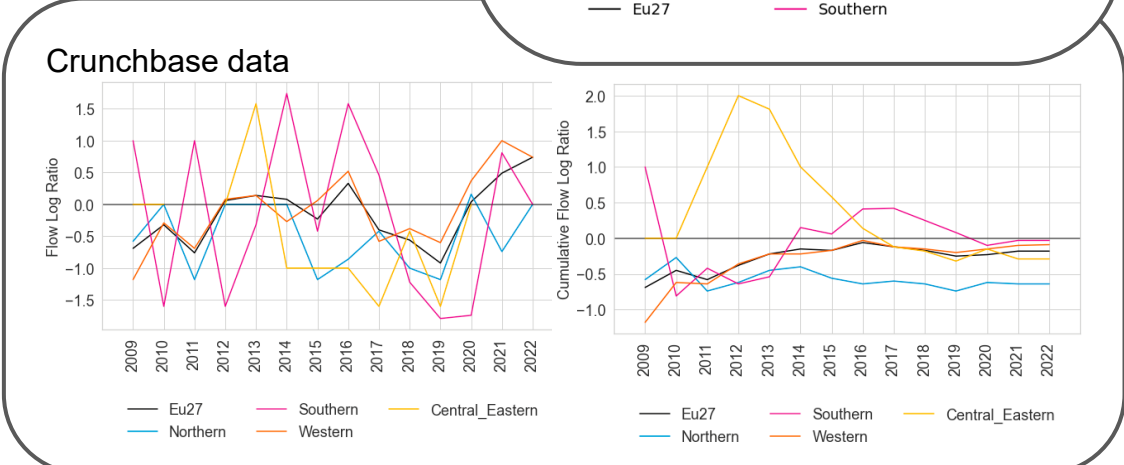
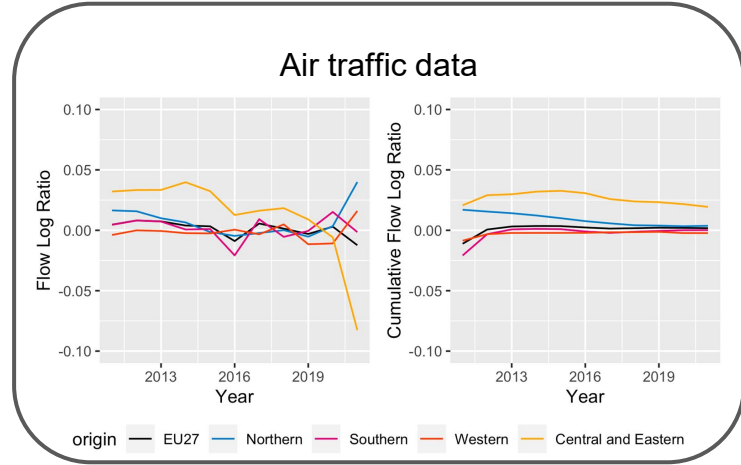
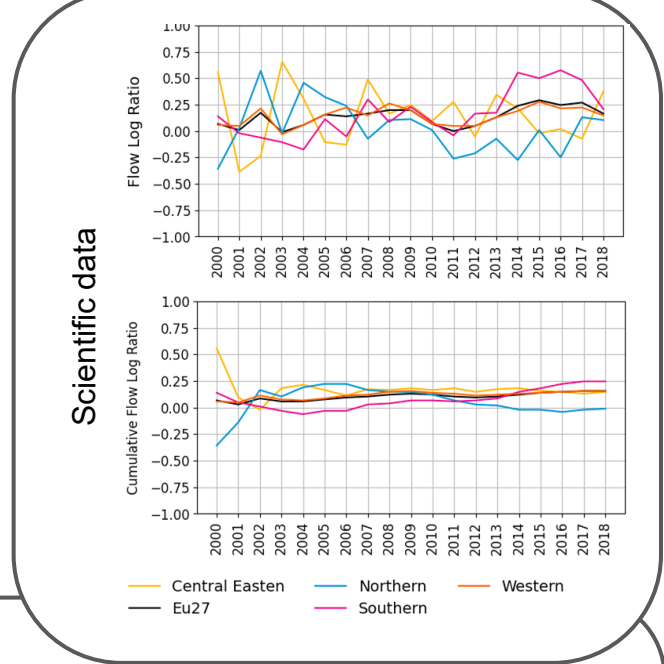
Alısık ST, Aksel DB, Yantaç AE, Baruh L, Salman S, Kayı I, İçduygu A, Bensason I (2019). UDMIT: an urban deep map for integration in Turkey. In Data for Refugees Challenge Workshop

Data sources: Other sources

- Highly skilled migration
 - Scientific flows - extracted from scientific publication data
 - EMAKG Dataset: doi.org/10.5281/zenodo.5888647
 - Crunchbase flows- computed from crunchbase.com
- Air traffic data
 - General mobility trends
- E.g. Brexit analysis

$$FLR(t) = \log_2 \frac{F_{C \rightarrow UK}(t)}{F_{UK \rightarrow C}(t)}$$

$$CFLR(t) = \log_2 \frac{\sum_{t_i \leq t} F_{C \rightarrow UK}(t_i)}{\sum_{t_i \leq t} F_{UK \rightarrow C}(t_i)}$$



Combining data sources

Some data sources are more amenable to combination.

We studied social media data + mobile phone data for investigating border rush.

Satellite data + mobile phone data for improving the interpretation and visualization of mobile data, as well as estimation of stocks.

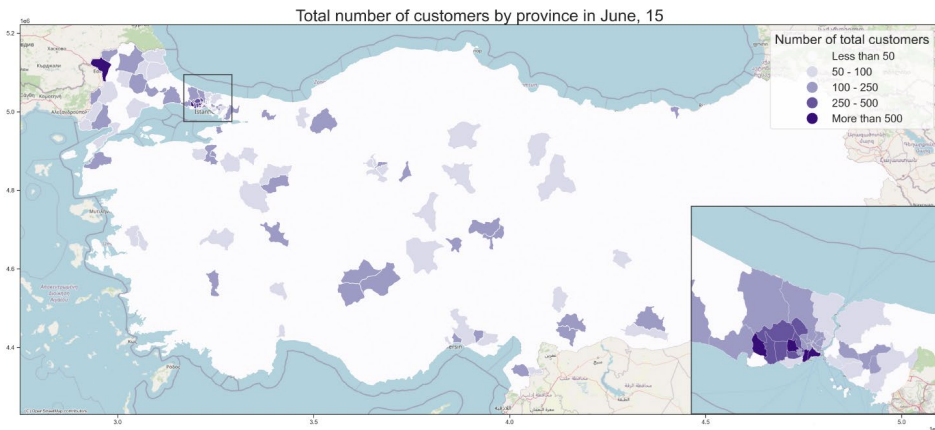


Figure 5: Total number of customers by province on June, 15. The figure only shows the distribution of people who were at the European border during the border rush in March, but stayed in the dataset after the rush. The city of Istanbul is zoomed for better visibility in the lower right corner.

C. Arcila Alderon, T. Bircan, B. Gürbüz, O. Öneş, B. Aydoğdu, A.A. Salah, A. Sirbu,
“Combining Twitter and mobile phone data to observe border rush”, in preparation

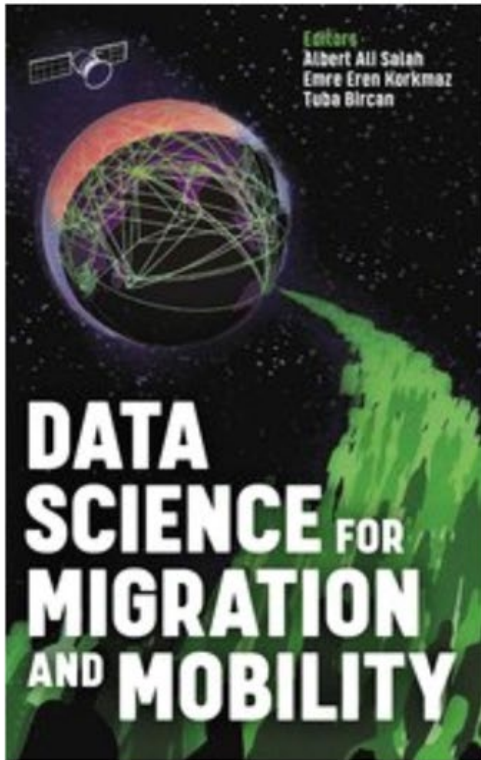
Ethical and legal issues

Social media data becomes 'small data' when focusing on migrants

- risk of re-identification
- need to minimise data and aggregate smaller groups

Individual versus group privacy - minimisation and aggregation of group data as well.

Legal/practical accessibility of data sources may change over time.



Data Science for Migration and Mobility

Edited by **Albert Ali Salah**, **Emre Eren Korkmaz**, and **Tuba Bircan**

British Academy

- Provides a multi-faceted perspective into data science for migration and mobility while guiding the reader through disciplinary jargon and conceptualisations.
- Helps migration scholars and students to understand the potential of new data sources with many case studies, and provides entry points for the available tools in the field.
- Written by leading experts from migration research, data and computer science, computational social science, sociology, demography, law, political science, economics, linguistics and psychology.

<https://hummingbird-h2020.eu/publications>



HumMingBird



COORDINATOR

HIVA - Research Institute for Work and Society, KU Leuven (BE)

PARTNERS

University of Utrecht, Department of Information and Computing Sciences (NL) • University of Salamanca, Faculty of Social Sciences (ES) • Otto-Friedrich-University of Bamberg, Geographic Migration and Transition Studies (DE) • CESSDA ERIC (NO) • EKKE - National Centre for Social Research (GR) • IEN - Institute of Economic Sciences, Data Center Serbia for Social Sciences (RS) • Malmö University, Malmö Institute for Studies of Migration, Diversity and Welfare (SE) • Turkcell Technology (TR) • GMV Innovating Solutions Ltd, Remote Sensing Services and Exploitation Platforms (UK) • MPG - Migration Policy Group (BE) • ENoMW - European Network of Migrant Women (BE) • White Research (BE) • University of Pisa, Department of Computer Science (IT) • CNR - National Research Council (IT) • VUB - Vrije Universiteit Brussel, Interface Demography, Department of Sociology (BE) • University of Zurich, Department of Social Anthropology and Cultural Studies (CH) • CILD - Italian Coalition for Civil Liberties and Rights (IT) • Caritas International (BE)

VISIT: <http://www.hummingbird-H2020.eu>

CONTACT US: HumMingBird@kuleuven.be

FOLLOW US:

