Big Data for SDG3 and SDG5
Promise and inequality traps

Claudia Abreu Lopes
Digital Health Week - 21 August 2019

#SDG5forHealth #Equality4Health #SDG3forSDG5
2019 SDG Gender Index

Source: Equal Measures 2030, 2019
Of the 14 indicators for SDG5, only 4 indicators have internationally accepted standards and data is regularly collected by most countries.

Of the 51 gender-related indicators across 14 SDGs, only one quarter have gender data available.

<table>
<thead>
<tr>
<th>Target</th>
<th>Goal 5: Achieve gender equality and empower all women and girls</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.1 End all forms of discrimination</td>
<td>5.1.1 Whether or not legal frameworks are in place to promote, enforce and monitor equality and non-discrimination on the basis of sex.</td>
</tr>
<tr>
<td>5.2 Eliminate all forms of violence against women and girls</td>
<td>5.2.1 Proportion of ever-partnered women and girls aged 15 years and older subjected to physical, sexual or psychological violence by a current or former intimate partner in the previous 12 months, by form of violence and by age.</td>
</tr>
<tr>
<td></td>
<td>5.2.2 Proportion of women and girls aged 15 years and older subjected to sexual violence by persons other than an intimate partner in the previous 12 months, by age and place of occurrence.</td>
</tr>
<tr>
<td>5.3 Eliminate all harmful practices such as child marriage and FGM</td>
<td>5.3.1 Proportion of women aged 20-24 years who were married or in a union before age 15 and before age 18.</td>
</tr>
<tr>
<td></td>
<td>5.3.2 Proportion of girls and women aged 15-49 years who have undergone female genital mutilation/cutting, by age.</td>
</tr>
<tr>
<td>5.4 Recognise and value unpaid care and domestic work</td>
<td>5.4.1 Proportion of time spent on unpaid domestic and care work, by sex, age and location.</td>
</tr>
<tr>
<td>5.5 Ensure women’s full and effective participation and equal opportunities for leadership</td>
<td>5.5.1 Proportion of seats held by women in (a) national parliaments and (b) local governments.</td>
</tr>
<tr>
<td></td>
<td>5.5.2 Proportion of women in managerial positions.</td>
</tr>
<tr>
<td>5.6 Ensure universal access to sexual and reproductive health and reproductive rights</td>
<td>5.6.1 Proportion of women aged 15-49 years who make their own informed decisions regarding sexual relations, contraceptive use and reproductive health care.</td>
</tr>
<tr>
<td></td>
<td>5.6.2 Number of countries with laws and regulations that guarantee women aged 15-49 years access to sexual and reproductive health care, information and education.</td>
</tr>
<tr>
<td>5a Undertake reforms to give women equal rights to economic resources</td>
<td>5a.1 (a) Proportion of total agricultural population with ownership or secure rights over agricultural land, by sex; and (b) share of women among owners or rights-bearers of agricultural land, by type of tenure.</td>
</tr>
<tr>
<td></td>
<td>5a.2 Proportion of countries where the legal framework (including customary law) guarantees women’s equal rights to land ownership and/or control.</td>
</tr>
<tr>
<td>5b Enhance the use of enabling technology, in particular ICTs</td>
<td>5b.1 Proportion of individuals who own a mobile telephone, by sex.</td>
</tr>
<tr>
<td>5c Adopt and strengthen sound policies and enforceable legislation for the promotion of gender equality</td>
<td>5c.1 Proportion of countries with systems to track and make public allocations for gender equality and women’s empowerment.</td>
</tr>
</tbody>
</table>
## Availability of sex-disaggregated values of SDG indicators at global or regional levels

### Programme area:  
- Reproductive, maternal and child health  
- Infectious and noncommunicable diseases  
- Injuries, violence and environmental risks  
- Health systems and financing

### Sex disaggregated values reported

<table>
<thead>
<tr>
<th>ID</th>
<th>Indicator</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.2.1</td>
<td>Under-5 mortality</td>
</tr>
<tr>
<td>3.3.1</td>
<td>New HIV infections</td>
</tr>
<tr>
<td>3.3.2</td>
<td>Tuberculosis incidence</td>
</tr>
<tr>
<td>3.4.1</td>
<td>NCD mortality</td>
</tr>
<tr>
<td>3.4.2</td>
<td>Suicide mortality</td>
</tr>
<tr>
<td>3.9.1</td>
<td>Air pollution mortality</td>
</tr>
<tr>
<td>3.9.2</td>
<td>Unsafe water and sanitation mortality</td>
</tr>
<tr>
<td>3.9.3</td>
<td>Poisoning mortality</td>
</tr>
<tr>
<td>16.1.1</td>
<td>Homicide</td>
</tr>
</tbody>
</table>

### Sex disaggregated values not currently reported

<table>
<thead>
<tr>
<th>ID</th>
<th>Indicator</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.2.2</td>
<td>Wasting in children</td>
</tr>
<tr>
<td>2.2.2</td>
<td>Overweight children</td>
</tr>
<tr>
<td>3.3.3</td>
<td>Malaria incidence</td>
</tr>
<tr>
<td>3.3.4</td>
<td>Hepatitis B prevalence</td>
</tr>
<tr>
<td>3.3.5</td>
<td>Need for NTD interventions</td>
</tr>
<tr>
<td>3.6.1</td>
<td>Road traffic mortality</td>
</tr>
<tr>
<td>3.8.1</td>
<td>UHC service coverage index</td>
</tr>
<tr>
<td>3.9.1</td>
<td>Nurse/midwife density</td>
</tr>
<tr>
<td>3.9.1</td>
<td>Dentist density</td>
</tr>
<tr>
<td>3.9.1</td>
<td>Pharmacist density</td>
</tr>
<tr>
<td>17.9.2</td>
<td>Cause-of-death data completeness</td>
</tr>
</tbody>
</table>

### Data collected at household or higher level

<table>
<thead>
<tr>
<th>ID</th>
<th>Indicator</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.1.1</td>
<td>Safe drinking-water coverage</td>
</tr>
<tr>
<td>6.2.1</td>
<td>Safe sanitation coverage</td>
</tr>
<tr>
<td>6.2.1</td>
<td>Water sector ODA</td>
</tr>
<tr>
<td>7.1.2</td>
<td>Clean energy coverage</td>
</tr>
<tr>
<td>11.6.2</td>
<td>Fine particulate matter in urban areas</td>
</tr>
<tr>
<td>3.8.2</td>
<td>Household health expenditure &gt;10%</td>
</tr>
<tr>
<td>1.1.1</td>
<td>Household health expenditure &gt;25%</td>
</tr>
<tr>
<td>3.8.2</td>
<td>Domestic government health expenditure</td>
</tr>
<tr>
<td>3.8.2</td>
<td>ODA medical research &amp; basic health sectors</td>
</tr>
<tr>
<td>3.8.2</td>
<td>International Health Regulations capacity</td>
</tr>
</tbody>
</table>

### Female specific

<table>
<thead>
<tr>
<th>ID</th>
<th>Indicator</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.7.1</td>
<td>Met need for family planning</td>
</tr>
<tr>
<td>3.7.2</td>
<td>Adolescent birth rate</td>
</tr>
<tr>
<td>3.1.1</td>
<td>Maternal mortality</td>
</tr>
<tr>
<td>3.1.2</td>
<td>Skilled attendance at birth</td>
</tr>
<tr>
<td>5.1.2</td>
<td>Intimate partner violence</td>
</tr>
</tbody>
</table>
Figure 1.3  Relationship between prevalence of child marriage and female enrolment in secondary education

Figure 1.4  Percentage of women (aged 15–49) not involved in major household purchasing decisions

Social norms influence women’s participation in household decision-making and health outcomes for children

Where women have a say in decision-making on major household purchases there are better health outcomes for children in terms of reduction of stunting (SDG Target 2.2)³³

Source: Calculations based on data from the Demographic and Health Surveys, and the multiple indicator cluster surveys, latest years⁶⁴
Big Data

• Volume [data exhaust]
• Variety [messy, multi-format data]
• Velocity [real-time data]
• Vinculation [data linkages]
• Veracity [measurement and interpretation]
Big data refer to data sources that require new tools or methods to capture, curate, manage and process them in an efficient way.
• Human sourced data
• Process mediated data
• Machine generated data
• Media sourced data
1. NO POVERTY
   Spending patterns on mobile phone services can provide proxy indicators of income levels.

2. ZERO HUNGER
   Crowdsourcing or tracking of food prices listed online can help monitor food security in near real-time.

3. GOOD HEALTH AND WELL-BEING
   Mapping the movement of mobile phone users can help predict the spread of infectious diseases.

4. QUALITY EDUCATION
   Citizen reporting can reveal reasons for student drop-out rates.

5. GENDER EQUALITY
   Analysis of financial transactions can reveal the spending patterns and different impacts of economic shocks on men and women.

6. CLEAN WATER AND SANITATION
   Sensors connected to water pumps can track access to clean water.

7. AFFORDABLE AND CLEAN ENERGY
   Smart metering allows utility companies to increase or restrict the flow of electricity, gas or water to reduce waste and ensure adequate supply at peak periods.

8. DECENT WORK AND ECONOMIC GROWTH
   Patterns in global postal traffic can provide indicators such as economic growth, remittances, trade and GDP.

9. INDUSTRY, INNOVATION AND INFRASTRUCTURE
   Data from GPS devices can be used for traffic control and to improve public transport.

10. REDUCED INEQUALITY
    Speech-to-text analytics on local radio content can reveal discrimination concerns and support policy response.

11. SUSTAINABLE CITIES AND COMMUNITIES
    Satellite remote sensing can track encroachment on public land or spaces such as parks and forests.

12. RESPONSIBLE CONSUMPTION AND PRODUCTION
    Online search patterns or e-commerce transactions can reveal the pace of transition to energy efficient products.

13. CLIMATE ACTION
    Combining satellite imagery, crowd-sourced witness accounts and open data can help track deforestation.

14. LIFE BELOW WATER
    Maritime vessel tracking data can reveal illegal, unregulated and unreported fishing activities.

15. LIFE ON LAND
    Social media monitoring can support disaster management with real-time information on victim location, effects and strength of forest fires or haze.

16. PEACE, JUSTICE AND STRONG INSTITUTIONS
    Sentiment analysis of social media can reveal public opinion on effective governance, public service delivery or human rights.

17. PARTNERSHIPS FOR THE GOALS
    Partnerships to enable the combining of statistics, mobile and internet data can provide a better and real-time understanding of today’s hyper-connected world.
Using Twitter data to monitor trends in VAW in Brazil
Monitoring disease outbreak in Uganda

https://diseaseoutbreaks.unglobalpulse.net//uganda/
Expanding health coverage in Malawi using mobile phone data

Turning radio conversations into data in Uganda

Table 1. Summary statistics for primary data sets. Phone survey data were collected by the authors in Kigali, in collaboration with the Kigali Institute of Science and Technology. Call detail records were collected by the primary mobile phone operator in Rwanda at the time of the phone survey. Demographic and Health Survey (DHS) data were collected by the Rwandan National Institute of Statistics. N/A, not applicable.

<table>
<thead>
<tr>
<th>Summary statistic</th>
<th>Phone survey</th>
<th>Call detail records</th>
<th>DHS (2007)</th>
<th>DHS (2010)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of unique individuals</td>
<td>856</td>
<td>1.5 million</td>
<td>7377</td>
<td>12,792</td>
</tr>
<tr>
<td>Number of questions in survey</td>
<td>75</td>
<td>N/A</td>
<td>1615</td>
<td>3396</td>
</tr>
<tr>
<td>Primary geographic units</td>
<td>30 districts</td>
<td>30 districts</td>
<td>30 districts</td>
<td>30 districts</td>
</tr>
<tr>
<td>Secondary geographic units</td>
<td>300 cell towers</td>
<td>300 cell towers</td>
<td>247 clusters</td>
<td>492 clusters</td>
</tr>
</tbody>
</table>

Fig. 3. Comparison of wealth predictions to government survey data. (A) Predicted composite wealth index (district average), computed from 2009 call data and aggregated by administrative district. (B) Actual composite wealth index (district average), as computed from a 2010 government DHS of 12,792 households. (C) Comparison of actual and predicted district wealth, for each of the 30 districts, with dots sized by population. (D) Comparison of actual and predicted rates of electrification, for each of the 30 districts. (E) Comparison of actual and predicted cluster wealth, for each of the 492 DHS clusters. CDR, call detail records.
Ethical and privacy issues

1. **Lawful, legitimate and fair use**
   Data should be obtained, collected, analysed or otherwise used through lawful, legitimate and fair means, taking into account the interests of those individuals whose data is being used.

2. **Purpose specification, use limitation and purpose compatibility**
   Any data use must be compatible or otherwise relevant, and not excessive in relation to the purposes for which it was obtained.

3. **Risk mitigation and risks, harms and benefits assessment**
   A risks, harms and benefits assessment that accounts for data protection and data privacy as well as ethics of data use should be conducted before a new or substantially changed use of data (including its purpose) is undertaken.

4. **Sensitive data and sensitive contexts**
   Strict standards of data protection should be employed while obtaining, accessing, collecting, analysing or otherwise using data on vulnerable populations and persons at risk, children and young people or any other data used in sensitive contexts.

5. **Data security**
   Robust technical and organizational safeguards and procedures should be implemented to ensure data management throughout the data lifecycle and prevent any unauthorized use, disclosure or breach of personal data.

6. **Data retention and data minimization**
   Data access, analysis or other use should be kept to the minimum amount necessary to fulfill the purpose of data use.

7. **Data quality**
   All data-related activities should be designed, carried out, reported and documented with an adequate level of quality and transparency.

8. **Open data, transparency and accountability**
   Appropriate governance and accountability mechanisms should be established to monitor compliance with relevant law, including privacy laws and the highest standards of confidentiality, moral and ethical conduct with regard to data use.

9. **Due diligence for third party collaborators**
   Third party collaborators engaging in data use should act in compliance with relevant laws, including privacy laws as well as the highest standards of confidentiality and moral and ethical conduct.
Risks, harms and benefits tool

Data innovation for development guide
RISKS, HARMs AND BENEFITS ASSESSMENT TOOL

CHECKLIST

Rationale for the checklist: Large-scale social or behavioural data may not always contain directly identifiable personal data and/or may be derived from public sources. Nevertheless, its use could potentially cause harm to individuals.

Data use should be always assessed in light of its impact (negative or positive) on individuals’ rights. This risk assessment tool (or checklist) outlines a set of minimum checkpoints, intended to help you to understand and minimize the risks of harms and maximize the positive impacts of a data innovation project (and is intended primarily for projects implemented within international development and humanitarian organizations).

When to use the checklist: The checklist should be considered before a new project is launched, when new sources of data or technology are being incorporated into an existing project, or when an existing project is substantially changed. In particular, this assessment should consider every stage of the project’s data lifecycle: data collection, data transmission, data analysis, data storage, and publication of results.

How to use the checklist: If possible, the questions raised by the checklist should be considered by a diverse team comprised of the project leader as well as other subject matter experts, including—where reasonably practical—a representative of the individuals or groups of individuals who could be potentially affected by the use of data. Consider consulting with data experts, including data privacy experts, and legal experts so that they can assist with answering these questions and help to further mitigate potential risks, where necessary.

Note that the checklist was developed by Global Pulse as part of a more comprehensive Risks, Harms and Benefits Assessment, consisting of Two Steps: (i) Initial Assessment and (ii) Comprehensive Risks, Harms and Benefits Assessment. This checklist is an initial assessment that should help to determine whether a comprehensive Risks, Harms and Benefits Assessment should be conducted.

Nature of the checklist: This checklist is not a legal document and is not based on any specific national law. It draws inspiration from international and regional frameworks concerning data privacy and data protection. The document provides only a minimum set of questions and guiding comments. The checklist and guiding comments are designed primarily as a general example for internal self-regulation. As this checklist offers only minimum guidance, you are encouraged to expand the list depending on the project’s needs, risks, or specific context, or in response to the evolving data landscape. Depending on the implementing organization (its legal status/nature) and applicable laws, the guiding principles, standards and basis for answering these questions may need to be changed.

For more information or to provide input on the checklist, please contact dataprivacy@unglobalpulse.org. This checklist is a living document and will change over time in response to the evolving data landscape. The latest version of the Risks, Harms and Benefits Assessment is available at www.unglobalpulse.org/privacy/tools.

For more information on the privacy protective and ethical use of data, please refer to the UN Principles on the Protection of Personal Data and Privacy and the UNDG Data Privacy, Ethics and Protection Guidance Note (2018).
Characteristics of Big Data

- Big (rare events, heterogeneity, small differences)
- Always-on
- Nonreactive
- Incomplete
- Inaccessible
- Non-representative
- Drifting
- Algorithmically confused
- Dirty
- Sensitive

BUT ...
Risks and challenges

• Elite capture: major risk of ‘black holes’ of data where entire demographics can be missed because of restricted access and use
• Restricted use: factors that constrain women’s ability to participate in social media or other interactive platforms
• Ambient sexism: perceived misogyny on social media
• Platform drifts: population, usage and system
• Ethical and privacy issues
South Asia has the world’s largest mobile gender gap

In India, 67% of men own mobile phones, but only 33% percent of women do.

South Asian countries in general are clear outliers among countries of similar levels of development, with India, Pakistan, and Bangladesh exhibiting some of the world’s highest gender gaps in access to technology.
Economic barriers

Normative barriers

Having a mobile phone makes me feel secure because when I go someplace outside and any problem occurs, I call my family members immediately and get the problem solved. My family becomes available wherever I am so I feel secure.

(Girl, 19, India)

coz sometimes phone aren't good :) especially if you are in social media, some people will judge you.

(Girl, 17, Philippines,

Through phone girls get involved in relationships and many things happen. A girl's life gets spoiled.

(Boy, 18, Bangladesh)

Real girls, real lives, connected – A global study of girls’ access and usage of mobile, told through 3000 voices. Girl Effect and Vodafone Foundation (2018)
Real girls, real lives, connected – A global study of girls’ access and usage of mobile, told through 3000 voices. Girl Effect and Vodafone Foundation (2018)

Reasons for not owning a phone

- Handset costs
- Parents’ safety concerns
- Fathers don’t allow
- Mothers don’t allow
- Family disapproval
- Data costs
- Own safety concerns
- Community disapproval
- Don’t know how to use phones
- Hard to register SIM
- Don’t have time
- Bad signal
Big Data and Artificial Intelligence

Female voices are less likely to be accurately processed by AI assistants because they are trained on male-biased data.

120 Hz
200 Hz
Figure 1: Africa personal data protection regulatory landscape

Delloite (2018). Privacy is Paramount | Personal Data Protection in Africa Personal Data Protection in Africa
Expanding the room...

Citizen-generated data (CGD) is data that people or their organisations produce to directly monitor, demand or drive change on issues that affect them.

It is actively given by citizens, providing direct representations of their perspectives and an alternative to datasets collected by governments or international institutions (Civicus, 2014).
Supplement big data with CGD

- To improve issues of representation and coverage
- Particularly for hard to reach populations, ensuring that NoLB
- To validate trends and insights generated by Big Data
- To involve citizens in the production, interpretation and use of data
WOMEN CONNECT: #LETGIRLSMAP

Humanitarian
OpenStreetMap
Team

the everyday sexism project

MOBISAF AIDS
Thank you!
@cabreulopes
Big Data for SDG3 and SDG5
Promise and inequality traps

Claudia Abreu Lopes
Digital Health Week - 21 August 2019

#SDG5forHealth #Equality4Health #SDG3forSDG5