MAX PLANCK INSTITUTE
FOR DEMOGRAPHIC
RESEARCH

MAX-PLANCK-INSTITUT
FÜR DEMOGRAFISCHE
FORSCHUNG

Web science

You are where you e-mail: Global migration trends discovered in e-mail data

U.S. emigration made visible for the first time.

Rostock, Germany. For the first time comparable migration data is available for almost every country of the world. To date, records were incompatible between nations and especially by gender and age, nonexistent. Emilio Zagheni from the Max Planck Institute for Demographic Research (MPIDR) in Rostock, Germany, for the first time provides a rich migration database by compiling the global flow of millions of e-mails.

"Where estimates of demographic flows exist, they are often outdated and largely inconsistent," says MPIDR researcher Emilio Zagheni. Official records are difficult to use for various reasons. Emigrants tend not to register after they move to a new country or do so very late. There is also no clear agreement between nations on how to actually define a migrant.
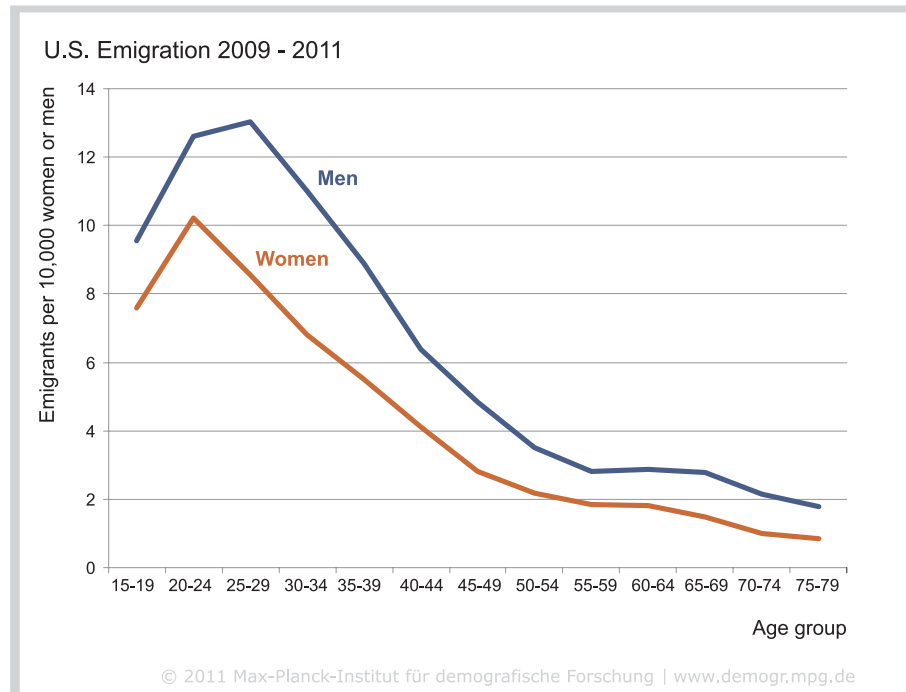
Official migration data is outdated and inconsistent

"Global internet data does not have these drawbacks," says Zagheni. "You are where you e-mail." Together with Ingmar Weber from Yahoo! Research he traced e-mails sent from Yahoo! accounts around the world to infer the residence of its sender. Every device which sends e-mail can be located at least at the country level by an internationally standardized code, the so-called IP address. Zagheni and Weber analyzed the countries derived from IP addresses for a set of messages sent by 43 million anonymous Yahoo! account holders between September 2009 and June 2011.

In addition to the date and geographical origin of each message they compiled the self-reported birthday and gender of the sender. When a person started sending e-mail from a new location permanently, it was assumed that he or she had changed residence. This way they were able to calculate rates of migration from and to almost every country in the world. Only anonymized data was used, so identifying individuals was impossible and no information about the recipients, the subject, or content of a message was accessed. The findings have now been published in the ACM Web Science Conference Proceedings.

The results not only are a proof of concept. They also reveal international migration characteristics never seen before. For the USA, Zagheni and Weber were able to produce the

MAX-PLANCK-GESELLSCHAFT

U.S. Emigration 2009 - 2011

Emigrants per 10,000 women or men

Men

Women

Age group

© 2011 Max-Planck-Institut für demografische Forschung | www.demogr.mpg.de

U.S. Emigration unveiled: by analyzing millions of e-mails the first consistent figure of those emigrating from the USA was made possible. The curves show those who sent most of their e-mails from the U.S. between September 2009 to June 2010 but consistently wrote the majority of their messages from abroad between July 2010 and June 2011.

first curve of emigration by age and sex ever. "In the U.S. many statistics are collected about people who move into the country, but there is no system that keeps track of people who move out," says Emilio Zagheni.

The potential of the e-mail statistics goes far beyond calculating gross country profiles. For instance, the researchers also looked into Mexico-US cross-border mobility. The data reveals how strongly both countries are demographically integrated: most people who moved from Mexico to the United States either spent time in the USA before emigrating north, or went back to visit Mexico soon after moving to the United States. Those in their 30s have the highest rate of mobility across the Mexico-US border, while the least mobile are those 50 and older.

Only the tip of the iceberg

The strength of Zagheni's and Weber's migration data comes not only from the vast number of e-mails available, but also from a mathematical model they set up to adjust for typical shortcomings of e-mail statistics: those who send e-mail are not representative of the entire population. Some groups, like the elderly, use e-mail less or not at all and are thus underrepresented. But the researchers managed to calculate adjustment factors for such groups by gauging their e-mail data against migration numbers from European countries, where official data is fairly reliable.

"What we addressed so far is only the tip of the iceberg," says Emilio Zagheni. With further fine-tuning of the adjustment factors and mining more digital data like twitter messages, more difficult questions could be tackled. For instance one could keep track of the short and long-term mobility

patterns before and after a crisis like that of the Japanese Fukushima reactors. Unquestionably, digital records give demographers the chance to gain a more accurate picture of population dynamics in regions they can so far only guess about, says Zagheni. "This research has the most potential in developing countries, where the Internet spreads much faster than registration programs develop."

About the MPIDR

The Max Planck Institute for Demographic Research in Rostock (MPIDR) investigates the structure and dynamics of populations. It focuses on issues of political relevance such as demographic change, aging, fertility, the redistribution of work over the course of life, as well as aspects of evolutionary biology and medicine. The MPIDR is one of the largest demographic research bodies in Europe and one of the worldwide leaders in the field. It is part of the Max Planck Society, the internationally renowned German Research Society.

www.demogr.mpg.de

Contact

Emilio Zagheni – MPIDR author of the article (speaks English, French and Italian)
PHONE   +49 381 2081 – 227
E-MAIL   Zagheni@demogr.mpg.de

Silvia Leek – MPIDR Press department
PHONE   +49 381 2081 – 143
E-MAIL   leek@demogr.mpg.de

This press release and high resolution graphics are available for download at:
www.demogr.mpg.de/en/news_press/press_releases_1916/you_are_where_you_e_mail_global_migration_trends_discovered_in_e-mail_data_2939.htm

Original article: Emilio Zagheni, Ingmar Weber: You are where you E-mail: Using E-mail Data to Estimate International Migration Rates, ACM Web Science Conference Proceedings.
http://www.demogr.mpg.de/publications%5Cfiles%5C4598_1340471188_1_Zagheni&Weber_Websci12.pdf

MAX-PLANCK-GESELLSCHAFT