data2x

partnering for a
gender data revolution

Mapping Gender Data Gaps

EXECUTIVE SUMMARY

Mayra Buvinic
Rebecca Furst-Nichols
Gayatri Koolwal

March 2014

Data2X, named for the power women have to multiply progress in their societies, aims to advance gender equality and women's empowerment through improved data collection and analysis that can provide a solid evidence-base to guide development policy. To provide basic information for a Data2X roadmap, this report maps gender data gaps in developing countries across five domains of women's empowerment: (1) health, (2) education, (3) economic opportunities, (4) political participation, and (5) human security.

Need, population coverage and policy relevance were the three criteria used to select which data gaps to map.

The report suggests "ways forward" to close these gaps using existing and new data sources, including censuses and micro-level surveys, service and administrative records, and the potential use of "big data" as a new source of gender data. These ways forward build on recent and ongoing data initiatives and are intended to inform the Data Revolution (High Level Panel Report, 2013) and the stand-alone gender equality goal (UN Women, 2013) called for in the post-2015 period.

Existing international databases often have data that could be disaggregated by sex and analyzed to address gender data gaps. These data sources should be mined before initiating new data collection efforts. Going forward, data mining and collection should be undertaken efficiently and in response to meaningful demand for and capability to use the data for policy purposes, and as a tool to drive social change.

Globally, close to 80 percent of countries regularly produce sex-disaggregated statistics on mortality, labor force participation, and education and training. Less than a third of countries disaggregate statistics by gender on informal employment, entrepreneurship (ownership and management of a firm or business), violence against women, and unpaid work. Reflecting these gaps, the UN Inter-Agency and Expert Group on Gender Statistics (IAEG-GS) has compiled a minimum set of 52 gender indicators and divided them into three tiers according to their conceptual clarity, international standards and regular production (UNSD, 2012). Educational indicators, followed by health, lead the way in terms of clarity, comparability and country coverage. Availability and coverage in the other domains highlighted in this report are poor, and this is especially the case for economic indicators. Throughout, this report makes reference to this minimum set of gender indicators as well as a proposed set to track a stand-alone gender equality goal in the post-2015 period (UN Women, 2013).

Laws and policies, particularly those related to marriage, property, and labor rights, influence health, education, and economic outcomes for women. However, comprehensive information on policies related to women is difficult to obtain. Databases that help document laws and policies affecting women's outcomes are reviewed in this report. Additional efforts to expand coverage and improve the robustness of these databases are warranted to round out the availability of policy relevant data on women.

Gender data gaps

Based on need (severity and disparities in outcomes affecting women), population coverage (number of girls and women that potential benefit from closing the gender data gap), and policy relevance, the table below lists the 28 gender data gaps mapped in this report by domain. For each data gap, the table classifies the type of gap according to the following desirable criteria for gender data: (1) coverage and regular country production; (2) cross-country comparability reflected in international data standards; (3) complexity (information that cuts across domains allowing users to understand patterns and determinants of specific variables); and (4) granularity (large and detailed data sources that can be disaggregated by demographic and other characteristics).

The data gaps by domain are as follows

(A) Health: Better vital registration data is needed to gather accurate information on maternal deaths, including causes of death by age, in high mortality low-income countries that do not currently collect and report this data (a main reason being that very large samples are needed to collect valid data on maternal mortality). Maternal mortality rates remain unacceptably high in these countries and accurate rates and conditions leading to maternal deaths are key information for building evidence-based policy. More and better information is also needed on maternal morbidity.

But women's health issues go beyond maternal conditions. There is substantial evidence that women lose more healthy life to disability compared to men, including excess disease burdens that are unrelated to motherhood such as Alzheimer's disease, dementia, depression, osteoarthritis, and other disabilities. Data collection efforts need to address largely unreported causes of women's excess disease burdens and parse out the contributions of sex and gender, and their interaction, in the etiology, onset, progression and prevention of these conditions.

Additional important data gaps in women's health are data on violence against women and mental health. These issues are both broad, affect large numbers of girls and women, and despite their impact are severely limited in data. A push to close these data gaps may create a snowball effect, where more data increases the visibility of these issues and provides an impetus for women to seek help and for service providers to offer more treatment options.

Adolescent health – including information on the social determinants of healthy behaviors among adolescent girls – also stands on its own as a data gap with important implications for policy. The determinants of healthy behaviors are particularly important during adolescence since actions at this age chart the path for the rest of women's lives.

Relatedly, more reliable data on women's utilization of maternal and non-maternal health services (underutilization that has been documented in many poor countries) would go a long way towards the design of better health interventions for girls and women.

(B) Education: Improving educational outcomes ensures that students, male and female, reap the social and economic returns to education, and may have a multiplier effect on enrollment. Having internationally comparable measures of learning outcomes disaggregated by sex should drive gender data efforts in this domain. Current measures of education quality across countries are largely based on inputs or use different exams to assess learning outcomes and are not sufficient to assess outcomes in a standardized manner.

A second data gap is better information on socially excluded girls – due to race, ethnicity, religion, location (rural vs urban) or disability – who are likely to suffer the double disadvantage of gender and social exclusion, resulting in lower enrollment levels and in poor learning outcomes for those who do enroll.

The third gap is global information on adolescent girls' transition from education to the workforce, as well as what happens to the large numbers of young women in developing countries who fail to make this transition. This information will allow for targeted policymaking to improve the relevance of schooling for the most disadvantaged girls in the educational system and to help with their incorporation into the workforce.

(C) Economic Opportunities: Having quality sex disaggregated data on informal employment is needed; women in the developing world are overrrepresented in work and enterprises that are not accurately or officially counted. Understanding women's experience in these areas requires having detailed data on their unpaid work, including reliable time use data, types and extent of informal employment, as well as entrepreneurial activities. Other data gaps include earnings disparities and opportunity costs of paid work for women, female labor migration (including age and other demographic characteristics, reasons for migrating, remittances sent and working conditions), employment mobility (that is, on those who are looking to move to paid work in the formal sector, and those who are transitioning from home and subsistence production into market work), asset ownership, and access to financial services. Better measurement of women's assets and financial constraints is essential to understanding their economic empowerment, but very few existing national surveys record this information at the individual level.

Data gaps also exist in the agricultural sector, particularly on women's stake in on-farm activities and conditions in agricultural informal employment. Measuring women's agricultural productivity and the factors determining this productivity, including access to land and agricultural resources, is essential for the design of gender-informed agricultural policies.

Women's access to child care services, whether formal or informal, affects their ability to work outside of the home, yet there is currently no consistent data in this area.

Access to information communications technologies, namely mobile phones and the internet, influences a number of areas of women's lives including their ability to communicate with peers, learn about employment opportunities, receive information about the prices of their products, conduct financial transactions, and learn new skills transferred through these technologies. This area rounds out the list of gender data gaps in economic opportunities.

(D) Political Participation: Of the existing data gaps on political and civic participation, the one with by far the greatest potential development significance is closing the gap on sex-disaggregated birth registrations at the national level and, concomitantly, providing national identity documentation. This information could then be used to track voter registration (and turnout) data disaggregated by sex. Closing these gaps would affect girls and women (and boys and men) in socially excluded groups in particular, who are often not counted and therefore unable to claim rights.

Tracking women's political representation at sub-national levels and their leadership roles in grassroots organizations and in key professions is another data priority, especially when information on representation is paired with other data to study the dynamics and outcomes of women's leadership.

(E) Human Security: Very limited data exists on the gender aspects of conflict, so better data collection overall in this area, including sex-disaggregated data on war-related mortality and morbidity, forcible displacement, adaptive responses to conflict, and conflict-related violence, is key.

There is also scant data on women's participation in peace and security efforts, particularly in leadership roles, while this information is basic to adequately monitor the implementation of UN Resolution 1325 at national levels.

Gender-relevant data initiatives

(A) Health: On mortality and morbidity data, the Global Burden of Disease (GBD) Study 2010 estimates disease burdens globally for 1990, 2005 and 2010 and is able to assess prevalence trends for morbidity and mortality disaggregated by sex, age, country, and region.

The UN Department of Economic and Social Affairs (UNDESA) has recently published guidelines for harmonizing the collection of data on violence against women using dedicated sample surveys, and will follow this publication with training of country statistical offices on their use.

The WHO convened Commission on Information and Accountability for Women's and Children's Health recommended establishing vital registration systems in 75 priority countries with high maternal and infant mortality rates.

www.data2x.org

"Countdown to 2015," a global movement, tracks these commitments. Adding to these efforts, global leaders in health statisitics, convened by WHO in 2013, have agreed to work together to improve health information, increase the collection of vital registration data and reduce reliance on estimates from statisitical models.

(B) Education: Efforts to improve metrics on learning outcomes, disaggregated by sex, include: (1) UNESCO Institute for Statistics' Literacy Assessment and Monitoring Program (LAMP), which aims to improve cross-national literacy and numeracy statistics; (2) the Learning Metrics Task Force, convened by the UNESCO Institute for Statistics (UIS) and the Center for Universal Education at the Brookings Institution, which is working on recommendations to help countries measure and improve learning outcomes; and (3) the OECD's Program for International Student Assessment (PISA), which assesses attainment in mathematics, reading, and science by testing students' cumulative learning at age 15, and includes indicators on sex, race and ethnicity. The PISA has been administered in 70, mostly middle-income countries, and plans to expand coverage to developing countries in the 2015 assessment cycle through the "PISA for Development" program.

To improve data on socially excluded children, The "Education for All" initiative published the World Inequality Database on Education (WIDE) in 2012, based on DHS and MICS data from over 60 countries. WIDE enables users to compare education attainment (years and completion of primary and lower secondary) between groups within countries, including by wealth, sex, race and ethnicity, and location.

(C) Economic Opportunities: The International Classification of Activities for Time Use Statistics (ICATUS), revised most recently in 2012, is an effort by national, regional and international time use survey experts to develop a standard classification of daily activities that is internationally comparable and relevant for both social and economic policies.

Through the EDGE project, the UN Statistical Division and UN Women collaborate on developing methodological guidelines to collect data on physical and financial assets disaggregated by sex; the same exercise will be done for entrepreneurship. The project will then pilot test gender data collection on assets and entrepreneurship in selected countries.

The Global Financial Inclusion (Global Findex) Database, an initiative of the World Bank and the IFC that started in 2011, measures how adults save, manage their finances, and cope with issue of access to financial services using 148 nationally representative country surveys. The next round of data at the country-level will be available in 2015.

On asset ownership, the Integrated Surveys on Agriculture, part of the World Bank Living Standards and Measurement Surveys (LSMS-ISA), collects detailed nationally representative household panel data showing the links between agriculture, socioeconomic status, and non-farm income activities. The LSMS-ISA also collects standardized individual-level disaggregated data on assets, including ownership, management and control of agricultural plots and livestock, as well as other assets and access to credit.

Adding to the collection of initiatives for better data on agricultural activities, USAID is developing the Core Agricultural and Rural Survey (CARDS) within the World Bank- and FAO- led Global Strategy to Improve Agriculture and Rural Statistics. CARDS will include farm and non-farm indicators for individual household members, disaggregated by sex, incorporating lessons learned from implementing the Women's Empowerment in Agriculture Index, a recent initiative of USAID, IFPRI and the Oxford Human Development and Poverty Initiative.

(D) Political Participation: Increasingly, countries are using mobile phones and biometric identification to provide a unique digital identity to citizens and record birth registrations. Noted examples include the National Population Commission of Nigeria, which uses decentralized monitoring to identify disparities in birth registration rates that can be accessed in real time, and the Universal ID program in India, which seeks to provide a unique digital identity to all citizens. Identity registration is closely linked to vital registration of births and deaths – a priority for health sector statisitics (see health domain).

UN Women is working to develop standards for measuring women's representation at the subnational level. The Inter Parliamentary Union (IPU) and IDEA collect some data on women candidates and voter turnout.

www.data2x.org

(E) Human Security: MICROCON is a five-year research program, funded by the European Commission and launched in 2007, that takes a micro-level approach to understanding the conflict cycle. Goals of MICROCON include compiling and collecting surveys and existing data, and advancing methods for qualitative and quantitative data on conflict at the individual, household, and group levels. Among MICROCON's thematic areas are the gender aspects of conflict. Norway's PRIO has set up a conflict database, and also focuses on the gender dimensions of conflict.

As part of the Resolution to Act (Res2Act) initiative, the Institute for Inclusive Security has developed a National Action Plan (NAP) Monitoring and Evaluation Toolkit that prompts data collection on the implementation of NAPs promoting inclusion of women in peace and security processes. The data generated from the Toolkit will help policymakers track implementation, as well as assist individuals in holding governments accountable.

Various parts of the UN system track different components of women's participation. UN Women periodically reports on the number of women negotiators and signatories in peace negotiations. The Department of Political Affairs tracks the number of women named to the positions of lead envoy and mediator to UN-brokered talks. The Department of Peacekeeping Operations tracks the number of male and female uniformed and civilian personnel serving in missions and at headquarters. The African Union (AU), European Union (EU), North Atlantic Treaty Organization (NATO), the Organization for Security and Cooperation in Europe (OSCE), and other multilateral organizations also gather sex-disaggregated data on personnel.

WHO, in coordination with UNAction, has developed a survey tool for women and men to measure different experiences of violence, perpetration, risk and protective factors, and impacts, including a section on mental health.

Ways forward

Bridging gender data gaps can be accomplished by mining existing household and administrative survey databases, enriching existing databases, or building new ones. Exploration of existing international databases for data that can be sex disaggregated and usefully analyzed, but are currently unused, is a first necessary step in bridging gender data gaps. Second, there are ways that existing data can be used to provide richer information on girls and women. Surveys covering different topics can be supplemented with one another, if they cover the same time period and context, to add data complexity and help tackle broader questions that may be difficult to address with just one source. Correlating data on outcomes with women's age, ethnicity, marital status, income, and other socioeconomic characteristics is important to add granularity to data sources, as indicators of women's status can vary substantially by these variables. Existing databases can also be enriched by adding specific modules with new questions to an existing survey instrument and sampling frame.

Big data – an umbrella term covering transactional and crowdsourcing data from mobile phones and the internet, including online search and social network feeds, as well as global remote sensing data from satellites – provides an exciting avenue to build new gender data sources with sufficient granularity to ask policy questions that involve disaggregating variables by sex, age, group attribution, and other characteristics.

Big datasets can help provide insight on mechanisms underlying policy trends – subjective issues that official data often does not include – similar to traditional qualitative data, but in real time. Big data can also contribute to development by providing interim evidence on different indicators between rounds of other official surveys. In particular, big data can help fill the following gender data and research gaps:

- Help to better understand the behavioral aspects of gender inequality as well as provide information on women's mobility, opinions about conditions that affect them, and their perceptions about policy.

- Complement more standard program evaluation data by offering the perceptions and opinions of women clients and beneficiaries.

- Capture information that is difficult to elicit through regular surveys due to response bias or limited access to respondents.

On **health**, ongoing initiatives to strengthen vital registration systems could benefit from mobile phone-based reporting to help gather more accurate data on maternal mortality rates and causes. Geographic information system (GIS) and remote sensing technologies can also enable mapping and visual representation of the distribution of risk factors, disease, and services. There may be ways of complementing this information with GBD estimates to obtain richer information on disease burdens and risk factors by sex and other demographic features.

The collection of data on mental health and violence against women ideally require dedicated surveys. For such sensitive and underreported health problems, however, mobile phones may complement dedicated surveys by providing an avenue for anonymous reporting. For example, mobile health interventions have significant potential to gather valuable information on health (including attitudes and behaviors) when they solicit queries from the user. Mobile phone reporting mechanisms are also being used in some countries to monitor locations and prevalence of sexual harassment and assault. Including targeted questions or existing survey modules is another option, as the DHS and the RHS do for measuring violence against women.

Sexual and reproductive health information and indicators for adolescent girls should be available through expanded coverage of existing survey instruments (the DHS) and complemented by data from mobile phones, particularly since adolescent girls are more likely than older women to adopt this technology in large numbers. Dedicated surveys covering multiple dimensions of adolescent girls' lives, with prospective panel cohorts that are followed over time, are highly desirable to fill policy gaps on adolescent girls' well-being and opportunities. GBD information can be exploited to get trends and patterns in adolescent health status.

Both client data (collected through traditional survey instruments and mobile phones) and facility-generated information from health service providers are also potentially rich information sources on women's health service utilization, especially if health service provider data is automated.

On **education**, as computerized recordkeeping in developing world school systems grows more widespread, digital data can contribute to understanding disparities in teacher effectiveness, attendance, exam scores, and completion rates by sex and other categories. Where such data systems are lacking, however, innovative forms of crowdsourcing may be equally important.

The Learning Metrics Task Force may chart the way forward to close the data gap in learning through lower secondary schooling. The expansion of the PISA to more developing countries in 2015 and beyond can increase the availability of comparable learning data for youth who should be transitioning to higher secondary education. The WIDE dataset should be expanded to include more countries to ensure information is demanded to capture learning and access for socially excluded girls.

Dedicated panel surveys focused on adolescent girls (see health domain) can be a rich source of information to examine girls' transitions from school to the workforce and to family formation, and can be complemented by information gathered from big data sources, such as mobile phones.

On **economic opportunities,** additional questions in official statistics are needed on unemployment, underemployment, informality, and looking for work, as well as on savings and assets and access to child care. Additionally, data from social networking feeds can help measure expectations and perceptions about the economy, factors affecting women's job mobility, and saving and spending patterns. Crowdsourced data can also measure "herd" effects such as trends in job changes among younger women in the economy who are more likely to be connected to these networks. Big data may be able to fill gaps regarding constraints on entrepreneur behavior, since moving into self-employment can be quite sensitive to changing economic conditions. In countries where mobile banking/transfer services are widely used, transactional data can also measure trends in economic activity linked to the user. Mobile data may also be good for collecting better information on access to financial services, distance traveled for work, and remittances and connections with others while working away from home.

A first policy priority is to count and make visible informal employment as well as understand the dynamics of women's work in this area. This includes tracking the proportion of women (including migrant women) in informal employment as well the numbers of young women whose first job after school is in informal work. In 2013, the ILO published a manual entitled "Measuring Informality" to guide country statistical offices and other organizations on standards for collecting micro-level data on informal employment (ILO, 2013). Recommendations include breaking down employment data into formal and informal work in both agricultural and non-agricultural sectors, as well as collecting better gender-disaggregated data on earnings across these areas.

On **political participation**, data efforts that currently collect information on women's representation at ministerial and parliamentary levels could be expanded to capture representation at subnational levels, although standardizing measures of local representation remains a challenge. National and international professional associations could be tapped to provide similar information for women in key professions. Big data could help with capturing some of this information. Voter registration and turnout data, important to track women's voices in the political process, requires harmonizing collection methodologies and resourcing a central body, endorsed by national Electoral Management Bodies, to coordinate the data gathering.

Digital recording of birth registrations and provision of identity records at the national level could play a significant role in ensuring citizens are counted and governments have accurate population data. The empowerment effects for women and other excluded population groups, and the development impact of this empowerment, could be large. An international effort, with donor support and South-South learning and collaboration could be rolled out targeting countries with particularly low birth registration rates. This data could also be used for information on voter registration and turnout.

Finally, on **human security**, conflict modules in household surveys are needed to gather data on the micro-level relationships between violent conflict and household welfare. In the absence of local capacity, donor support for carrying out national household surveys, along with technical assistance to utilize non-standard sampling techniques, is needed in order to gather this data on individual and household-level effects, especially secondary effects. This includes data on sexual and gender based violence in the context of conflict.

Remote sensing and crowdsourcing information about the locations of victims in disaster-stricken and conflict settings, as well as estimates of displaced persons and population movements, can be an important source of information about the scale and scope of a disaster or conflict. User-generated and remote sensing data is particularly useful in conflict settings where traditional research design must be altered due to lack of sampling frames, population mobility, and insecurity faced by researchers and respondents.

More consistent and thorough reporting on women's leadership in peace and security processes is needed. Res2Act's framework for monitoring and evaluating National Action Plans on Women, Peace, and Security could be used for data on women's participation in peace and security processes.

Table Gender Data Gaps Highlighted in this Report

	Type of gap			
	Lacking coverage across countries and/or regular country production	Lacking international standards	Lacking complexity (information across domains)	Lacking granularity (large detailed datasets allowing for disaggregation)
Health				
Maternal morbidity/mortality	■			■
Women's excess disease burdens	■			
Violence against women	■		■	
Mental health	■		■	■
Adolescent health	■		■	■
Utilization of health services by women	■			■
Education				
Learning outcomes	■	■		
Excluded girls	■	■		
Transition rates	■			
Economic Opportunities				
Unpaid work	■			
Informal employment	■			
Earnings and opportunity cost of paid work	■	■		
Conditions of migrant workers	■	■		
Employment mobility	■	■	■	■
Entrepreneurship	■		■	■
Asset ownership	■		■	■
Productivity in agriculture	■	■	■	■
Access to financial services	■			
Access to child care	■	■	■	■
Access to ICT (mobile phones & Internet)	■	■	■	■
Political Participation				
Representation in local governance, political organizations & the professions	■	■		
National identity documentation	■			■
Voter registration and turnout	■	■		
Human Security				
Conflict-related mortality and morbidity	■			■
Forcibly displaced and migrant profiles	■			■
Impact of conflict on gender variables; women's adaptive responses to conflict	■		■	■
Conflict-related sexual and gender-based violence	■		■	■
Women's participation in peace and security processes	■	■	■	■