

Joint Managing Editors:

- Solon Arditis (Eurasylum)
- Frank Laczko (International Organization for Migration – IOM)

Editorial Advisers:

- Joanne van Selm (Eurasylum)
- Karoline Popp (International Organization for Migration – IOM)

Editorial Coordinator:

- Valerie Hagger (International Organization for Migration – IOM)

Editorial Assistants:

- Miguel De Lim (International Organization for Migration – IOM)
- Anna Lyn Constantino (International Organization for Migration – IOM)

Editorial Committee:

- Aderanti Adepoju (Human Resources Development Centre, Lagos, Nigeria)
- Richard Ares Baumgartner (European Agency for the Management of Operational Cooperation at the External Borders of the European Union – FRONTEX, Warsaw)
- Peter Bosch (European Commission, Brussels)
- Juan Carlos Calleros (Staff Office of the President of Mexico)
- Howard Duncan (Metropolis, Ottawa, Canada)
- Neli Esipova (Gallup World Poll, New York)
- Araceli Azuara Ferreiro (Organization of American States – OAS, Washington, D.C.)
- Philippe Fargues (Migration Policy Centre – MPC, Florence)
- Lukas Gehrke (International Centre for Migration Policy Development – ICMPD, Vienna)
- Shahidul Haque (Ministry of Foreign Affairs, Government of the People’s Republic of Bangladesh)
- Maia Welbourne (Citizenship and Immigration Canada – CIC, Ottawa)
- Chris Hedges (United Kingdom Border Agency – UKBA, London)
- Michelle Leighton (International Labour Office – ILO, Geneva)
- William McClure (Australian Department of Immigration and Border Protection)
- Jennifer McDonald (Passport, Immigration and Citizenship Agency, Ministry of National Security, Jamaica)
- Sankar Ramasamy (Department of Labour, New Zealand)
- Dilip Ratha (World Bank, Washington, D.C.)
- Cécile Riallant (EC-UN Joint Migration and Development Initiative, Brussels)
- Nand Kishore Singh (Member of the Indian Parliament, New Delhi)
- Simon Tonelli (Council of Europe, Strasbourg)
- Adriana van Dooijeweert (Dutch Advisory Committee on Migration Affairs – ACVZ, The Hague)

MIGRATION POLICY PRACTICE

A Bimonthly Journal for and by Policymakers Worldwide

ISSN 2223-5248

Vol. IV, Number 2, April–June 2014



IOM trains city welfare and development staff of Cadiz City, Negros Occidental on how to use the IOM Vulnerability Index forms (18 Feb). © IOM 2014 (Photo by Alan Motus)

CONTENTS

2	Introduction <i>Solon Arditis and Frank Laczko</i>
4	For a truly European immigration policy <i>Martin Schulz</i>
8	Working with cities on mobility, diversity and prosperity <i>Colleen Thouze</i>
16	Migrants or expatriates? US citizens as a migrant group <i>Amanda Klekowski von Koppenfels</i>
20	Can Big Data help us achieve a “migration data revolution”? <i>Frank Laczko and Marzia Rango</i>
30	Publications
32	MPP Readers’ Survey

Published jointly by the International Organization for Migration (IOM) and Eurasylum Ltd.



International Organization for Migration (IOM)





Can Big Data help us achieve a “migration data revolution”?

Frank Laczko and Marzia Rango¹

Introduction

“Big Data” is generated every time we send a text message, make a mobile phone call, run an Internet search, transfer money online or interact on social media. This article looks at the potential of using such data to improve data on migration at a time when there are growing calls to improve the migration evidence base. To what extent can such new data sources provide policymakers with more timely and comparable data on migration?

At the 2013 UN High-level Dialogue on Migration and Development, it was widely agreed that policymakers needed much better data on migration and development. For example, the best migration data often comes from censuses, which can be several years old and, thus, outdated. Only 12 countries in sub-Saharan Africa have conducted a census in the last decade. The lack of migration data often leads to public misperceptions about the scale of migration and its impact. Poor official statistics on migration makes it difficult for decision makers around the world to develop effective policies. The lack of robust migration data also makes it more difficult to argue for the inclusion of migration indicators in the post-2015 development agenda.

Calls for better data on migration are not new, but they come at an important time when the development community is recommending the need for a “development data revolution” (UN High-level Panel report on the post-2015 development agenda, 2013). Bill Gates, for example, recently argued that improving the lives of the poor will require a revolution in development data:

“Ultimately, the better the data available in the development field, the higher the quality of people’s lives in poor countries.”²

1 Frank Laczko is Head of the Migration Research Division at IOM Headquarters in Geneva; Marzia Rango is a Research Intern in the Migration Research Division at IOM Headquarters in Geneva.

2 Visit www.adb.org/features/big-data-vital-statistics-development-for-more-information-on-Big-Data.

A “data revolution” would mean developing new approaches, new partners and new sources and uses of data, and could add impetus to calls for better data on migration and development. As the 2013 High-level Panel (HLP) report notes:

“This is not just about governments. International agencies, CSOs and the private sector should be involved. A true data revolution would draw on existing and new sources of data to fully integrate statistics into decision-making, promote open access to, and use of, data and ensure increased support for statistical systems.” (HLP, 2013).

Given this context, governments and the development community are becoming increasingly interested in exploring the potential of using Big Data, the unprecedentedly large amount of data automatically generated through the use of digital devices or web-based platforms and tools. The UK Government, for example, plans to spend over USD 100 million on research to explore the potential of using such data. Innovations in technology and reductions in the prices of digital devices worldwide mean that digital data is being produced in real time and at an unprecedented rate. In particular, there are now more than 6 billion cellular phones, 5 billion of which are used in developing countries. There has been exponential growth in mobile phone penetration in developing countries, reaching 89 per cent in 2013.³

This article explores how Big Data might help to improve our understanding of migration trends around the world. We consider some of the advantages and disadvantages of using such data and illustrate through concrete examples its potential uses in the study of migration. The article outlines ways in which Big Data can be used to: (a) track post-disaster displacement using call detail records (or “CDRs”); (b) identify modalities and determinants of mobile money transfers through other kinds of mobile phone data; (c) estimate and predict migration flows and rates through the Internet protocol (IP)

3 See www.itu.int/en/ITU-D/Statistics/Documents/facts/ICTFactsFigures2013-e.pdf for more information.

addresses of website logins and sent e-mails; (d) infer migration trends and compare patterns of internal and international migration using geo-located social media data; and (e) analyse transnational networks and diaspora groups or migration-related public discourse through social media content.

Defining “Big Data”

The term “Big Data” usually refers to the vast amount of data generated by the use of digital devices and web-based tools and platforms (Latouzé, 2012). Big Data can also come from other types of digital sensors and meters, such as satellite imagery. More broadly, Big Data also refers to the growing analytical capacity and increasingly powerful computation methods available today to analyse complex and huge amounts of digital data. The main characteristics of Big Data are commonly referred to as the “Three Vs”: volume, velocity and variety (Laney, 2001), with the recent addition of a fourth “V,” standing for the potential and, typically, financial value that can be extracted from the collection, aggregation and analysis of Big Data.⁴

The exponential growth in the use of mobile phones, social media and Internet-based services worldwide means the “volume” of data available is larger than ever before in human history, but Big Data does not exclusively owe its name to its size. In fact, any “passively collected data deriving from daily usage of digital services” constitutes Big Data – a phenomenon also known as “data exhaust” (Latouzé, 2012). “Velocity” refers to the unprecedented speed with which data is generated, and “variety” to its complexity, as Big Data is constituted by structured and unstructured data – an example of the former are CDRs, which can be objectively arranged and analysed, while any type of social media content constitutes unstructured data, as it is subjective and non-verifiable.

4 McKinsey Global Institute estimated the financial value that can potentially be extracted from the use of Big Data across different sectors: Gains would be in the order of USD 300 billion for the US health-care sector each year, and EUR 250 billion for the European public administration sector (McKinsey Global Institute, 2011).

Growing interest in the potential of using “Big Data”

Since the mid-1990s, when the concept of Big Data in its present connotation was arguably first introduced,⁵ businesses, academics and policymakers have paid increasing attention to ways of harnessing the potential of this kind of data in their respective fields. Interest in using Big Data in the field of international development has exploded in recent years. There has been a surge of publications dedicated to Big Data and international development such as, notably, the World Economic Forum’s “Big Data, Big Impact: New Possibilities for International Development⁶” and the UN Global Pulse’s “Big Data for Development: Challenges and Opportunities.”⁷ The UN Global Pulse initiative was established in 2009 with the specific purpose of “exploring how digital data and new real-time analytical technologies can provide a better understanding of changes in human well-being and emerging vulnerabilities.”⁸ Research from UN Global Pulse focused, among other aspects, on the use of social media as early warning signals for commodity price volatility or spikes in unemployment rates in specific countries, and there are plans to set up a network of “Pulse Labs” aimed at establishing public–private partnerships to harness the potential of Big Data in development program planning and monitoring. Big Data could also be used to improve decision-making in health care, natural disaster and resource management, and economic productivity (Hilbert, 2013). Institutions, including the Harvard School of Public Health and the Qatar Computing Research Institute, offer useful resources on Big Data for international development.

Potential uses of Big Data to study migration

Discussion of the potential uses of Big Data to fill migration data gaps is still in its infancy. Papers presented at the 2014 Global Forum on Migration and Development and the UN High-level dialogue on

5 See <http://bits.blogs.nytimes.com/2013/02/01/the-origins-of-big-data-an-etymological-detective-story> for more on the history of Big Data.

6 This report may be downloaded from www3.weforum.org/docs/WEF_TC_MFS_BigDataBigImpact_Briefing_2012.pdf.

7 This report may be downloaded from www.unglobalpulse.org/sites/default/files/BigDataforDevelopment-UNGlobalPulseJune2012.pdf.

8 UN Global Pulse official website: www.unglobalpulse.org.

Migration and Development in 2013 barely mentioned the subject. The number of migration studies drawing on Big Data is still relatively limited but rapidly increasing. Some examples are presented below (see Table 1 for a summary of the different types of studies which have been conducted in the migration field).

CDR data to track forced displacement or infer internal migration patterns

CDRs are digital records automatically generated and collected by mobile network operators every time a mobile phone call is made. They generally include information about the time and duration of the call, the calling and receiving numbers and, most important for migration research, the approximate locations of the caller and the receiver (i.e. the locations of the cell towers to which the SIM cards connect during the call). The identification of either location is more or less accurate depending on the distance between cell towers (smaller in urban areas relative to rural areas).

Several studies have used CDR data to track population movement in the aftermath of disasters, given the rapid and uncontrolled character of this kind of displacement, and detect patterns of internal and circular migration, which are typically hard to capture using traditional sources of data like national censuses and household surveys. Bengtsson et al. (2011) estimated the magnitude of and the trends in population movements out of Port-au-Prince following the 2010 Haiti earthquake. The authors obtained (anonymized) records from one of the country's largest mobile network operators on 2.8 million SIMs that made at least one call during the pre- and post-earthquake periods. Records included the location of the mobile phone tower to which each SIM connected during the call, which made it possible to follow mobile phone subscriber's movements. The estimates were close to early estimates by the national civil protection agency in terms of number (though quite different in terms of distribution). More importantly, they were similar to retrospective population-based data, showing the predictive power of geo-referenced mobile phone call data. A parallel study was carried out in New Zealand after the Christchurch earthquake with similar results (ACAPS, 2013).

CDR data uses are not confined to studies on post-disaster displacement. For example, Blumenstock (2012) used mobile phone records to infer internal migration patterns in Rwanda. The potential of Big Data for the study of internal and circular migration

“Records included the location of the mobile phone tower to which each SIM connected during the call, which made it possible to follow mobile phone subscriber's movements.”

is particularly significant, given that traditional data collection tools such as government censuses and household surveys do not typically capture such patterns. Along the same lines, Eagle et al. (2009) looked at differences in movement between rural and urban residents. Also, Frias-Martinez et al. (2010) studied the impact of socioeconomic status on migration in one Latin American city by looking at ICT-generated data, and Lu Wetter et al. (2013) analysed the predictability of human mobility using a similar methodology.

Identify modalities and determinants of mobile money transfers using mobile phone data

Apart from call records, existing research shows that other kinds of mobile phone data can also be collected, analysed and used in several ways. Blumenstock et al. (2013) collected the comprehensive log of all mobile phone activity – calls, text messages, money transfers and purchases – of 1.5 million subscribers from Rwanda's primary telecommunications operator during a four-year period. More specifically, they used data from interpersonal transfers between subscribers of the network – date, time, value and anonymized identifiers for senders and receivers – to understand how money sent through phones is used to share risk in the aftermath of natural disasters (in particular, the 2008 Rwanda earthquake) and the determinants of mobile money transfers.

Using IP addresses of website logins and e-mail sending to estimate international migration trends

State et al. (2013) and Zagheni and Weber (2012) provide examples of innovative approaches to estimate migration flows using Big Data in the form of IP address-based location data. In the first paper,

short- and medium-term migration flows were estimated using IP addresses from repeated logins to Yahoo! Services by over 100 million anonymized users during a one-year period. The authors constructed the sample following a data-cleaning protocol aimed at reducing the inaccuracies (“noise”) normally associated with the use of IP geo-location data – for instance, users might use a proxy server to connect to the Internet, in which case their physical location would not correspond to the IP-based geo-location. By constructing a statistical model for the prediction and analysis of migrant (as well as short-term tourist) flows, the authors were able to infer global mobility patterns on the basis of “conditional probabilities of migration,” or else the likelihood that a migrant from one country will go to another country. This model was also able to capture patterns of circular or “pendular” migration which cannot be identified using traditional data sources such as national censuses or household surveys.

In Zagheni and Weber (2014), IP addresses were used to map the geographic locations from where 43 million anonymized users sent e-mail messages within a given period. The authors linked such information to users’ self-reported age and gender information to estimate age- and gender-specific migration rates, with the assumption that a user’s country of residence is the one from where he or she sent most of his or her e-mails. In order to correct for selection bias – the fact that the sample considered is normally quite large, but not necessarily representative of the whole population due to differences in Internet usage patterns across countries and age groups – the researchers constructed a model which took into account differentials in Internet penetration rates by age and gender across countries. The reason for this is that in countries where Internet penetration is low, Internet users might be a select group of highly educated or mobile people, and, thus, mobility might be overestimated. As the authors argue, this type of research has the potential to revolutionize the way in which migration statistics are compiled, given its global scale, the relatively short time-frame in which it can be carried out given the velocity of data generation, the real-time collection, and the limitations associated to the use of demographic registration systems, as migrants often do not register in the destination country (when they do, there is often a lag between their movement to the destination country and actual registration).

Using geo-located social media (Twitter) data to infer migration trends

Social media is becoming an important source of information to study migration trends, as it makes it possible to identify and (anonymously) follow users over time through geo-referenced posts. Zagheni et al. (2014) used geographic information (latitude and longitude) derived from Twitter posts to estimate migration in countries of the Organisation for Economic Co-operation and Development (OECD). In particular, the researchers downloaded geo-located “tweets” (i.e. Twitter posts) of about 500,000 users who posted at least one tweet in the sample period using Twitter Application Programming Interface, and mapped users to countries, dividing them into those who “tweeted” from one, two, three countries or more. They then selected a sample of users for which they had detailed and consistent information over the period considered – at least three geo-located tweets in each of the four-month periods over two years – scaling down the sample to about 15,000 users. Their main methodological innovation is the use of a difference-in-difference strategy to reduce the “selection bias” inherent in the use of social media data for statistical inference, given that Twitter (as well as other social media platforms) users do not constitute a representative sample of the whole population. In other words, assuming that the composition of Twitter users changes in a similar way across countries, one can extrapolate migration trends by comparing relative changes in migration rates for a single country with relative changes for the group of reference at two points in time.

Geo-referenced Twitter data allows for comparisons between patterns of internal and international migration, because it makes it possible to differentiate between country residents, that is, those who post tweets from the reference or origin country, and migrants, that is, those who “tweet” from another country. Also, this data has great potential in migration forecasting because it is available well before information from national censuses or surveys. More work from the scientific community is needed, however, first, to overcome selection bias in the use of social media to track mobility, thereby ensuring the external validity of studies based on this kind of data; and second, to make it possible to predict mobility patterns with highly complex, unstructured and hard-to-verify social media data.

Alternative uses of social media content in migration research

Social media represent another Big Data source of useful information for migration studies. As previously explained, social media platforms such like microblogging site Twitter can be used to infer mobility patterns using the geographic information linked to users' posts. However, the potential of social media data in the field of migration studies is not confined to the identification or prediction of mobility trends. Social media contain unstructured and often publicly available data (unlike e-mail messages, which are private in nature), which can be used in a variety of ways as a growing academic literature shows.

A special issue of the *Journal of Ethnic and Migration Studies* entitled "Migration and Diaspora in the Age of Information and Communication Technologies" (No. 9, Vol. 38, 2012) was entirely dedicated to issues that emerge from the intersection of ICTs and migration, such as diaspora and transnationalism, as well as to ways in which ICTs impact migrants' lives. Empirical studies have investigated the role of social media platforms in the construction of transnational networks (Nedelcu, 2012). Social media have been found not only to ease communication between migrants, but to actively transform the nature of migrant networks and, in some instances, facilitate migration: potential migrants can use social media to establish ties with people already residing in the potential destination country, or to collect information through informal channels (Dekker and Engbersen, 2013). Social media can also be useful to investigate the political activism of migrants and minority groups (Conversi, 2012; Kissau, 2012), migrants' integration into the host society (Rinnawi, 2012), as well as the transformation of family relationships and the impact of migration on migrants' well-being, for example.

Lastly, social media can be rich sources of information for analysing people's attitudes towards immigrants and immigration, as well as migrant integration in the host society. For instance, the project Unite Europe aims to analyse social media content generated by citizens to inform local integration policies.⁹ However, there are serious challenges associated with this kind of research, as information gathered on social media might represent a very small sample of public opinion or might be exaggerated or in any case not correspond to people's actual views (Mc Gregor and Siegel, 2013).

9 Unite Europe official website: www.uniteeurope.org.

To sum up, the most common applications of Big Data in the context of migration seem to be in the use of mobile phone call data to track and assist displaced populations in regions exposed to disaster risk. There are several ongoing studies of this type. Academic researchers at the Flowminder Foundation work to improve public health outcomes in disaster-, conflict- and disease-affected regions through the collection, aggregation, analysis and dissemination of mobile phone data to track displacement and therefore improve relief responses.¹⁰ The United Nations University's project, "Mobile Data, Environmental Extremes and Population" (MDEEP), seeks to use mobile phone call data, among other sources, to study the impact of cyclones and household responses in Bangladesh. A current project led by Maastricht University is focusing on the potential of using new technologies in responding to human displacement.¹¹ Georgetown University and the University of Sussex are jointly conducting research on the uses of Big Data to create an early warning system to prevent forced population displacement.¹²

“As previously explained, social media platforms such like microblogging site Twitter can be used to infer mobility patterns using the geographic information linked to users' posts.”

10 Flowminder Foundation official website: www.flowminder.org.

11 See "ICT and population displacement: potentials of new technologies in responding to human displacement" at www.maastrichtuniversity.nl/web/Main/Education/InnovationAndExcellence/ExcellenceProgrammes/PREMIUM/Projects1.htm.

12 See www.georgetown.edu/research/news/isim-forced-migration-warning-system.html for more information.

Table 1: Summary of the alternative uses of Big Data in studying migration trends

Big Data type	Description	Examples of past uses	Pitfalls/Challenges	Further potential
Call detail records (CDRs)	Digital records automatically generated and collected by mobile network operators every time a mobile phone call is made. CDRs include information about the time and duration of the call, the calling and receiving numbers and the approximate locations of the caller and the receiver.	Tracking post-disaster displacement Bengtsson et al. (2011) estimated the magnitude of and the trends in population movement out of Port-au-Prince following the 2010 Haiti earthquake through anonymized location records on 2.8 million SIMs that made at least one call in the pre- and post-earthquake period.	Privacy and ethical issues; personal security issues in conflict situations; selection bias and external validity; data collection and analysis; integration of innovative with traditional data; contextual knowledge.	Creation of early warning systems for forced migration and population displacements ¹³ ; migration forecasting; studying internal and temporary/circular migration patterns.
Other types of mobile phone data	Log of all mobile phone activity, including calls, text messages, money transfer and online purchases.	Identifying modalities and determinants of mobile money transfers Blumenstock et al. (2013) used data on interpersonal money transfers between 1.5 million subscribers to Rwanda's main telecommunication operator (date, time, value and anonymized identifiers for senders and receivers) to understand how money is used to share risk in post-disaster situations (in this case, the 2008 Rwanda earthquake) and identify the determinants of mobile money transfers.	Privacy and ethical issues; security issues in conflict situations; selection bias and external validity; data collection and analysis; integration of innovative with traditional data; contextual knowledge.	Studying patterns of remittance-sending activities.
IP addresses of website logins and sent e-mails	Geo-referenced data based on IP-addresses of repeated logins to the same website or e-mail-sending activity.	Estimating and predicting migration flows/rates State et al. (2013) estimate short- and medium-term migration flows from repeated logins to Yahoo! Services website of over 100 million anonymized users during a one-year period. They also built a model to infer global mobility patterns based on the likelihood of migrants moving. Zagheni and Weber (2012) estimated age- and gender-specific migration rates using IP-addresses to map the geographic locations from where e-mails were sent and the self-reported age and gender information of 43 million anonymized users.	Privacy and ethical issues; selection bias and external validity; data collection and analysis; accuracy of IP-based geographic information at the local and regional levels; integration of innovative with traditional data.	Migration tracking and forecasting; studying internal and temporary/circular migration patterns.

13 Visit www.georgetown.edu/research/news/isim-forced-migration-warning-system.html and www.ehs.unu.edu/mdeep, respectively, for more information about Georgetown University's and United Nations University's current research projects.

Big Data type	Description	Examples of past uses	Pitfalls/Challenges	Further potential
Geo-located social media data	<p>Geographic information (latitude and longitude) derived from users' geo-located posts on social media.</p>	<p>Inferring migration trends and comparing patterns of internal and international migration</p> <p>Zagheni et al. (2014) estimated migration trends in OECD countries using geo-located posts on Twitter of 15,000 users with an established minimum level of activity and for which they have consistent information over time, distinguishing between residents, who were tweeting from one country, and migrants, who were tweeting from different countries.</p>	<p>Privacy and ethical issues; selection bias and external validity; data collection and analysis; accuracy and truthfulness of self-provided information; integration of innovative with traditional data.</p>	<p>Migration tracking and forecasting; study of internal and temporary/circular migration patterns; creation of early warning system for forced migration and conflict-induced displacement.</p>
Social media content	<p>Data generated by users' activity on social media (posts, comments, pictures, etc.), partly publicly available.</p>	<p>Analysis of transnational networks and diaspora groups</p> <p>Nedelcu (2012) investigated the role of social media in the construction of transnational networks by considering Internet use by Romanian professionals in Toronto and their transnational families.</p> <p>Oiarzabal (2012) analysed 90 websites used by the Basque diaspora in 16 countries as part of his research on the influence of online activities on offline Basque diaspora members.</p> <p>Analysis of migration-related public discourse</p> <p>The project Unite Europe aims to analyse social media content generated by citizens to inform local integration policies.</p> <p>The Oxford Migration Observatory (2013) conducted a quantitative analysis of the language used by 20 British newspapers over two years to determine prevalent migration discourse in the United Kingdom.</p>	<p>Privacy and ethical issues; selection bias and external validity; accuracy and truthfulness of self-provided information and social media content; data collection and analysis of complex unstructured data.</p>	<p>Analysis of migrants' networks, political mobilization, community and identity formation; assessment of migrant integration in the host country; analysis of public attitudes towards immigrants and immigration; evaluation of public confidence in government immigration policies.</p>

Using Big Data to study migration: Issues and challenges

The use of Big Data in migration research comes with significant challenges. First, there are serious privacy, ethical and human rights issues related to use of data inadvertently generated by users of mobile devices and web-based platforms. Risks to individual rights to privacy can even threaten personal security in conflict situations (Letouzé, 2012). Public concerns over the use of Big Data for any purpose, including research, need to be identified and adequately addressed by policymakers, perhaps through the creation of a regulatory system setting out conditions and limits to access to and use of certain kinds of data.

There are also more general concerns associated with the selectivity and partiality of Big Data. This is sometimes referred to as the “Digital Divide,” which refers to the fact that populations’ access to and use of digital data tend to vary according to factors such as age, gender and financial condition. The availability of a potentially very large sample does not ensure that the sample will be representative, because data only refers to ICT users – a self-selected sample of the whole population. The risk of “selection bias” inherent in the use of ICT-generated data in migration research undermines the external validity of such studies – although the normally large sample size makes them more likely to be accurate. As a consequence, policies could be informed by partial evidence and non-representative data.

In specific cases, such as in the analysis of information gathered from social media, the impossibility of verifying the truthfulness of the information provided by users can also threaten the internal validity of empirical studies. Data collection and analysis constitute particularly challenging tasks, given the complexity, velocity and volume of Big Data. Further challenges are represented by the integration of innovative kinds of data with traditional ones, the insertion of new studies in existing migration theoretical frameworks (or the creation of new ones) and, more generally, the need to ensure that Big Data-derived knowledge is put in context for policymaking purposes. Finally, infrastructural challenges in data-sharing, management and security require more work and communication between researchers and policymakers (King, 2011).

Conclusions

It is widely recognized that policymakers lack timely and comparable data on migration. It is also the case that much migration data is scattered within and between countries and is not always used effectively by policymakers. Big Data alone is not the solution to these challenges, but the examples presented above suggest that Big Data can provide very useful evidence on emerging migration trends. The challenge remains how best to harness the potential of using Big Data. There are still many questions, including how we can make greater and more effective use of Big Data in the study of migratory trends within and across countries, and how Big Data can inform analyses of the impact of migration on the development of sending and receiving States, as well as on migrants’ well-being.

The opportunities presented by Big Data in migration research are matched by an equally significant number of issues and challenges which need to be tackled: individual privacy and ethical issues; personal security issues in conflict situations; selectivity and partiality of collected data; infrastructural challenges hindering data collection, aggregation and sharing among researchers for replication of studies; on the methodological side, selection bias and external validity issues; integration of innovative with traditional kinds of data for comprehensive analysis; contextual knowledge for informed policymaking; and, finally, accuracy and truthfulness of users’ self-provided information on social media. ■

References

ACAPS

- 2013 *Call Detail Records: The use of mobile phone data to track and predict population displacement in disasters*. Available from <http://www.gsdr.org/go/display&type=Document&id=5445>.

Ben-David, A.

- 2012 The Palestinian Diaspora on the Web: Between de-territorialization and re-territorialization. *Social Science Information*, 51: 459–474.

- Bengtsson L. et al.
2011 Improved response to disasters and outbreaks by tracking population movements with mobile phone network data: A post-earthquake geospatial study in Haiti. *PLOS Medicine*, 8(8): e1001083, doi:10.1371/journal.pmed.1001083.
- Blumenstock J., N. Eagle and M. Fafchamps
2013 *Motives for Mobile Phone-Based Giving: Evidence in the Aftermath of Natural Disasters*. Available from www.jblumenstock.com/files/papers/jblumenstock_mobilequakes.pdf.
- Dekker, R. and G. Engbersen
2013 How social media transform migrant networks and facilitate migration. *Global Networks*, doi: 10.1111/glob.12040.
- Eagle, N., Y. de Montjoye and L.M.A. Bettencourt
2009 Community computing: Comparisons between rural and urban societies using mobile phone data. *Computational Science and Engineering*, 4:144–150. Available from <http://citeseerx.ist.psu.edu/viewdoc/download?rep=rep1&type=pdf&doi=10.1.1.183.1723>.
- Hilbert, M.
2013 Big Data for Development: From Information to Knowledge Societies. Available from <http://ssrn.com/abstract=2205145>.
- King, G.
2011 Ensuring the data-rich future of the social sciences. *Science*, 331(11):719–721, doi: 10.1126/science.1197872.
- Kirkpatrick, R.
2013 Big data for development. *Big Data*, 1(1):3–4, doi:10.1089/big.2012.1502.
- Laney, D.
2001 3D Data Management: Controlling Volume, Velocity, and Variety. META Group File 949. Available from <http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf>.
- Latouzé, E.
2012 “Big Data for Development: What May Determine Success or Failure?” OECD Technology Foresight presentation, October 22, 2012, Paris. Available from www.oecd.org/mwg-internal/de5fs23hu73ds/progress?id=O7fS9xGecB.
- Lu, X., et al.
2013 Approaching the limit of predictability in human mobility. *Scientific Reports*, No. 3, Article 2923. Available from www.nature.com/srep/2013/131011/srep02923/full/srep02923.html.
- Mc Gregor, E. and M. Siegel
2013 Social Media and Migration Research. UNU-MERIT Working Paper Series #2013-068. Available from www.merit.unu.edu/publications/wppdf/2013/wp2013-068.pdf.
- Nedelcu, M.
2012 Migrants’ new transnational habitus: Rethinking migration through a cosmopolitan lens in the Digital Age. *Journal of Ethnic and Migration Studies*, 38(9):1339–1356.
- Oiarzabal, P.J.
2012 Diaspora Basques and online social networks: An analysis of users of Basque institutional diaspora groups on Facebook. *Journal of Ethnic and Migration Studies*, 38(9):1469–1485.
- Oxford Migration Observatory
2013 Migration in the News: Portrayal of immigrants, migrants, asylum-seekers and refugees in national British newspapers, 2010-2012. Report. Available from <http://www.migrationobservatory.ox.ac.uk/reports/migration-news>.
- Rinnawi, K.
2012 ‘Instant nationalism’ and the ‘Cyber Mufti’: The Arab Diaspora in Europe and transnational media. *Journal of Ethnic and Migration Studies*, 38(9):1451–1467.

State B., I. Weber and E. Zagheni

- 2013 “Studying international mobility through IP geo-location.” In: Proceedings of the sixth ACM international conference on Web search and data mining, pp. 265–274.

UN Global Pulse

- 2012 *Big Data for Development: Challenges and Opportunities?*, UN Global Pulse, New York Paper. Available from www.unglobalpulse.org/sites/default/files/BigDataforDevelopment-GlobalPulseMay2012.pdf.

UN High-level Panel

- 2013 Report on the Post-2015 Development Agenda 2013. Available from www.un.org/mwg-internal/de5fs23hu73ds/progress?id=pV59d7bffc.

Zagheni, E. et al.

- 2014 Inferring International and Internal Migration Patterns from Twitter Data. *WWW 2014 Companion*, 7 – 11 April 2014.

Zagheni, E. and I. Weber

- 2012 “You are where you e-mail: Using e-mail data to estimate international migration rates.” In: ACM Web Science Conference proceedings, 25 June 2012.