USING ADVERTISING AUDIENCE ESTIMATES TO IMPROVE GLOBAL DEVELOPMENT STATISTICS

Ingmar Weber[1], Ridhi Kashyap[2], Emilio Zagheni[3]
[1]Qatar Computing Research Institute, [2]University of Oxford, [3]Max-Planck Institute for Demographic Research

Abstract – *The United Nations Sustainable Development Goals (SDGs) are a key instrument in setting the agenda around global development until 2030. These goals come with a set of 232 indicators against which countries should monitor their progress with respect to the SDGs. Existing data sources to measure progress on the SDGs and global population trends however are often (i) outdated, (ii) lacking international comparability, (iii) lacking appropriate disaggregation, or (iv) missing completely. These problems are often especially acute among less developed countries. In this paper we describe how anonymous, aggregate data from the online advertising platforms of Facebook, LinkedIn and other services can be used in combination with existing data sources to improve global development statistics. We illustrate the process of using and validating such non-representative data through two case studies looking at (i) Internet access gender gaps, and (ii) international migration statistics.*

Keywords – Facebook, global development statistics, online advertising, SDG monitoring.

1. INTRODUCTION

The Sustainable Development Goals (SDGs) are a set of 17 global goals set by the United Nations in 2015 to serve as guiding principles to set the course for a better future. The goals include aspects of economic, social and environmental development. To give the goals a practical meaning, they come with a total of 169 targets, which, in turn come with 232 indicators to serve as a global "benchmark" to monitor progress on the SDGs.[1]

However, even before the introduction of these indicators that countries are expected to monitor, many countries are lacking even the most basic development statistics. In a 2014 publication, the World Bank reports that 16 countries did not participate in the 2010 census round, leading to census data which is at least 10 years old [2]. Even when data is available, data quality is often poor, which basically renders it useless [3]. Issues of limited data availability and poor data quality are particularly pronounced in less developed countries, which at the same time have a lot to gain from proper data use [4].

To address the challenges related to "data poverty" and to potentially leapfrog the development of traditional data acquisition infrastructure several researchers and organizations have explored the use of "Big Data" approaches for improving development statistics and spark a data revolution [5].

The sources of big data most commonly used include satellite data [6-8], call detail records [9,10], web search data [11,12] and public social media posts [13-15].

Here we describe a new data source: anonymous online advertising audience estimates provided by companies such as Facebook, LinkedIn, Google, Twitter and others. In a nutshell, companies providing targeted online advertising support potential advertisers in their campaign planning with information on how many of their users match certain targeting criteria. This serves as a kind of real-time census over the user base of the considered services, providing aggregate answers to questions such as "how many male Facebook users from Germany currently live in Geneva?".

[1] However, not all of the indicators already have agreed-upon measurement criteria [1].

In this article, we describe this type of data, which has complementary strengths and weaknesses to both traditional data and other big data sources, in detail and show how it can be used to improve global development statistics.

In the next sections, we first describe this rarely used data source. Next, we summarize case studies that use this type of data to (i) nowcast Internet access gender gaps around the globe, and (ii) improve international migration statistics. After presenting these encouraging findings, we then discuss important limitations and challenges, including the issues of selection bias and privacy concerns.

2. ONLINE ADVERTISING AUDIENCE ESTIMATES

Online advertising is the main revenue for companies such as Facebook, Twitter and Google. To support potential advertisers in planning budgeting requirements for an advertising campaign, all of these platforms provide services to estimate the potential audience reach for a given set of targeting criteria. As an example, Fig. 1 shows a screenshot of Facebook's Ads Manager[2]. The screenshot shows on the right the "Potential reach", 1,100 people, when targeting Facebook users living in Geneva who are men and expats from Germany. Similar interfaces are provided by the advertising platforms of Twitter, LinkedIn and Snapchat among others, though the targeting criteria supported differ. Google's advertising interface differs in that it shows estimates for *ad impressions*, rather than users, where more active users will create more ad impressions.[3] Most platforms provide an application programming interface (API) to facilitate collecting audience reach estimates programmatically.

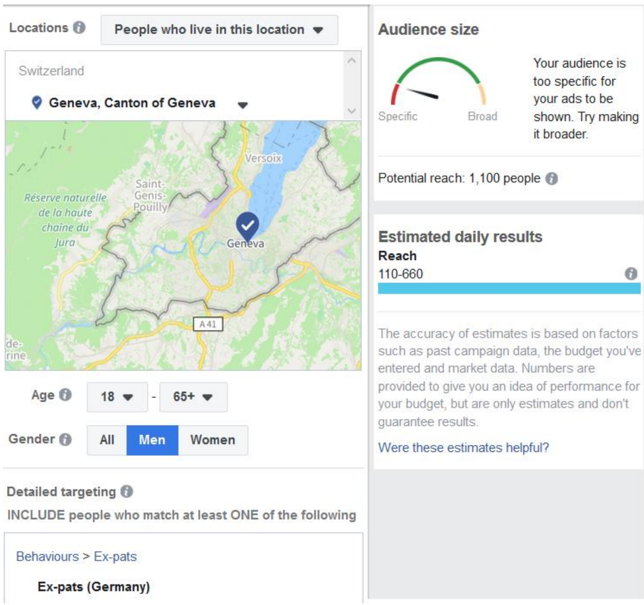


Fig. 1 –Screenshot of Facebook's Ads Manager. The potential reach (1,100) is an estimate of the number of monthly active users (MAUs) living in Geneva who are men and expats from Germany. The estimate of MAUs can be obtained programmatically via Facebook's Graph API. The screenshot has been edited to fit into a single column.

These estimates are available in near real time and can be disaggregated by age and gender, as well as other characteristics. The spatial resolution differs between platforms and the country targeted and ranges from state level to ZIP codes at the sub-city level. However, to avoid re-identification of individual users, platforms apply techniques related to k-anonymity [16]. This means that the smallest audience estimates that are presented have to include 100-1000 users, depending on the platform. Furthermore, the estimates are rounded to further limit potential re-identification attempts.

[2] Any reader with a Facebook account can access this interface following the steps at https://www.facebook.com/adsmanager/creation until reaching the audience selection step.

[3] See https://support.google.com/google-ads/answer/6320?hl=en for details on the definition of an "impression".

In the following sections, we describe two case studies that both use online advertising audience estimates from Facebook. However, the same general approach has been applied using LinkedIn to obtain statistics on gender skill gaps in the US [17].

3. CASE STUDY 1: INTERNET ACCESS GENDER GAPS

Achieving gender equality by 2030 is Goal #5 of the Sustainable Development Goals. One of the targets for this goal is to "enhance the use of enabling technology, in particular information and communications technology (ICT), to promote the empowerment of women"[4]. Corresponding indicators of interest relate to gender-disaggregated statistics on Internet usage and mobile phone ownership. The adoption of these SDG targets related to gender and ICT use acknowledges that even as the use of ICTs has rapidly expanded, significant gender gaps in access to these technologies persist. The International Telecommunication Union (ITU), the UN's specialized agency for ICTs, estimates that some 200 million fewer women are online compared with men, with Internet use gender gaps being significantly greater in less developed countries [18]. However, the paucity and irregular production of gender-disaggregated data on Internet and mobile phone access, particularly in less developed country contexts, present significant challenges towards monitoring the progress towards these goals.

Currently, the best available data source for gender-disaggregated statistics on Internet use is compiled by the ITU, based on surveys fielded by the national statistical agencies of ITU member states [19]. Although year-to-year availability of this data for different countries vary, estimates for at least one year in the period 2011 to 2015 was available for 84 countries (out of 218 in the world). The most limited coverage was for low-income (data available for 2 out of 31 countries) and lower-middle income countries (data available for 11 out of 53 countries). Data availability for countries in sub-Saharan Africa (4 out of 48) and South Asia (1 out of 8) are especially limited. Statistics on mobile phone ownership by gender are even more limited than that on Internet access. Data from the GSMA, a trade body representing the interests of mobile phone operators worldwide, on gender gaps in mobile phone ownership were available for only 22 countries. In light of these shortcomings of existing data sources, the Data2X initiative at the UN Foundation identified gender-disaggregated data for access to Internet and mobile phones was identified as one of the most pressing gender data gaps.[5]

In this data sparse context, data from Facebook's online advertising audience estimates, which can be queried for aggregate statistics on the number of users of the platform by gender, age and device type, have been leveraged to measure and 'nowcast' gender gaps in Internet and mobile access [19]. Nowcasting refers to the idea of 'predicting the present', especially in the case of indicators where real-time information is useful but where there is likely to be a significant delay or lag in producing it [20].

The Facebook data was used to generate a "Facebook Gender Gap Index" (FB GGI), an indicator of the number of female to male Facebook users in a given country. For example, in Belgium we observed an equal number of 3.1M female and male monthly active Facebook users, whereas for India there were 40M female and 133M male monthly active Facebook users as of October 2018.

While the FB GGI reflects gender gaps in Facebook use and not Internet access per se, in practice these Facebook indicators are highly correlated with officially reported statistics on Internet (from the ITU) and mobile phone gender gaps (from the GSMA) for the countries for which this data is available. This suggests that these online, Facebook-derived indicators are useful measures that can be used to inform predictive regression models that are validated against official statistics. Furthermore, the FB GGI appears to capture gender inequalities in Internet access most effectively in less developed countries where access to the Internet is most unequal by gender.

[4] See the description of goals, targets and indicators at https://unstats.un.org/sdgs/indicators/indicators-list/.

[5] See https://www.data2x.org/what-is-gender-data/gender-data-gaps/.

The predictive power of Facebook indicators can be further enhanced when combined with other offline, development-related measures associated with gender inequalities in Internet and mobile access (e.g a country's GDP per capita, the Global Gender Gap Report (GGGR) measures of gender gaps in literacy or economy). When comparing the performance of regression models predicting gender gaps in Internet use using online Facebook indicators with those using 1) offline variables only, and 2) a combination of online Facebook variables and offline variables, models using Facebook data did better than those using offline indicators alone, and those combining online-offline indicators did the best.

To quantify the prediction quality of different models, Table 1 reports three different evaluation metrics, namely (i) Adjusted R-squared, (ii) mean absolute error, and (iii) symmetric mean absolute percentage error (SMAPE). Adjusted R-square is a measure of model fit that quantifies the percentage of the variance, i.e. variability, in the ground truth data that can be "explained", i.e. modeled, using a linear combination of features. The mean absolute error reports the average absolute prediction error. The SMAPE normalizes the absolute prediction error by the average of the true and predicted values, i.e. SMAPE = 2*|true - predicted|/|true + predicted|.

Table 1 highlights how all measures of predictive fit are best for the online-offline model followed by the online and then offline models. A significant strength of the online model, which uses a single Facebook indicator only, is that it enables prediction for the most number of countries, with the biggest gains in coverage made for less developed countries. Fig. 2 visually shows the coverage gain compared to ITU data, in particular for sub-Saharan Africa.[6]

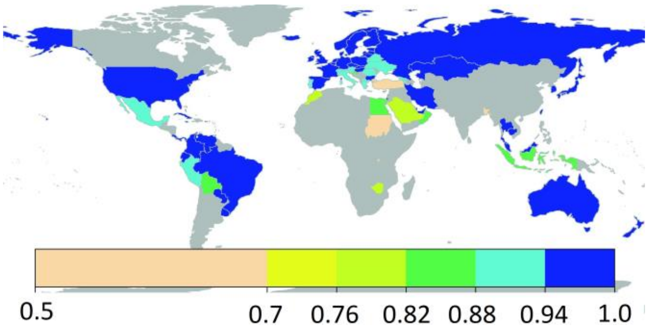
Table 1 – Summary of results for three regression models predicting ITU Internet gender gap using (i) a single online Facebook variable; (ii) online and offline variables; (iii) offline variables. See [19] for additional details and statistical measures.

	Online Model	**Onl.-Offl. Model**	**Offline Model**
Intercept	.933***	.932***	.933**
FB GGI	.071***	.093***	
log(GDP per capita)		.018*	
GGGR - Literacy		-.018	
GGGR - Education		-.019	
Internet Penetration			.040***
GGGR - Tertiary Educ.			.032
GGGR - Economy			.043**
GGGR - Score			-.024
Adjusted R^2	.691	.791	.615
Mean Abs. Error	0.0325	0.0288	0.037
SMAPE	3.92%	3.90%	4.97%
# predicted countries	152	127	132

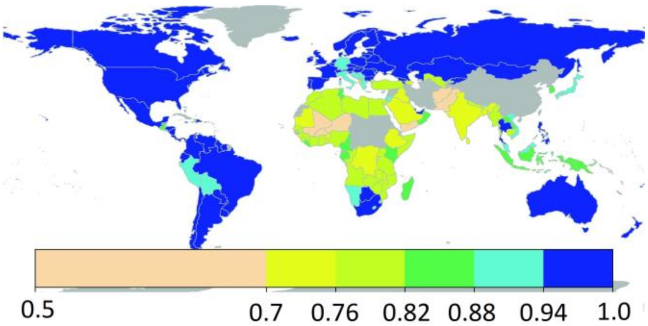
***$p < 0.001$, ** $p < 0.01$, * $p < 0.05$.

In addition to their real-time availability, another advantage offered by the Facebook data source is the finer geographical resolution for which this data is available. Facebook's advertising audience estimates have been used to generate measures of subnational digital gender inequality in India and this approach can be extended to other countries [21]. Gender gaps in Facebook may also serve as a measure for other aspects of gender inequality more generally, including domains such as education, health and economic opportunity, as indicated by the correlation between Facebook gender gap measures and those of the World Economic Forum's gender gap indicators [22]. Facebook gender gap measures may also help predict changes in economic gender inequality, as suggested by findings in [22] that countries with a lower Facebook gender gap in 2015 saw an overall increase in economic gender equality in 2016.

[6] Recent and automatically updated Internet access gender gaps predictions are available at https://www.digitalgendergaps.org/.



(a) Internet access gender gaps according to 2015 ITU data.



(b) Internet access gender gaps according to 2017 Facebook data.

Fig. 2 – Two world maps showing the ratio of (percentage of women with Internet access)/(percentage of men with Internet access) on a per-country basis. ITU data from 2015(top) is compared to model predictions of the online model (see Table 1) using Facebook data from 2017 (bottom). The model manages to largely reproduce ITU ground truth data while substantially improving global coverage. See [19] for additional details.

4. CASE STUDY 2: INTERNATIONAL MIGRATION STATISTICS

Migration is not included directly among the Sustainable Development Goals. However, migration is one of the driving forces behind demographic changes among the globe and SDG #10, *Reduce Inequalities*, includes the target to "facilitate orderly, safe, regular and responsible migration and mobility of people, including through the implementation of planned and well-managed migration policies". The implementation and monitoring of such policies is tightly linked to the availability of accurate and timely data on migration, as emphasized by the Global Compact for Safe and Orderly Migration[7], an intergovernmental agreement negotiated with the support of the United Nations, and designed to cover migration in a comprehensive way.

Data on migration, especially flows, are often unavailable for a number of countries. When the data does exist, it is typically outdated or inconsistent across countries. This is because some countries rely on survey data, others on census data, registration systems or various other administrative sources. Such traditional data has various degrees of under-reporting issues and the statistics obtained from the data is often published with a substantial delay.

In this context, data from Facebook's advertisement platform offers a new source of information that is timely and that can be considered as a continuous census of a large portion of the world population. The main challenge is that Facebook users are not necessarily representative of the underlying populations. In order to understand how the selection bias changes across demographic groups, Zagheni et al. [23] used the following linear regression model:

$$\log(\text{ACS foreign-born pop}^z_{ij}) = \beta_0 + \beta_1 \log(\text{Facebook expats}^z_{ij}) + \beta_2\, I(\text{origin } 1) + \dots + \beta_{30}\, I(\text{origin } 30) + \beta_{31}\, I(\text{age group } 1) + \beta_{38}\, I(\text{age group } 8) + \varepsilon^z_{ij} \quad (1)$$

where $\text{ACS foreign-born pop}^z_i$ is the number of people in the age-sex group z born in country i and living in US state j, according to the American Community Survey. $\text{Facebook expats}^z_{ij}$ is the respective quantity from Facebook data. The indicator variables add controls for different age groups and countries of origin. The mean absolute percentage error (MAPE) for the 'naive' model that does not include indicator variables, is on 56 percent. The MAPE for the age-origin model, that explicitly models biases related to differences in countries of origin or in age groups, is significantly lower at 37 percent. This indicates that biases show regularities along the dimensions of age and country of origin. These regularities can be modeled and the biases can be corrected, at least partially.

Facebook data offers timely, but possibly biased data. Traditional surveys offer time series of representative statistics, but often outdated or from relatively small samples. Combining the two types of sources holds promise for generating timely and reliable migration statistics. [24]

[7] Details at https://www.iom.int/global-compact-migration .

Note that our analysis also shows the potential for obtaining subnational estimates for stocks of migrants. While countries like the United States have good migration stocks data at both the national and the subnational level, other countries, especially less developed ones, might only have reliable data at the national level and could potentially benefit from our approach for creating spatially disaggregated estimates.

While our current research focuses on improving estimates for stocks of migrants at the traditional time scale of one year, similar approaches also hold promise for monitoring short-term flows of migrants. As an example, recent data on Facebook users in Spain [25] showed a surprisingly high number of likely migrants from Venezuela, higher than the numbers in the most recent official statistics, but very plausible given the current crisis in Venezuela.

5. LIMITATIONS AND CHALLENGES

When deriving any type of insights from online data the question of data bias naturally comes up. Here we discuss two of the most important types of bias, as well as other limitations and challenges related to using online advertising audience estimates for monitoring global development.

Arguably the most frequently cited bias is the selection bias where, say, insights derived from online data are more likely to represent the behavior and the needs of the well-off than the people most in need as the latter are less likely to contribute to the data in the first place. Irrespective of economic status, people not on social media would not contribute to the generation of the data and hence the analysis. This can lead to distorted results such as the finding that in countries with fewer women than men on Google+, those women who are online have a higher online status than the men [26].

However, this type of bias and data distortion is not necessarily problematic. Firstly, it is often exactly *the missing data that is the signal*. For example, in our research on gender gaps it is the fact that women are *not* found in the data at the same rate as men that provides a signal on gender inequalities. Secondly, approaches using supervised machine learning such as regression models treat the (biased) data merely as a signal to predict a particular quantity of interest, e.g. stocks of migrants [27]. As long as the signal has high predictive power, it has potential value for the task. For such approaches, selection bias is only a challenge when it is non-systematic, e.g. when different countries exhibit different mechanisms underlying the selection bias and where these mechanisms cannot be understood and modeled, and hence corrected for, through available data. Nevertheless, machine-learning models when applied to social and demographic outcomes can produce predictions that are inconsistent with prior domain knowledge. For example, in the case of demographic rates, certain empirical regularities by age and gender are well documented. Age-specific migration rates tend to peak in early adulthood. Combining substantive and theoretical knowledge about the drivers of migration, or patterns of gender equality, with machine-learning based approaches is a challenge that is important to address. A further challenge also emerges when several, different estimates for the quantity to be predicted, or in other words, different ground truth measures, are available, which come from different surveys or other data sources. In this case, applying methods to combine and generate estimates that account for different types of error and uncertainty across different data sources is an important area for future research.

Another more fundamental type of bias affecting a lot of data for development research is the so-called “streetlight effect” where a particular issue is studied not because it is the most pressing issue, but because data is readily available. This type of bias certainly does extend to advertising audience estimates which, say, are unlikely to provide important insights for improving statistics on Goal #14, life below water. Our work does by no means advocate stopping the collection of traditional data or stop collecting data on pressing issues with poor data availability.

More fundamentally, the streetlight effect is also related to the problem that not everything that counts is countable.[8] Though we certainly believe that better data can contribute to better solutions,

[8] Also see the article in https://www.theguardian.com/global-development-professionals-network/2014/dec/17/data-revolution-limitations-in-images for a discussion of general shortcomings of the data revolution vision.

data alone is not the same as knowledge or truth. Ultimately, data cannot replace moral judgement on which actions to take.

Another important challenge is the fact that the data provided to advertisers comes from proprietary black boxes without an easy way for academics or others to audit aspects related to data quality. Whereas some user attributes such as age and gender are most likely derived from self-declared information, other attributes such as Facebook's "Ex-pats (Germany)" are based on a proprietary inference algorithm with an unknown accuracy. However, as advertising is the main revenue source for Facebook and others, one would assume that they invest appropriate resources in satisfying the advertisers' needs, i.e. accurately inferring user attributes. As mentioned previously, the typical use of the audience estimates is also *not* as a census, i.e. as an exhaustive count, but as an input signal for a regression task. In this latter setting one might still be able to obtain accurate predictions for variables of interest, even without full transparency on the precise specification of the input variables.

Last but not least, the issue of user privacy is important to consider, in particular in light of the recent Cambridge Analytica scandal. Other researchers have shown that previous versions of Facebook's advertising platform leaked personally identifiable information [28]. This leakage was possible through an exploitation of the audience estimates for so-called "custom audiences" that involve targeting a particular set of users, identified by name, email or phone number. In its current version, audience estimates can only be obtained for anonymous, aggregate user groups such as male Germans living in Geneva. As such, most of the privacy concerns are similar to those surrounding population estimates for census tracts. At the same time, due to the possibility of (i) dynamically targeting different sub-populations, and (ii) doing so in a repeated manner, it cannot be ruled out that despite the aggregation and rounding of the returned audience estimates a sufficiently skilled attacker could abuse this data source and obtain attributes for individual users. However, none of the data collected in any of the described or proposed work contains individual level information and could not be used to obtain such information.

Apart from privacy concerns for individual users, there is the harder to address issue of group-level profiling. For example, one could potentially abuse the platform to create density maps of users with Facebook interests such as "Jewish prayer" or "gay bar". Using interests strongly correlated with protected attributes such as race can also be used to exclude certain parts of the population to view, say, ads related to housing, employment or financial services [29]. For our own research we never use the "custom audiences" and only perform secondary analysis of anonymous and aggregated data.

6. CONCLUSIONS

The case studies above demonstrate the value that online advertising audience estimates hold for complementing existing traditional data sources such as surveys and censuses for improving global development statistics. Note that all the data sources described here are publicly available free of cost, which helps to reduce latency for near real-time estimates. Furthermore, this helps to democratize data access as, traditionally, personal contacts at big companies would be required for accessing similar types of data.

We do not see big data as a silver bullet to overcome the challenges of significant data gaps. However, we do believe that when used (i) in combination with, not in replacement of, existing data sources, and (ii) in an ethical and responsible manner, then online advertising audience estimates can help to fill data gaps on important topics such as gender gaps and international migration. Better data on these topics will hopefully support better policy making and lead to better resource allocation.

ACKNOWLEDGEMENT

Ridhi Kashyap and Ingmar Weber are supported by a Data2x "Big Data for Gender Challenge" grant. Details at http://data2x.org/big-data-challenge-awards/#digital.

REFERENCES

[1] Barbara Adams, Karen Judd: The Ups and Downs of Tiers: Measuring SDG Progress. Global Policy Watch, 2018. https://www.globalpolicywatch.org/blog/2018/04/26/tiers-measuring-sdg-progress/

[2] World Bank: A Measured Approach to Ending Poverty and Boosting Shared Prosperity: Concepts, Data, and the Twin Goals. Policy Research Reports, 2014.

[3] Morten Jerven: Poor Numbers: How We Are Misled by African Development Statistics and What to Do about It. Cornell University Press, 2013.

[4] Kofi Annan: Data can help to end malnutrition across Africa. Nature 555, 7 (2018).

[5] Independent Expert Advisory Group on a Data Revolution for Sustainable Development: A world that counts. UN Report, 2014, http://www.undatarevolution.org/wp-content/uploads/2014/11/A-World-That-Counts.pdf.

[6] Tobias G. Tiecke, Xianming Liu, Amy Zhang, Andreas Gros, Nan Li, Gregory Yetman, Talip Kilic, Siobhan Murray, Brian Blankespoor, Espen B. Prydz, Hai-Anh H. Dang: Mapping the world population one building at a time. ArXiv, 1712.05839v1, 2017.

[7] J. Vernon Henderson, Adam Storeygard, David N. Weil: Measuring economic growth from outer space. American Economic Review, vol. 102, no. 2, pp. 994-1028, 2012.

[8] Tom Bundervoet, Laban Maiyo, Apurva Sanghi: Bright Lights, Big Cities Measuring National and Subnational Economic Growth in Africa from Outer Space, with an Application to Kenya and Rwanda. Policy Research Working Paper 7461, 2015.

[9] Joshua Blumenstock, Gabriel Cadamuro, Robert On: Predicting poverty and wealth from mobile phone metadata. Science Vol. 350, Issue 6264, pp. 1073-1076, 2015.

[10] Jessica E. Steele, Pål Roe Sundsøy, Carla Pezzulo, Victor A. Alegana, Tomas J. Bird, Joshua Blumenstock, Johannes Bjelland, Kenth Engø-Monsen, Yves-Alexandre de Montjoye, Asif M. Iqbal, Khandakar N. Hadiuzzaman, Xin Lu, Erik Wetter, Andrew J. Tatem, Linus Bengtsson: Mapping poverty using mobile phone and satellite data. Journal of the Royal Society, Volume 14, issue 127, 2017.

[11] Hyunyoung Choic, Hal Varian: Predicting the Present with Google Trends. Economic Record, vol. 88, Iss. 1, pp. 2-9, 2012.

[12] Nikolaos Askitas, Klaus F. Zimmermann: Google Econometrics and Unemployment Forecasting. Applied Economics Quarterly, Vol. 55, No. 2, pp. 107-120, 2009.

[13] Giuliano Resce, Diana Maynard: What matters most to people around the world? Retrieving Better Life Index priorities on Twitter. Technological Forecasting and Social Change, 2018.

[14] Dolan Antenucci, Michael Cafarella, Margaret Levenstein, Christopher Ré, Matthew D. Shapiro: Using Social Media to Measure Labor Market Flows. NBER Working Paper No. 20010, 2014.

[15] Alejandro Llorente, Manuel Garcia-Herranz, Manuel Cebrian, Esteban Moro: Social Media Fingerprints of Unemployment. PLOS ONE 10(5): e0128692, 2015.

[16] Latanya Sweeney: k-Anonymity: A model for protecting privacy. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, Vol. 10, No. 05, pp. 557-570, 2002.

[17] Karri Haranko, Kiran Garimella, Emilio Zagheni, Ingmar Weber: Professional Gender Gaps Across US Cities. ICWSM, p604-607, 2018.

[18] Broadband Commission. 2013. Doubling Digital Opportunities - Enhancing the Inclusion of Women and Girls in the Information Society. Technical Report. UNESCO;ITU. http://www.broadbandcommission.org/Documents/working-groups/bb-doubling-digital-2013.pdf.

[19] Masoomali Fatehkia, Ridhi Kashyap, Ingmar Weber: Using Facebook Ad Data to Track the Global Digital Gender Gap. World Development, vol. 107, p189-209, 2018.

[20] Enrico di Bella, Lucia Leporatti, and Filomena Maggino. "Big data and social indicators: Actual trends and new perspectives." *Social Indicators Research* 135, no. 3 (2018): 869-878.

[21] Yelena Mejova, Harsh Rajiv Gandhi, Tejas Jivanbhai Rafaliya, Mayank Rameshbhai Sitapara, Ridhi Kashyap, and Ingmar Weber. "Measuring Subnational Digital Gender Inequality in India through Gender Gaps in Facebook Use." In *Proceedings of the 1st ACM SIGCAS Conference on Computing and Sustainable Societies*, p. 43. ACM, 2018.

[22] David Garcia, Yonas Mitike Kassa, Angel Cuevas, Manuel Cebrian, Esteban Moro, Iyad Rahwan, Ruben Cuevas: Analyzing gender inequality through large-scale Facebook advertising data. Proceedings of the National Academy of Sciences, 2018.

[23] Emilio Zagheni, Ingmar Weber, Krishna Gummadi: Leveraging Facebook's Advertising Platform to Monitor Stocks of Migrants. Population and Development Review, vol. 43, Issue 4, p721-734, 2017.

[24] Emilio Zagheni, Kivan Polimis, Monica Alexander, Ingmar Weber, and Francesco C. Billari. "Combining Social Media Data and Traditional Surveys to Nowcast Migration Stocks." In Annual Meeting of the Population Association of America 2018.

[25] Spyridon Spyratos, Michele Vespe, Fabrizio Natale, Ingmar Weber, Emilio Zagheni, Marzia Rango: Migration Data using Social Media: a European Perspective. EC Technical Report, 2018. http://publications.jrc.ec.europa.eu/repository/handle/JRC112310

[26] Gabriel Magno, Ingmar Weber: International Gender Differences and Gaps in Online Social Networks. SocInfo 2014: 121-138.

[27] Emilio Zagheni, Ingmar Weber: Demographic Research with Non-Representative Internet Data. International Journal of Manpower, vol. 36 (1), p13-25, 2015.

[28] Giridhari Venkatadri, Athanasios Andreou, Yabing Liu, Alan Mislove, Krishna P. Gummadi, Patrick Loiseau, Oana Goga: Privacy Risks with Facebook's PII-based Targeting: Auditing a Data Broker's Advertising Interface. Symposium on Security and Privacy, pp: 221-239, 2018.

[29] Till Speicher, Muhammad Ali, Giridhari Vekatadri, Filipe Nunes Ribeiro, George Arvanitakis, Fabricio Benevenuto, Krishna Gummadi: Potential for discrimination in online targeted advertising. PMLR 81:5-19, 2018