# **Using big data** for insights into the **gender digital divide for girls:**

# A discussion paper

unicef
for every child

# **Using big data** for insights into the **gender digital divide for girls:**

# A discussion paper

Alexandra Tyers-Chowdhury
and Gerda Binder
UNICEF

Ridhi Kashyap
and Mariana De Araujo Cunha
University of Oxford

Reham Al Tamime
University of Southampton

Ingmar Weber
Qatar Computing Research Institute

# Introduction

## the gender digital divide for adult women

Digital technology is rapidly evolving and increasingly indispensable in our lives. This is true not only for adults, but also for children and adolescents. For young people, digital technology provides tools to engage in entrepreneurship, network with peers, and access critical health, career and financial information. For women and girls, digital adoption and use offers an opportunity to overcome hurdles they may face in the physical world. It can offer skills needed to excel in education and work, as well as increased access to the digital jobs of the future.[1]

**Unfortunately, a large gender divide disfavours women in digital access and use.**

In recent years, a large increase in primary data has widened understanding of this gender digital divide for the adult women cohort. Key statistics and methodologies are shown in Box 1. However, the gender digital divide community of knowledge is still overwhelmingly focused on women over the age of 18.

**There is much less data and evidence on the girl cohort, leaving a significant knowledge gap regarding the digital reality and shifts for today's generation of girls.**

1 Tyers, A and Banyan Global, 2020

Box 1

## The gender digital divide for adult women

**More than 50% of the world's women are offline.[2]**

This percentage is even larger in low-and middle-income countries (LMICs), where the internet penetration rate for adult women is 41%, compared to 53% for men.[3]

**Roughly 393 million adult women in LMICs do not own mobile phones. Globally, women are 8% less likely to own a mobile phone than men.[4]**

There are also stark regional differences. For instance, the gender gap in mobile ownership is much larger in South Asia (23%) and sub-Saharan Africa (13%).[5] There are also differences in device ownership: women are more likely than men to borrow or share mobile phones (often within a household or from a male family member) and are rarely the primary owners of a device.[6] Women are more likely to have simpler phones that do not support mobile internet use, and women are 20% less likely than men to own a smartphone.[7]

This gender gap in digital access is accompanied by a gender gap in meaningful digital use. Women tend to use mobiles and the internet differently than men. For example, they often own less expensive and sophisticated handsets, use a smaller range of digital services (often primarily voice and SMS), use digital services less frequently and less intensively, and use the internet less frequently and for fewer things.[8] These disparities in usage limit women's access to the full range of opportunities offered by digital connectivity.[9]

2 ITU, 2019
3 ITU, 2019
4 GSMA Connected Women, 2020
5 GSMA Connected Women, 2020
6 EQUALS, 2019
7 GSMA Connected Women, 2020
8 Web Foundation, 2015; Web Foundation, 2016; Web Foundation, 2020; LIRNEasia, 2019; GSMA Connected Women, 2020
9 Tyers, A and Banyan Global, 2020

## The main data sources for the adult women cohort are:[10]

- Annual quantitative surveys from GSMA Connected Women across up to 21 LMICs since 2018. The GSMA defines the digital gender gap as the difference between men and women as a proportion of the rate for men.
- National statistics offices that report to the International Telecommunication Union (ITU) for its annual Facts and Figures quantitative report (although only 69 countries submit sex-disaggregated data). Like the GSMA, the ITU defines the digital gender gap as the difference between men and women as a proportion of the rate for men.
- The Web Foundation, which publishes quantitative and qualitative primary data sets in a few select LMICs (although not annually). Unlike the GSMA and the ITU, the Web Foundation measures the digital gender gap as the difference between men and women as a proportion of the rate for women, making it female-centered.
- After Access, which publishes nationally representative primary quantitative datasets across 16 countries in Africa, Asia and Latin America (although not annually).
- The Demographic and Health Surveys (DHS), which has started collecting nationally representative sex-disaggregated data since Phase 7 of the survey programme on internet use (if someone has ever used it, or has used it in the past 12 months) and mobile ownership in over 80 LMICs countries. These generally cover women aged 15-49 as well as men.
- The Multi-Indicator Cluster Surveys (MICS), a household survey program on children and women conducted by UNICEF, which has also started collecting sex-disaggregated ICT data in 118 countries in Wave 4.
- In addition, Facebook's 2020 Survey on Gender Equality at Home captures household gender dynamics in 500,000 households during the COVID-19 pandemic using online surveys, indicating an interest in this space among big tech companies.

However, despite increased availability of data about adult women, there are challenges in methodology. Most research is limited to relatively small-scale studies of selected countries rather than studies at a global scale. In many countries, no sex-disaggregated data exists at all. Measurement methods also differ between data sets, leading to different data points. Many data sets focus on digital access only rather than both digital access and use, while many do not collect qualitative data (which means that data lacks nuance and depth). Quantitative data can highlight what is happening, while qualitative data can highlight why it is happening.

In addition, data sets are often reliant on end user surveys, which can be time-consuming and expensive. Because of this, the surveys are often not run regularly and so are not longitudinal, which means it is difficult to track changes over time and measure progress in closing the gender digital divide.

10 Tyers, A and Banyan Global. 2020

For more detailed information, see **USAID's Gender Digital Divide Desk Review Report.**

The only multi-country study that examines the gender digital divide for adolescent girls is the 2018 Girl Effect and Vodafone Foundation 'Real girls, real lives: connected' study, which was the first comprehensive study into adolescent girls and mobile technology. It drew on experiences of more than 3,000 girls and boys from 25 countries - predominantly LMICs in Africa, Asia, the Middle East and North and South America - to understand how girls access or try to access mobile phones and the internet.

The study indicates a pattern of lower digital access and use for girls, similar to that for women.

However, there is little global insight beyond this. Some of this is due to methodology. The vast majority of the global statistics quoted in Box 1 come from survey data such as GSMA Connected Women, whose studies only sample adults over the age of 18. This is partly because GSMA studies draw on data collected by the GSMA Intelligence team as part of their annual Consumer Survey, which only samples adults. In addition, GSMA defines a mobile user as someone who has sole or main use of a SIM card,[11] though laws in many countries prevent anyone under the age of 18 from purchasing a SIM card.[12]

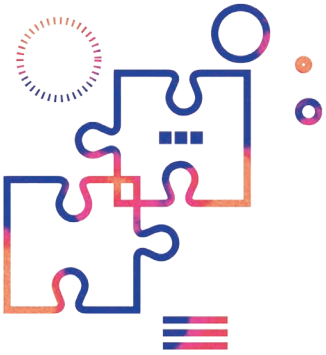
## SIM registration laws also present a challenge for leveraging other data sources to understand digital adoption for girls under the age of 18.

Many statistics bases use mobile (or mobile internet) subscriber data to understand digital penetration, but this data is rarely sex-disaggregated and is unlikely to include people under 18 because they do not appear in statistics that use SIM registration as a proxy for access.

The other often-quoted source of gender digital data is the ITU, which collects data from national statistics offices. While the ITU reports some youth data (e.g., aged 15–24), it is not clear whether their sex-disaggregated data covers adult women only (aged 18 and over) or if it includes girls aged 15 and over (as other ITU data sets would suggest). The ITU also does not appear to have publicly available age- and sex-disaggregated data for girls aged 15–18.

Even the Girl Effect and Vodafone Foundation study, which expressly looks at adolescents, has some limitations in that its insights are very limited for non-mobile or non-internet girl users. This is because the study sample only covered current mobile users, and did not include the perspectives of non-mobile users.

11 GSMA Connected Women, 2020
12 GSMA, 2018



It can also be challenging to conduct research with adolescent girls for a range of reasons. Adolescent girls may have more parental restrictions and tend to be less visible than their male peers. Often, a research team may need to involve partner networks such as schools or non-governmental organisations (NGOs), which can broker introductions. These necessary, but more complicated, approaches may make the logistics of data collection more time-consuming, and may act as a deterrent for researchers.

These methodological challenges may explain the lack of data for adolescent girls and why less research is dedicated to understanding their patterns of digital access and use. But we urgently need more data. For adult women, sex-disaggregated digital data helps stakeholders shape action to close the gap in digital divide as a result of these insights.

**For girls aged under 18, we know so little about their patterns of digital access and use.**

**Without deeper insights, it is very difficult to design meaningful approaches and action to address the gap. We need more data and evidence of the gender digital divide for girls in order to raise attention and leverage action for closing the gap.**

There are, however, alternative data sources. This paper describes the findings of an innovative approach to data collection using one of these alternative data sources: using big data to try to contribute to the body of evidence for the gender digital divide for adolescent girls.

# Using big data

## as a methodology to evidence the gender digital divide

Big data is increasingly recognised as having an essential role to plug digital evidence gaps and generate evidence of progress towards Sustainable Development Goal (SDG) 5: Gender equality, especially related to digital technology access and use.[13] It is increasingly a useful avenue to contribute insights into the gender digital divide.

With support from Data2X and as a part of the Big Data for Gender Challenge, the University of Oxford and Qatar Computing Research Institute (QCRI) set up the Digital Gender Gap Portal. As part of their project, they use a novel source of big data – specifically social media advertising data from platforms such as Facebook, Google, Snapchat and others – to monitor global digital gender gaps.

Social media advertising data provide aggregate counts of users of different online platforms, which are normally provided to advertisers on these platforms. These aggregate counts of users are available for different countries and can be disaggregated by sex, age, and location as well as behaviour (such as interests) and can be pulled from the platform regularly to track and measure in real time. These data can also provide information related to device types used for accessing social media platforms, for example mobile devices, which is helpful to understand the different forms of digital access.

**The Digital Gender Gap Portal draws on a peer-reviewed methodology[14] and uses Facebook advertising data in combination with other data sources to measure global gender gaps in internet and mobile access.**

13 UN Women, 2018
14 Fatehkia, Kashyap and Weber, 2018

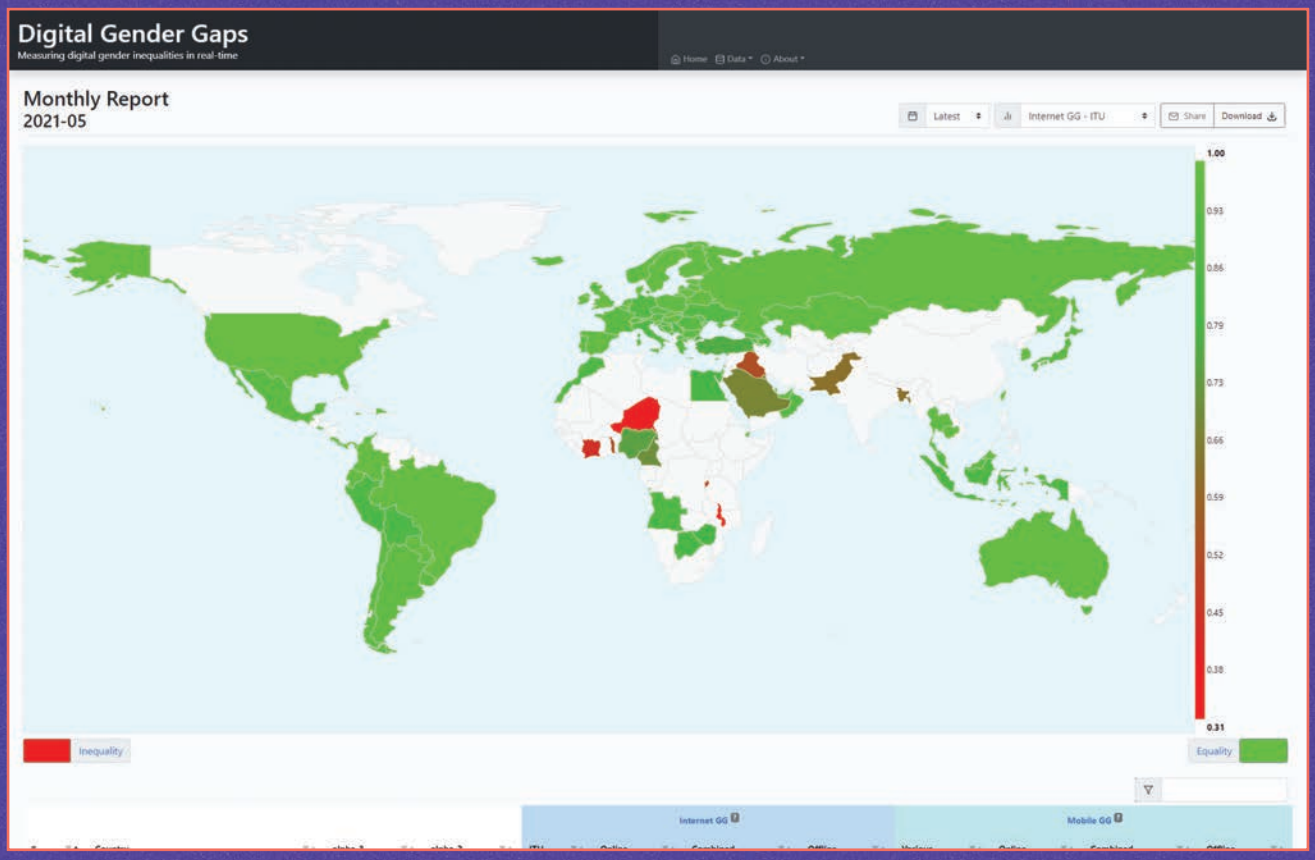



Figure 1 The digital gender gap portal

## Box 2
## The digital gender gap big data methodology

The University of Oxford and QCRI have used big data from Facebook and Google advertising platforms to extrapolate internet data access overall, and generated indicators of country-level gender gaps from Facebook and Google data, which measures the ratio of female to male monthly active users in a given country.[15] The team then built regression models to predict gender gaps in internet use and digital skills in different countries and using this model, extrapolated estimates of the internet gender gap to expand global coverage of this indicator beyond those for which the ITU survey data are available.[16]

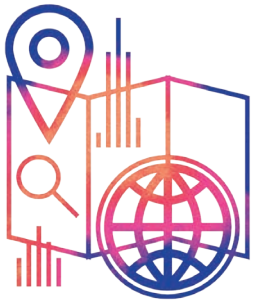
15 Kashyap et al, 2019; Kashyap et al, 2020
16 Kashyap et al, 2020

Facebook and Google online indicators were very strongly correlated with the ITU's survey data on internet access and gender gaps.[17] In both the ITU and the Facebook data sets, the internet gender gap is much larger in sub-Saharan Africa and south Asia in both data sets, and women are much less online compared to men.

This strong correlation with ITU survey data clearly shows that a big data methodology is reliable.

Therefore, this methodology, which builds statistical models using big data, can help us understand the digital gender gap in more depth and in more detail. It can help overcome a lot of the methodological challenges in survey data collection by allowing us to generate longitudinal data to measure changes over time in a more cost-effective way. It can also help not only understand gaps for countries where there is no data, but also understand gaps and patterns at a subnational and community level.

**The Oxford and QCRI team state with a high degree of certainty that when women are strongly under-represented on social media platforms, they are unlikely to have equal internet access in these countries in general.[18]**



## Big data from social media platforms has provided insights for adult women, suggesting that it also has potential to be leveraged to study patterns of digital access for the girl cohort.

Moreover, since large social media platforms have global coverage and a large number of users, we are able to capture significant online populations in real time through social media advertising data. Because the Oxford and QCRI big data methodology for measuring the gender digital divide for adult women correlated so strongly with survey data, our hypothesis was that the same data source could also be used to help serve as a first proxy measure of digital gender gaps for adolescents under the age of 18.

17 Kashyap et al, 2019
18 Kashyap et al, 2019; Kashyap et al, 2020

While we cannot predict internet access gender gaps using statistical (regression) models in the same way for the adolescent age group due to limited data availability of survey-based sources for this age group, the big data enable us to gain a first proxy of digital access for this age group in a global, comparative way.

## Big data can provide a perspective of different platforms like Facebook or Snapchat and can help us understand how adolescent digital gender gaps compare with adult digital gender gaps.

Big data can also help us understand how use of specific types of social media may be gendered.

Another advantage in using big data from social media advertising platforms as an option for generating data on adolescent girls is that it is not reliant on SIM registration laws or SIM data, as it utilises social media platforms instead. It also helps overcome the challenges of gaining access to adolescent girls for surveys. On the other hand, unlike surveys, a potential limitation with these data is that they are provided with limited metadata or descriptions about how the different categories are identified and applied to different users, and/or whether they are inferred using algorithms. As age and gender are often the most basic information required to set up a profile on these platforms, these data likely directly reflect these fields as provided by users. Another limitation with these data is that while they are able to tell us about the aggregate numbers of users of these platforms, they are unable to tell us how they are used, and whether girls and boys use them differently.

# Big data and the gender digital divide for adolescent girls: key findings

**Social media platforms allow users aged 13 and above to open an account, and so social media advertising platforms have data for adolescent users aged 13 and above, disaggregated by gender. However, this data has not been analysed yet to look for digital gender gaps for teenage users.**

## Box 3
## Methodology to identify digital gender gaps for adolescents

This was a unique opportunity to work together to apply the Oxford and QCRI big data methodology to mine the data that already exists to really understand the gender gaps for adolescents in a way that has not been done before. Doing an exercise like this could shed light on not only patterns of digital access, but also on patterns of engagement with different platforms, to understand gendered practices and bring new evidence to a space that is severely lacking in data.

To do so, the Oxford and QCRI team used big data from Facebook and Snapchat advertising platforms for users aged 13 to 17 on both platforms from 2019 to 2020. From this, they calculated a Teenage Gender Gap Index (TGGI) for both platforms for as many countries across the world as these data were available. For Facebook, this is the ratio of female to male monthly active users of Facebook, and for Snapchat this is calculated by dividing the average female to average male ratio of audience on Snapchat by the female to male ratio of the population.

A TGGI value below 1.0 indicates a larger gender gap disfavouring girls, with more male users than female users, while a TGGI value at 1.0 indicates gender parity, and a value greater than 1.0 indicates female users exceeding the number of male users. TGGI values for both of these platforms can serve as a proxy for digital access amongst teenagers.

# 6 key insights from data analysis

## 1 Facebook and Snapchat audiences are much younger in LMICs


According to big data analysis, teenagers (age 13 to 17) account for a larger percentage of total Facebook users in LMICs countries compared to higher income countries (HICs). Since many LMICs have younger populations, the age structure of the populations was factored in, and data continued to show that the teenage Facebook population of LMICs tends to be proportionately larger. Similarly, teenagers account for a larger percentage of total Snapchat users in LMICs: 25% of total users in Europe and North America vs. 25 - 40% of total users in Latin America and parts of Asia.

## 2 Facebook shows a significant global gender digital divide for girls under 18

Analysis of Facebook data across 184 countries revealed a digital gender gap for girls aged 13 to 17. This follows a similar pattern to that of the adult population. 94 of the 184 countries (51%) analysed had a TGGI value below 1.0, which indicates a gender gap, with more male Facebook teenage users than female Facebook users in these countries.

## 3 The Facebook digital gender gap is more pronounced in certain regions

The Facebook data revealed similar regional disparities to the adult population. There were smaller gender gaps in countries in Europe and North America, and larger gender gaps disfavouring girls in countries in sub-Saharan Africa, the Middle East, and Asia.

The countries that indicated the greatest gender gaps - with much higher male teenage use relative to female use - were Tajikistan, Yemen, Afghanistan, Pakistan, Chad, India, Sri Lanka, Azerbaijan, Iraq and Zimbabwe.

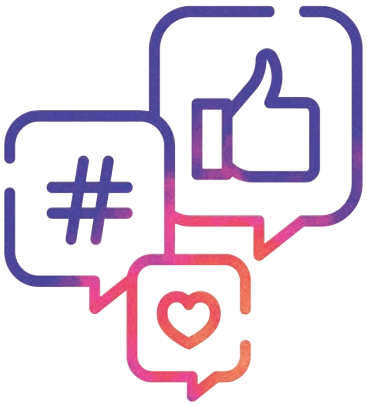
# 6 key insights from data analysis

## 4 Younger girls tend to be online on the two platforms more than older girls

For both Facebook and Snapchat, there is a larger gender gap among older teens (aged 15 to 17) than among younger teens (aged 13 and 14). This pattern is consistent across countries for Snapchat users, while there is some variation between countries among Facebook users.

This general trend of a smaller gender gap in the younger age groups could signal a generational change towards greater digital gender equality in the future.

## 5 The more girls are online, the more they engage



Comparing monthly active users (MAUs) to daily active users (DAUs) of Facebook can provide insight into different patterns of digital gender inequality in terms of intensity of use. For countries with a TGGI higher than 1.0 - i.e., those with more female than male Facebook users - the DAU-based TGGI estimates are higher than the MAU-based estimates.

This suggests that in countries where female users have more access to Facebook, they use it more regularly, and their intensity of use tends to exceed that of male users.

## 6 Snapchat use is also gendered - but in a very different way

Snapchat TGGI values appear generally higher in LMICs than in HICs. This indicates that the gender gaps in LMICs are lower than in HICs. There appears to be a greater teenage female audience on Snapchat relative to a teenage male audience in LMICs than in HICs. Most countries in Europe, North and Latin America have lower TGGI values than certain countries in Asia, indicating that Asian countries have higher gender parity in Snapchat use. The countries with the lowest Snapchat TGGI values - and therefore greater male audience relative to female audience - include Norway, Ireland, and Denmark. Countries with the highest Snapchat TGGI - and therefore greater female audience relative to male audience - include Indonesia, Russia and Japan.

## These 6 key insights show that there is a global gender digital divide for adolescent girls that echoes the patterns seen in adult women. More boys are online than girls at a global level, and the gap is larger in LMICs.

We could surmise that the countries where the gender gaps are identified as being larger tend to be countries with stronger patriarchal norms, where girls face more social, economic and cultural barriers - but this would need further investigation. The suggestion that frequency and intensity of digital use increases when girls have more access to digital platforms echoes the findings of the Girl Effect and Vodafone Foundation study, and could be a useful insight when designing digital literacy interventions for girls, especially when related to organic learning rather than formalised approaches. The finding that younger girls have more access is particularly interesting. It may be worth tracking this over time to see if the gap widens as these younger girls get older, and if this is a global trend or only in certain countries, and undertaking further research to understand the factors that might influence that.

The initial analysis of these big data Facebook and Snapchat data sets has already provided new insights on digital adoption for adolescent girls at a global level. However, it is difficult to draw firm conclusions from this data set alone. In many ways, it raises more questions about why those patterns exist. This is most likely because this was a short initial exploratory exercise to test the hypothesis that big data can be used to generate reliable and cost-effective insights about the gender digital divide for adolescent girls.

To draw firmer conclusions and to trigger action to close the gender digital divide for girls, additional data mining activities may be required. One shortcoming with the big data methodology is its reliance on quantitative data only - it tells us the 'what' but it does not tell us the 'why'. It may also be worth including supplementary qualitative studies to generate deeper insights into the quantitative trends identified from the big data in specific countries.

## A potential limitation with the big data methodology is that users, especially teenagers, may lie about their age to make themselves look older to join the platform.

This results in what has been called 'invisible social media users'. While we have no way of verifying this via the big data methodology, it is unlikely that what we think are teenagers are actually adults. In fact, it is more likely that these users are actually younger than we think, and under the age of 13.

# Conclusions

The findings from this exercise show that big data can be a useful data source to provide some insights about girls' digital access and use, as it helps overcome some of the methodological challenges in other data collection methods and has the potential to produce regular measures of digital access by gender and age.

**It can help us understand whether girls are actually online or not, who is online, and in which countries or regions.**

---

There is also great potential to look at other platforms such as Instagram to understand trends and differences across more platforms.

**Big data is useful because we can potentially use it to track longitudinal changes, and measure how the gender digital divide for adolescent girls is changing over time.**

The Oxford and QCRI team periodically share data on the SDGs Today portal to help track progress against SDG 5: Gender Equality for adult women, and so this data could also be leveraged to track progress for adolescent girls as well as a separate set of indicators.

There is also potential to look at big data at a national and subnational level, and focus on specific countries. Data sets are also available for device types as well as content consumed on different platforms,

**which can help understand both the patterns of access as well as the patterns of online behaviour and any gendered practices**

and how this changes over time. It may also be possible to seek validation of the key themes that have emerged from other gender digital divide literature for girls, such as privacy and security (by looking at privacy settings for users, for example), and including supplementary qualitative research to uncover insights identified in the quantitative trends. The big data methodology may also be a helpful tool in validating any findings from survey data collected at a country level.

## There is still much we do not know about girls' meaningful use of digital technology and the internet.

We hope to explore further how big data can help generate more insights and evidence, and how it could be used alongside other data sources.

We are also interested in hearing from others working in this space, and are using innovative ways of generating insights into the gender digital divide for girls.

- What is your view on using big data to understand the gender digital divide?
- Do you work on big data and how it relates to digital use?
- Are you using any other alternative (and cost-effective) data sources?
- How could big data be applied to understand the impact of digital interventions that you are planning or have implemented?
- Are you interested in collaborating, and being part of a big data and gender and innovation community?

**Get in touch! Reach out to the authors:**

**UNICEF**
Gerda Binder
gbinder@unicef.org

Alex Tyers-Chowdhury
atyers@unicef.org

**University of Oxford**
Ridhi Kashyap
ridhi.kashyap@nuffield.ox.ac.uk

**University of Southampton**
Reham Al Tamime
rat1g15@soton.ac.uk

**QCRI**
Ingmar Weber
iweber@hbku.edu.qa

# References

EQUALS, 2019.
**Taking stock: data and evidence on gender equality in digital access, skills, and leadership.**
Available at https://www.itu.int/en/action/gender-equality/Documents/EQUALS%20Research%20Report%202019.pdf

Girl Effect and Vodafone Foundation, 2018.
**Real lives, real girls, connected.**
Available at https://www.girlsandmobile.org/

GSMA, 2018.
**Access to mobile services and proof of identity: the undisputed linkages.**
Available at https://www.gsma.com/mobilefordevelopment/resources/access-mobile-services-proof-identity-global-policy-trends-dependencies-risks/

GSMA Connected Women, 2020.
**The mobile gender gap report 2020.**
Available at https://www.gsma.com/mobilefordevelopment/wp-content/uploads/2020/02/GSMA-The-Mobile-Gender-Gap-Report-2020.pdf

ITU, 2019.
**Facts and figures.**
Available at https://itu.foleon.com/itu/measuring-digital-development/home/

Fatehkia, M, Kashyap, R, and Weber, I, 2018.
**Using Facebook ad data to track the global digital gender gap.**
Available at https://www.sciencedirect.com/science/article/pii/S0305750X18300883

Kashyap, R, Weber, I, Fatehkia, M, Knowles I and Al Tamime, R, 2019.
**Using Facebook and Google advertising data to measure the gender digital divide.**
Available at https://data2x.org/resource-center/big-data-and-gender-brief-series/

Kashyap, R, Fatehkia, M, Al Tamime, R and Weber, I, 2020.
**Monitoring global digital gender inequality using the online populations of Facebook and Google.**
Available at https://www.demographic-research.org/volumes/vol43/27/default.htm

LIRNEasia, 2019.
**After access Asia report 3.0**
Available at https://lirneasia.net/2019/05/afteraccess-asia-report3/

OECD, 2018.
**Bridging the digital gender divide.**
Available at https://www.oecd.org/going-digital/bridging-the-digital-gender-divide.pdf?mc_cid=9cbd34bb84&mc_eid=86778a9e72

Plan, 2020.
**Bridging the gender digital divide.**
Available at https://plan-international.org/education/bridging-the-digital-divide

Tyers, A and Banyan Global, 2020.
**The gender digital divide desk review report.**
Available at https://www.marketlinks.org/weege-wiki/gdd-desk-review-report

The Web Foundation, 2015.
**Women's rights online.**
Available at http://webfoundation.org/docs/2015/10/womens-rights-online_Report.pdf

The Web Foundation, 2016.
**Women's rights online: digital gender gap audit.**
Available at http://webfoundation.org/about/research/digital-gender-gap-audit/

The Web Foundation, 2020.
**Women's rights online: closing the digital gender gap for a more equal world.**
Available at https://webfoundation.org/research/womens-rights-online-2020/

UNESCO, 2019.
**I'd blush if I could - closing gender divides in digital skills through education.**
Available at https://en.unesco.org/Id-blush-if-I-could

UNICEF, 2019.
**Gender counts: East and Southeast Asia.**
Available at https://www.unicef.org/eap/reports/gender-counts-east-and-southeast-asia

UN Women, 2018.
**Gender equality and big data.**
Available at https://www.unwomen.org/en/digital-library/publications/2018/1/gender-equality-and-big-data

**UNICEF Gender and Innovation**
Evidence briefs - Insights into the gen g divide for girls

unicef
for every child