Ladysmith

Building alliances for gender equality:

How the tech community can strengthen the gender data ecosystem

—

Building alliances for gender equality:
How the tech community can strengthen the gender data ecosystem

Authors

Dr Tara Patricia Cookson

Dr Julia Zulver

Dr Lorena Fuentes

Dr Melissa Langworthy.

Design

Marc Robinson

Table of Contents

Acknowledgements

The team at Ladysmith would like to extend our sincere gratitude to all of the experts–across a range of public, private, non-profit and multilateral institutions–who generously contributed their time to engaging with us about gender data and the SDGs. This report would not have been possible without this generous input from such a diverse range of research, policy, service delivery and advocacy organizations.

As a feminist research organization, we are acutely aware of the ways in which women and gender equality advocates use digital technology to organize their everyday lives and advance the causes they care about. Women use social media to stay in touch with family members in the context of migration. As small-business owners, they (including the authors of this report) advertise their products and services through online platforms. They support one another through private 'groups' where they share information about the challenges of motherhood, affordable childcare services, and how to leave violent relationships. They use online platforms to fundraise for and raise awareness about feminist causes, and to run for public office. They combine ride-share services with public transportation to build safe routes to and from work and school, and they make up the majority of hosts on certain housing rental platforms. For these reasons, we believe that this conversation–about technology, data, and gender equality-- is an important one.

Find this report at:
http://ladysmithcollective.com

Introduction: Gender Data and The Tech Community

Lucian Alexe @lucian_alexe

Five years ago, gender equality advocates called for a 'Gender Data Revolution' to accompany the 2030 Agenda for Sustainable Development. They wanted to make visible the many facets of gender inequality that all too often go unaddressed because they are not systematically recorded. The call to bridge 'gender data gaps' quickly found favor in public institutions, philanthropic foundations, and non-governmental and civil society organizations. This is because gender data have many uses: they can guide the just and efficient allocation of resources; enable the design of more effective policies, programs and services; serve as a tool for citizens to hold governments accountable to their promises; and help track progress on the Sustainable Development Goals (SDGs).

Gender data come in many forms (see Box 1), and in the digital age we would be remiss to overlook the ways in which gender equality advocates might leverage the large and unconventional datasets produced, for example, by social media platforms, real-time sensors, and electronic records ("big data"). Yet gender equality advocates and technology companies rarely sit at the same table. And it is crucial that they do sit at the same table if we are to take advantage of the novel insights that digitally produced data offer. This is because the production of quality gender data is not a simple exercise in counting women–it requires methodological rigor and a deep familiarity with the ways that power operates in the lives of women and men and girls and boys.[1]

Box 1

What is Gender Data? A Definition

Gender data reflect and make visible differences in the experiences, needs, opportunities, or contributions of women and men, and girls and boys, in all areas of life[2]. The term "gender data" is not synonymous with "data about women," and it isn't synonymous with sex-disaggregated data. Sex relates to biological differences, and gender to socially constructed characteristics along a spectrum. Gender data can be quantitative–for example, sex-disaggregated statistics on the use of paternity leave benefits. It can also be qualitative–for example, data from interviews and participant observation with mothers and fathers might reveal why those benefits are or aren't used. The generation of high quality gender data requires researchers and data scientists to demonstrate a grasp of gender biases embedded in "raw data"; in definitions, classifications[3], and hypotheses; in research questions; in data collection methodologies; and in population samples.

Of course, in order to use gender data, the gender data must first be produced. Various institutions have mapped 'gender data gaps' across the Sustainable Development Goals indicator framework.[4] For example, data gaps exist around gender indicator 5.b.1, the proportion of people who own a mobile phone, by sex, and indicator 11.2.1, the proportion of the population that has convenient access to public transport, by sex, age, and persons living with disabilities. The gaps identified in prior work inform this report, which also draws on the latest research, and consultations with gender and development practitioners regarding the gender data gaps that are most pressing to them. Their accounts highlight the existence of gender data gaps that cut across the SDGs and that are not easily captured in straightforward indicators. This includes, for example, data that is needed in order to understand and respond to the impacts of climate change on women's and men's access to decent work, and the impacts of specific laws, policies and programs on women's exposure to gender-based violence in humanitarian crises.[5]

Identifying Gender Data Needs

The specific purpose of this report is to bring the needs and concerns of the gender and development community into conversation with the intentions, capabilities and data-expertise of the tech community. This report shares findings from original research that asked, through literature review, dataset analysis and expert consultation, the following two questions:

01 **what are common gender data needs–and concerns–within the international community working to achieve the SDGs?; and**

02 **how might the tech community contribute its data, processing power, or human resources to meeting these needs?**

We investigated thematic areas across the SDGs where gender data is sparse, and identified common concerns about data generation and governance shared by women's rights stakeholders and the tech community. We collected examples of existing 'data for good' initiatives that serve as examples for what works, what is under-utilized, and where incremental improvements or entirely new approaches are needed.

Box 2

Research Methods: This report is based on:

A scoping review of **207 academic articles** and policy reports on gender data and the SDGs.

34 publicly available datasets related to gender equality or including a strong gender element.

49 interviews with gender and development experts from across world regions, SDG themes, and public, private, academic, non-governmental, and civil society organizations.

5 roundtable events with participants from the tech community, and public, private, academic, non-governmental, and civil society organizations.

Unlike prior reports, we specifically sought to identify gender data gaps which the tech community might be well-placed to address. We considered, for example, the novel questions that might be answered by the tech community's unconventional data regarding how resources are shared within social networks, or data that might shed light on the patterns of abuse women face in different areas of their lives. Our analysis also reflects an awareness that in some cases, relying on private tech companies to bridge data gaps could introduce unjustifiable risk of unintended consequences. This might be true for example around murdered and missing women and girls, or migrant workers with insecure immigration status—two topics with significant gender data gaps, which we chose not to feature in this report.

Despite these constraints, we found much for the tech community to work on. We identified seven thematic areas, cutting across the SDGs, where data is needed to address humanity's grand challenges (see Box 3). To be clear, the gender data gaps identified here are not exhaustive; they reflect substantive intersections and opportunities to impact the lives of women and girls.

Box 3

Gender data gaps that limit progress on humanity's grand challenges

Recognizing that it is impossible to catalogue all the gender data gaps that exist in relation to the SDGs, we chose to focus on where the tech community could contribute to 1) issues of longstanding feminist concern, and 2) emerging issues that will shape gender equality in the future:

✔ **Climate change and migration (SDGs 8,10,13,14,15):** Enormous gender data gaps exist around both the environment and displacement, and at their intersection. Emerging evidence suggests that women and men migrate differently in response to climate change. Gender data is critically needed to inform climate action and rights-based policies: Why, and when, are people moving, and what supports do they need?

✔ **Family life (SDGs 1,3,5,8):** Gender inequalities within households are hard to capture in conventional household surveys. Complementary social network data could shed light on how family members–who often live apart–share resources and expend unpaid labour caring for one another. With a rapidly aging population, the need for this data is ever more pressing.

Last-mile healthcare (SDGs 1,3,5): Incredible progress has been made on generating gender data on health inequities. A vital, emerging frontier has to do with data on how to reach people, with high quality care, in the world's most isolated communities. For example, how far do women have to travel to reach a rural health clinic? What is needed to ensure quality of care upon arrival?

'Hidden' economies (1,8,9,16): Just allocation of economic resources depends on accurate data. Current data gaps on women and men's informal employment hinder policymakers' ability to extend social protections like maternity leave and pensions. And data on the impact of tax evasion and avoidance on gender equality could spur better regulations.

Gender-based violence and discrimination online (SDGs 3,5,16): The amount of time spent online is increasing, especially for young people. But we don't yet know how best to ensure women's and girls' rights online, or even the specific ways in which bullying and harassment, on the one hand, or trafficking on the other, occur. New methods and privacy-preserving ways of sharing data are needed.

STEM 'school to power' pipeline (SDGs 4,8,9,17): Global, comparable data on gender gaps in specific STEM jobs, such as data science, are scarce–but what does exist tells a concerning story. Data is needed that shows when and why girls educated in STEM fields drop or are pushed out of STEM careers, and what policies are needed to ensure they rise to power.

Shifting social norms (cross-cutting the SDGs): Political, economic and social life is rapidly changing, but to what extent are social norms–attitudes, beliefs and practices–changing too? For example, how are movements for gender equality impacting on men and boys? What works to change discriminatory beliefs about women's political participation?

In crafting this analysis, our intention is to bridge the gap between two communities that are not sitting at the same table often enough. The report aims to serve as a guide to feminist allies within the tech community who seek gender equality, not only within their own companies, but in the wider world.

Avenues for the Tech Community to Strengthen the Gender Data Ecosystem

This report considers four avenues through which the tech community could strengthen the gender data ecosystem for the achievement of the SDGs. We present these various avenues in light of what we learned about the current state of affairs regarding the permeation of data science in gender equality efforts (see Box 4). With this context in mind, we suggest ways the tech community could contribute to gender data projects through: sharing privacy-preserved data; processing and visualizing existing, external data for new insights; new data products and tools; and strengthening the capacity of feminist and women's organizations.

Box 4

Data science bottlenecks faced by feminist and women's organizations

Data science has yet to permeate gender equality efforts, and all of our informants offered ideas about why this might be the case. The issues raised by people working within feminist organizations tended to have relatively little overlap with the hypotheses we heard from the tech community. Here's what we heard from the gender and development community:

The available data isn't always useful to feminist and women's organizations. Often this is because of embedded biases and limitations to representativeness. It's also because different organizations have distinct data needs. Partnerships with feminist and women's organizations are needed from the outset to determine what is needed and to ensure its usefulness.

Lack of resources may be the biggest barrier to social sector organizations using unconventional data sets. Data scientists are expensive, and feminist organizations are chronically underfunded–this means that they often can't afford to hire the necessary in-house expertise.

There's a lack of knowledge sharing. Gender organizations often have very low visibility into what specific datasets tech companies possess, the terms under which data might be responsibly shared, and what other data-relevant resources the tech community might be in a position to make available. More dialogue is needed to create a baseline of mutual understanding from which to brainstorm impactful interventions.

The kind of partnerships that are required to leverage large datasets for gender equality require trust and credibility among involved parties. Gender equality advocates are all too aware that data alone isn't enough to achieve gender equality, and so they want to see the tech community taking steps to guarantee the rights of women and girls as they work in tech sector supply chains and conduct their lives using digital platforms.

One of the most obvious ways in which the tech community could contribute is by sharing its own data in a privacy-preserving manner. While conventional data sources such as national censuses and qualitative research form the foundation of public policymaking and service delivery, the data produced through digital technologies is distinctive in terms of timeliness, granularity, and the novel questions it may answer. When combined with conventional data sources, the digital trace data generated by, for example, social media platforms, video games, Fitbits, mobile phones, air travel, and power grids may help us understand new dynamics about people's everyday lives, and how inequalities are produced, sustained, and ultimately overcome.

Yet sharing its own data isn't the only way the tech community can contribute. Tech organizations could also extend their computational power and human resources to process and visualize existing, external data for new insights. For example, they might build models to predict the spread of specific gendered impacts of climate change. A third option is for the tech community to work with feminist organizations to understand what new data products and tools would be useful, and then build these. Meaningful consultation with feminist organizations in the conceptualization and design of data products and tools may yield surprising insights and is critical for ensuring that they

have on-the-ground impact (see Box 5). Fourth, they could contribute resources that strengthen the capacity of feminist and women's organizations to make use of unconventional datasets.

This final area of work is perhaps the one with the greatest potential for structural and sustainable change. Today, stark power imbalances exist between tech companies and social sector institutions. While the former possess vast resources in terms of data, computational processing power, and data-relevant human talent (e.g. engineers, data scientists), the latter possess thematic expertise, proximity to affected populations and grounded theories of change, but often lack the resources for large-scale data science projects. The production and use of quality gender data requires significant, system-wide investments that extend beyond the offices of tech companies, to the public sector, universities, non-governmental organizations, and women's rights groups.

Key elements of productive gender-data partnerships

The creation of datasets, tools, and products that achieve high-impact usage by gender equality organizations requires productive partnership from the outset. Three principles are key:

01 **Partnership from ideation to dissemination:** All too often, assumptions are made about what it is that gender equality advocates and service providers need in order to do their work more effectively or efficiently. By contrast, bringing feminist partners into the design process from the outset is more likely to result in a final product that becomes a high-value tool for organizations on the ground.

02 **Feminist influence over outcomes:** The participation of gender equality organizations should go beyond simply being in the room, to having power over the features of the dataset, tool or product.

03 **Get closer to the problem:** National and sub-national government bodies and more niche NGOs and civil society organizations can make great partners. These groups are not only closest to the problems, they're often closest to the potential solutions.

To be sure, bridging a data gap doesn't ensure that the lives of women and girls will improve. This is because gender equality isn't only inhibited by a lack of data.[6] Some of the most significant gaps exist in the political will to take action on problems about which there has long been ample data. There is perhaps no better example than in the case of gender-based violence, for which plenty of data has long existed.[7] Feminist and women's organizations also have concerns about the extent to which tech companies are willing to uphold women and girls' rights beyond data-for-development initiatives, making the privacy, safety and security of women and girls a priority in their corporate products and services. While these concerns are in some respects distinct from the topical themes we highlight below, they are all too relevant to the update of new data products and partnerships.

While data alone cannot guarantee the resolution of a problem, ensuring its usability by feminist academics and gender equality organizations is a critical step in ensuring its impact. In the following section we consider, from a feminist perspective, areas in which the tech community might help to address specific gender data gaps. In addition to longstanding issues of feminist concern, we highlight emerging demographic and environmental challenges for which we are only beginning to understand the gendered impacts, but all signs point to the need for swift action.

Gender data gaps across the SDGs

Inherent to the 2030 Agenda is a call for data-driven development. The demand for data has a direct and significant impact on National Statistical Offices, the backbone of public policymaking and service delivery, which are notoriously under-funded. It also introduced new power hierarchies in data production, dissemination, and access, as new data technologies and expertise were increasingly found in the private sector.[8] Within this landscape, tech companies have emerged as powerful 'data holders' and actors whose potential to contribute to achieving the SDGs is significant, but as of yet far from fully realized.

When the international development community talks about 'gender data gaps,'' they are typically referring to 'missing' or unavailable data as it relates to i) the UN Sustainable Development Indicator framework; or ii) the drivers of gender inequality and how to create positive change.

The 232 indicators that make up the SDG indicator framework are widely used to identify inequalities and to monitor global progress on the SDGs. Significant gender data gaps exist across the framework's 54 'gender indicators', which call for disaggregation by sex or refer to gender equality as an underlying objective.[9] Indeed, there exists sufficient and regularly collected data for only 10 of these 54 indicators. The 44 indicators without sufficient data include, for example, incidences of sexual violence (SDG 5), number of people living below 50 per cent of the median income by

sex and disability (SDG 10), and mechanisms for increasing capacity to address climate change that include women and youth (SDG 13).[10] Some SDGs have few or no gender indicators at all (see Box 6), which

feminist have suggested poses a barrier to global progress on gender equality in a development context where 'what is counted, counts'.[11]

Gender data gaps within the SDG indicator framework

The SDGs that lack gender indicators altogether include SDG 6: clean water and sanitation, SDG 7: affordable and clean energy, SDG 9: industry, innovation, and infrastructure, SDG 12: responsible consumption and production, SDG 14: life below water, and SDG 15: life on land. Yet the lack of gender indicators does not mean that gender does not have a bearing on these SDGs, nor that we should overlook gender data collection on the various facets of life that they encompass. Two examples highlight how collecting gender data on 'gender blind' SDG indicators could provide policymakers and funders with actionable insights.

Bridging a gender data gap around access to sanitation, SDG 6
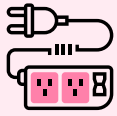Contribution type: Processing and visualizing external data

SDG Target 6.2 calls for access to adequate and equitable sanitation and hygiene for all, and to end open defecation, paying special attention to the needs of women and girls and those in vulnerable situations. But the indicator for tracking progress on this target only calls for the proportion of the population using safely managed sanitation services.[12] The experts we spoke with stressed the need for sex-disaggregated data on distance to safe and clean sanitation facilities, and on access to such facilities in the context of humanitarian crises. Such data could help researchers resolve key questions regarding the relationship between menstrual hygiene management (MHM) and girls' education. And sex-disaggregated data on the availability of "female-friendly toilets,"[13] including distance to them, could guide infrastructural investments and planning processes that address the safety risks that many women and girls face when accessing sanitation facilities that are not located in their own homes.

Check out these data-driven MHM initiatives
FACE Africa: http://www.faceafrica.org
Gender, Adolescent Transitions and Environment (GATE) research program at Columbia University: http://bit.ly/PeriodPosseResearch

Bridging a gender data gap around access to energy, SDG 7

Contribution type: Processing and visualizing external data

While SDG Target 1, ensure universal access to affordable, reliable and modern energy services, lacks gender indicators,[14] the reality in many households is that women and girls are responsible for sourcing fuel, and when that fuel source is hard to come by or is 'dirty,' suffer negative consequences on their time and health.[15] Sex and location disaggregated data on access to electricity and on primary reliance on clean fuels and technology is needed to push the energy sector towards more equitable and sustainable investments.

Existing tech initiatives around energy access show how processing and visualizing external data can produce new insights. Facebook's Electrical Distribution Grid Maps is a predictive model for energy access used by governments, NGOs and the private sector to plan infrastructure development. The model draws on imagery from the Visible Infrared Imaging Radiometer Suite (VIIRS) product, a NASA satellite imagery dataset that shows the earth at night; the MODIS land cover dataset, a global product providing land cover classifications for any location in the world; and roadway locations from OpenStreetMap. This initiative doesn't include an explicit gender component, but there are clear opportunities for it to be applied in service of gender equality. This kind of tool could be integrated into efforts by feminist engineers, grassroots women's organizations and national women's ministries to advocate for and design the extension of energy systems that alleviate the time and health burdens borne by women and girls where electricity is lacking.

Read more about this initiative at:

https://dataforgood.fb.com/tools/electrical-distribution-grid-maps

Yet some of the most pressing gender data gaps actually exist outside of the SDG indicator framework. These gaps pose a barrier to understanding the drivers of gender inequality (e.g. why are men more likely to own a mobile phone than women?), the dynamics of complex phenomena (e.g. how does deforestation impact women and men's livelihoods?), and how to create change (e.g. what works to change discriminatory attitudes?). These kinds of nuanced and context-specific gender data are deeply important for governments, NGOs, academics and private sector institutions to take informed action on the equitable allocation of resources, and for crafting effective policies, laws, and services.

These gender data gaps often cut across the SDGs, and bridging them sometimes requires attention to other data concerning for example income, geography, citizenship status, ethnicity and race, ability, sexuality, and religion (see Box 7). While it would be impossible to map all of the gender data gaps that have a bearing on the achievement of the SDGs, in what follows we highlight key areas of feminist concern where the tech community may be well-placed to contribute. These areas aren't definitive or exhaustive–they are meant to spark creative thinking and productive conversation.

Dilemmas around disaggregation

The more disaggregated a dataset is, the more useful it is likely to be for accurately answering tricky social questions. But from a practical perspective, it would be incredibly challenging (and possibly not very productive) to disaggregate all datasets by all of the possible factors by which inequalities are experienced (e.g. sex, age, location, race, ethnicity, citizenship status, ability, sexuality, etc.). We asked interviewees which factors, in addition to sex, tend to be the most useful for policymaking and service design. The most common responses were disaggregation by age and location (for example, see sections on family life, migration, and healthcare). What is most useful is highly context dependent, however, and partnerships with researchers and organizations closest to the affected communities are key to identifying where to start.

When it comes to user data, however, disaggregating by multiple factors also introduces safety risks. The higher levels of resolution–the more factors disaggregated–the easier it becomes to identify individuals. These tradeoffs need to be discussed transparently with partners in the gender and development community so that all parties can make informed, rights-based decisions about data production and use.
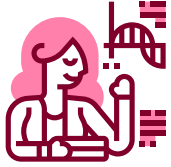
STEM 'School to Power Pipeline'

SDGs: quality education (4), decent work and economic growth (8), industry, innovation and infrastructure (9).

Women's under-representation in the highest levels of science, technology, engineering and mathematics (STEM) is resulting in innovations that fail to advance women's interests or reflect their lives and needs.[16] This disparity impacts the creation of new technologies, artificial intelligence, medical advancements, and infrastructure, and it is particularly acute in some STEM professions, such as data science. The sparse data available on gender disparities in specific STEM fields suggests a concerning story: for example, in the US, it is estimated that only 16% of all data scientists are women.[17] This is a problem of global and far reaching scale, as increasing the representation of women in STEM can have

impacts across the SDGs, such as the environmental goals (13, 14, and 15).[18]

In addition to focusing on the scale of the gaps themselves, there is a need for data on why they exist and what works–and doesn't work–to close them. For example, researchers are working to understand why women remain grossly under-represented in STEM professions, even in places where girls seem to be outperforming boys in STEM education.[19] In order to make effective policy recommendations, they need data on the "pain points"[20] –where and why women are stepping out, or being pushed out, of STEM career paths (see Box 8).

Box 8

Data-sharing to address persistent gender gaps in STEM fields
Contribution type: Sharing privacy-preserved datasets

Experts we spoke with voiced a need for cross-corporation, longitudinal data on recruitment, sex-disaggregated number of applicants, hiring, promotions and promotion ratios, retention (at the 1, 3, 5 year-mark), and use of family-friendly policies such as maternity/paternity leave, family leave and childcare. If the top ten tech companies in the world were to provide this sort of privacy-preserved data, the insights gleaned could provide critical new insights into the "bright spots"–what works to retain and advance women in STEM, as well as the pain points.

Examples of tech companies using their data to identify and understand gender gaps already exist. For example, LinkedIn partnered with the World Economic Forum to determine the gender gap in artificial intelligence skills.[21] Utilizing LinkedIn data, they found a significant gender gap in AI professionals (22% women) and gender differences in reported AI skills sets and occupational roles, with women more likely to be employed as data analysts or in research and teaching positions. Men showed higher AI skills and representation in more lucrative and senior positions, such as software engineers, heads of IT, and CEOs. More than a gender gap in AI skills, these results show that women are less involved in the creation and strategic direction of AI systems. Identifying field-specific gaps like the one in AI is a critical first step. The next step will be conducting research to understand exactly how these gaps are produced, and what can be done to address them.

Read more about what researchers at the Harvard Kennedy School are doing to understand pain points in the tech sector on their Gender Action Portal: https://gap.hks.harvard.edu/

The gender gap in STEM professions also impacts the capacity of social sector organizations to engage with large datasets because the pool of data scientists that care deeply about gender equality isn't as big as it could be. Part of the solution could be increasing the number of data scientists who can work competently and passionately in partnership with social sector organizations (see Box 9).

Box 9

Increase support for (gender) data scientists

Contribution type: Capacity building

Several non-profit initiatives exist that seek to **generate interest among girls in data science** from a young age, such as Girls Who Code (supported by Lyft and Uber), Girl Develop It, and TechGirlz. There are also mentorship programs, such as non-profit Girls in Tech Phoenix's LeadforCareer, which uses AI (including user data from LinkedIn) to match women in technology with young women hoping to work in STEM. And under its #MakeWhatsNext campaign, Microsoft generates interest and provides experiential learning and mentorship opportunities for girls in STEM on a significant scale.

There is also a need, however, to **increase the share of data scientists who use their skills to achieve gender equality.** While there are many ways this could be done, the gender equality experts we spoke with expressed a need for long-term, sustainable data science support. This means that models based on volunteering may not be always be appropriate. Voluntary labour can be less predictable than formal paid labour, and onboarding new volunteers, who many only donate a few hours a week, requires significant inputs of time and resources from NGO staff. The short-term nature of many volunteer engagements can also mean that easier to solve problems may be favoured over more complicated problems that require patient commitment.[22]

Generating sustainable models of (gender) data science support will require some innovation, but potential solutions may not be so far off. For example, many tech organizations already have successful internship programs. These could be leveraged to host data scientists to pursue work on issues related to gender inequality. Tech companies could also offer full-time secondment of their own data scientists to social and public sector organizations for a minimum of six month periods. And they could provide unrestricted resources to social and public sector organizations so that they can competitively hire in-house talent.

Yet another part of the solution is tackling the fact that social sector organizations struggle to compete with much better resourced industries, including the tech industry, for data science talent. This is especially the case for organizations with a mandate to achieve gender equality: the revenue used to fund UN Women's global work in 2019 was reported at USD 404.7 million. For comparison, this is less than 2 per cent of the annual revenues of the smallest tech company to make Fortune Magazine's Global 500 ranking.[23] To really invest in the gender data ecosystem, corporate social responsibility initiatives could place funding in the hands of the gender equality experts, to hire their own in-house expertise, and set their own data-related agendas.

Climate Change and Migration

SDGs: No poverty (1), decent work (8), responsible consumption and production (12), climate action (13), life below water (14), life on land (15), justice, peace, and strong institutions (16).

The World Bank projects that in three major world regions alone–Latin America, sub-Saharan Africa, and South Asia–unmitigated climate change will push around 143 million people to migrate by the year 2050.[24] Indeed, for all world regions, mitigating and adapting to climate change, and ensuring that migration is safe, orderly and regular, are two of humanity's biggest contemporary and future challenges. Yet the scope of these challenges is matched with a troubling dearth in information: these are two areas where accurate and nuanced gender data to guide policymaking and regulatory innovation is desperately needed.

This is especially so with regards to the environment. Qualitative feminist research has long shown that environmental change can have different impacts on men and women, and that these need to be acknowledged and addressed. For example, recent research indicates that climate change is driving gender-differentiated migration patterns,[25] and an increase in gender-based violence.[26] Yet quantitative, technical and physical-sciences-based information lacking a gender perspective is often privileged when it comes to solving environmental problems.[27] Where the application of big data has been explored in the context of environmental monitoring (e.g. combining government data with geospatial data producers and user communities in the US), gender isn't typically considered.[28] Effectively–and swiftly–bridging gender data gaps with actionable information will require productive partnerships and multiple forms of data brought into conversation in new ways (see box 10).

Box 10

Bridging gender data gaps for climate change response
Contribution: Processing external data; new tools and products

An enormous opportunity exists to develop new methodologies for producing gender data that can help governments, the private sector, and NGOs mitigate and respond to climate change. Currently, at the global and national levels, comparable pictures of the 'gender-environment nexus' are generated mostly through proxy data– for example, proportion of the population with a safe drinking water source, or sex-disaggregated data on the agricultural labour force. Partnerships between the tech community and academics, NGOs and public sector actors at the gender-environment nexus could explore initiatives to generate gender data that prompt positive action on climate change, such as data on:

✔ **Women's participation in sector-specific environmental governance bodies (e.g. communal land, water, forest, national energy utilities).**

✔ **Number and percentage of women and men who access employment or increase their incomes due to climate change adaptation or mitigation activities.**

✔ **Time saved in collecting and carrying water, fuel, and forest products due to environmentally sustainable and climate change adaptation activities'.[29]**

✔ **The impact of climate change on patterns of migration, disaggregated by sex and age.**

Gender data gaps regarding migration are similarly vast, including: sex-disaggregated data on the dynamics of internal migration (that which occurs within a country), which is more common than international migration; the extent to which all migrants, including those with irregular status, have access to social protections such as maternity leave and pensions; and the impact of migration on various elements of family life, including unpaid care of children and the elderly.[30]

Many experts we spoke with highlighted gender data gaps around experiences of human trafficking and other forms of sexual and gender-based violence in contexts of migration, which have related indicators tied to SDG 16.[31] These serious gender data gaps are correlated with the inadequate allocation of resources to track and respond to gender-based violence in crisis settings.[32] The tech community may have a role to play here, though the experts we spoke with

emphasized that bridging these gaps would require close partnership with feminist researchers and organizations with deep experience in sexual and gender-based violence. In any case, when moving populations are vulnerable, data generation comes with specific risks that must be taken seriously in order to ensure their human rights are upheld. Existing data-sharing projects for humanitarian response illustrate how this can be managed (see Box 11).

Data-sharing partnerships that protect human rights in disasters
Contribution: New tools and products

In the wake of a natural disaster, Facebook's Disaster Maps share real-time data on population movement with humanitarian response agencies. The data help these organizations determine things like whether communities have access to power and cellular networks, if they have evacuated, and what services and supplies they need most. Given that crisis affected populations are especially vulnerable, Disaster Maps use statistical techniques to maintain individuals' privacy, and Facebook only produces the Maps for carefully selected partners such as UN agencies and NGOs.

This type of data initiative could be used to support the efforts of organizations focused specifically on recognizing women's needs and guaranteeing women's rights in crises, such as CARE. Their Rapid Gender Analysis is designed by and for gender equality practitioners to quickly and practically gather context-specific gender data to inform humanitarian response.[33] The gender data that CARE has available is often scarce and imperfect, due to the nature of a crisis. Were sex-disaggregated migration data to be available on a large scale, organizations such as CARE could be better equipped to advocate for the efficient, effective, and equitable allocation of resources and provision of services (e.g. health, child and elder care, and gender-based violence prevention and response) in times of crisis.

Read more about Disaster Maps here:
https://dataforgood.fb.com/tools/disaster-maps
Read more about CARE's Rapid Gender Analysis here:
https://insights.careinternational.org.uk/images/documents/rapid-gender-analysis/GIE-Guidance-Note-Rapid-Gender-Analysis.pdf

Last-mile healthcare

SDGs: no poverty (1), good health and wellbeing (3), gender equality (5), industry, innovation and infrastructure (9), reduced inequalities (10).

Over half of the world's population lacks access to affordable healthcare.[34] NGOs, governments and donors are innovating to reach people in the most isolated corners of the globe, a concern that spans low, middle and high-income countries. While improved maternal care is a top concern in sub-Saharan Africa, care for a rapidly aging population will be one of this century's greatest challenges in rural areas in high-income countries. Healthcare is an inherently gendered area of life: not only do women interface with healthcare institutions because of their biological functions (they get pregnant and bear children), but also because they bear social responsibility for the health of others. For example, they provide the majority of unpaid care for children and elderly persons, and they make up the majority of Community Health Workers in lower-income settings. This means that when access to healthcare is poor, women often 'bridge the gaps' with their own unpaid or poorly paid labour.

While we increasingly have data about healthcare outcomes, there are gaps when it comes to understanding current barriers to access and how to efficiently overcome them, especially in low-resourced settings. Gender data for last-mile healthcare delivery requires a higher level of geographical granularity in order to account for health disparities that exist between and within urban and rural settings.

For example, the World Bank reports that 50-60% of people in low income countries live more than 8km from a healthcare facility.[35] Yet distance alone doesn't tell us enough about what it takes to access care. In rural areas, for example, lack of road infrastructure and public transportation delay women's arrival to hospitals, and in urban areas, delays can also occur when ambulances cannot enter dense slums.[36] Moreover, we need data on the availability and quality of care at health facilities; just because a facility exists does not mean that it is adequately staffed and stocked with the supplies needed to meet women's needs (see Box 12). Ensuring the universal right to healthcare requires layers of data including district-level data on distance and travel time to access a health facility, as well as information on the availability of services upon arrival.

Box 12

When access to a clinic is no guarantee of access to quality care

Contribution: Sharing privacy-preserved datasets

Guaranteeing the right to healthcare requires a lot more than brick-and-mortar clinics. The What Women Want campaign asked 1.2 million women from 114 countries what they wanted most in terms of maternal and reproductive healthcare. The top five demands included respectful and dignified care; water, sanitation and hygiene; medicines and supplies; more and better supported midwives and nurses; and fully functional health facilities located closer to home. These demands highlight needs and experiences that are all too often excluded from official healthcare indicators. But they have a serious impact on the ability of women to realize their right to healthcare, and to be supported as they care for dependents. For example, a study in the Journal of Surgical Research found that more than a third of the hospitals assessed in low- and middle-income countries did not have running water, meaning that they cannot safely perform surgeries.[37] Poor quality and inaccessible services also impact negatively on women's time, as they may need to travel long distances to reach a clinic, or return multiple times before the clinic is open, adequately staffed, or equipped with the necessary supplies. District level and even more granular data on the availability and quality of services can shed light on persistent health inequities and provide critical impetus for more just and efficient allocation of resources.

Read more about the What Women Want findings at:

http://bit.ly/WWWCampaign

Innovative healthcare organizations and donors are finding creative uses of data to map where healthcare coverage gaps exist, and to build models to understand what is required to close them. Some NGOs are using technology in savvy ways to build predictive models of healthcare access and then partner with Ministries of Health to implement improvements (see Box 13). Overall, however, the inability to afford in-house data scientists limits the ability of most frontline organizations to fully leverage all that the data has to offer (see Box 14).

Box 13

Data to build healthcare access models in isolated places

Contribution: Processing external data

Pivot is an organization that delivers systematic, data-driven health systems strengthening programs in rural Madagascar.[38] They have begun using DHS data on mortality and access to care, and layering this with rainfall data, elevation data, land-use data, time-use data, and Open Street Maps data that show all foot trails in the district. This mapping reveals barriers to care in a much more precise manner than would be possible with GPS data alone. Layered information enables Pivot to make predictive models about which people and communities face the greatest risk of lacking access to life-saving care; they can then use these models to improve healthcare delivery. Other NGOs like Medic Mobile and Village Reach also work on how to use new technologies in smart ways to meet healthcare delivery challenges in low-resource contexts. The tech community could offer its processing power to help bring these initiatives to scale around the world.

Read more about these health equity organizations here:

Pivot: https://pivotworks.org

Medic Mobile: https://medicmobile.org

Village Reach: https://www.villagereach.org

Box 14

Supporting the capacity of resource-constrained Last-Mile Healthcare NGOs

Contribution: Strengthening capacity

Data-driven 'last mile healthcare' organizations tend to be resource-constrained, meaning that although datasets around healthcare do exist, there is a capacity issue in terms of affecting real change. These NGOs often lack the resources to hire in-house data scientists to make creative use of existing data sets that could improve access to quality healthcare and as a result, promote gender equality. Tech companies could meaningfully engage in two, related ways; (1) they could second experienced data scientists to support those NGOs with existing data infrastructure; (2) they could provide "data catalyst grants" to boost creative data projects focusing on algorithm development, deployment, and iteration.

Family life and unpaid care work

SDGs: No poverty (1), good health and wellbeing (3), gender equality (5), decent work and economic development (8).

Feminist economists have long worked to develop accurate portraits of the division of unpaid care and domestic work within households and families. Unpaid care and domestic work includes activities such as grocery shopping, cooking meals and fetching water; minding, bathing, soothing and teaching children; doing laundry; cleaning the house; and caring for elderly parents and dependents (including in separate households) who are ill or living with a disability.[39] Accurate age and sex-disaggregated data on how much time people spend on care of other household members, and on dependents in other households, such as elderly parents, is critical for policymakers and employers to design just and effective policies and services, such as paternity leave, family leave, care grants, childcare and long-term care services. Data about how family members in the same and different households share financial resources and assets such as income, cars, technology, food and expenditure on health, education, and child and elder care, can also shed light on gender inequalities.[40]

Conventional methods for measuring 'time spent' on unpaid care and domestic work, such as household surveys and time diaries (what did you do from 8am-9am? And 9am to 10am?), are critical tools for policymakers, but they are also time-intensive, expensive, and can lack granularity. For example, people may undertake several activities at once, like cooking a meal, and "passively" minding a child, but they may only report on the cooking, because it was the more physically intensive activity.[41] Data that capture this multi-tasking could help shine a light on the "black box" of the family. Moreover, there is a need to develop methods that capture the time spent thinking about, planning, and managing care of dependents who may not be in physical proximity, for example an elderly parent who lives in another household (see Box 15).

Box 15

Using social network data to get inside the "black box" of the family.

Contribution: Sharing privacy-preserved datasets

Social network data could complement household surveys, contributing to what we know about the dynamics of family life within and outside household walls. This might include decision-making and care-taking dynamics such as income sharing and remittances, especially in the context of global population aging, in which adult children in one household care for elder parents in another, and migration, where many families are stretched across borders. Survey data is particularly limited when it comes to understanding situations where family life and care work are spread across multiple households.[42] But social media data and call-record data could help us understand family and care taking dynamics, including how resources are shared when families do not cohabit.

Putting multiple types of data into conversation is often required to achieve accurate insights into complex social issues. Take, for example, the question of time use and women's economic empowerment. Quantitative time-use data indicates that women spend a lot more time on childcare than men, in addition to their paid work, and this impacts negatively on their economic empowerment.[43] But qualitative data reveals that childcare isn't the only thing that women need: they also want decent work, and leisure time to rest.[44] Combining multiple data collection methods can give us further insight into family life and unpaid care work:[45] for example, time-use and household surveys, qualitative field research, and social network data could provide novel insights (see Box 16).

Box 16

Innovative uses of data to understand how family patterns are changing

Contribution: Sharing privacy-preserved datasets; new data tools and products

The Gender Equality Observatory within the Economic Commission for Latin America and the Caribbean (ECLAC) provides cutting-edge insights into the lives and autonomy of women and men. Their surveys have a clear analytical agenda; for example, the "Total Work Time" dataset highlights the importance of documenting both paid and unpaid labour in relation to each other, because the latter

acts as a barrier to women's participation in the formal labour market that could bring economic resources and increased autonomy.[46] Their "Femininity Index of Poor Households" compares the percentage of poor men and women, and highlights the reasons that women's lack of economic autonomy means that they are more likely to live in poverty.[47] To offer richer insights into the patterns of family dynamics for organizations like ECLAC, the tech community could (1) help develop new methods to get quantitative, granular insights on multi-tasking and time spent "on call" for care, (2) develop a targeted, online time use survey with attention to multi-tasking, and (3) share data on how people spend their time online to complement data on how people spend their leisure time.

'Hidden' economies

SDGs: No poverty (1), gender equality (5), decent work and economic growth (8), industry and innovation (9).

Plenty of work takes place outside of the formal, recognized sphere. In many countries formal employment does not cover even a majority of work. There are an estimated 740 million women in the informal economy worldwide, and in Africa, 85.8% of labour is in the informal economy.[48] Even those who work in the formal sector may concurrently work in the informal sector. Indeed, women make all kinds of contributions to economic life that aren't formally captured or recognized: for example, informal labour (in markets and family businesses), the production of artisan goods (selling on Etsy, through Facebook platforms), and in family farming.

'Hidden' economies are one of the areas where social media platforms have a huge impact on people's lives. It is possible that through privacy-preserved data sharing, the tech community could help feminist economists better understand the dynamics of how, where, and when women engage in informal employment, using this data to address gaps in access to social protection including health insurance and maternity leave. We also know that funding for social programming for women is stymied by tax avoidance and evasion; measuring the impact of tax evasion on gender inequality could be a big support to ensuring stronger regulation (see Box 17).

Box 17

Tax justice and funding for gender equality

Contribution: New products and tools

Around the world, a widely embraced strategy for reducing women's unpaid labour and addressing other inequalities is for states to progressively tax their citizens and use the revenue to fund social programs.[49] A study in Guatemala, Honduras, and the Dominican Republic found that their gender-blind tax policies effectively deepen inequalities between men and women.[50] In many countries direct taxation, and corporate and wealth taxes, as well as systems that prevent elite tax avoidance and evasion, help fund services like childcare and social protection programs that redistribute the burden of women's unpaid care work. Our informants urged the tech community to develop methods for measuring the impact of tax avoidance on gender inequality. In doing so, they could help make visible the social protection needs of those women working in 'invisible' economies as well as spur better tax regulation.

Shifting social norms

All of the SDGs

Donors like the Bill and Melinda Gates Foundation and the United Kingdom's Department for International Development (DFID) are funding new work on social norms, including gender norms, because they hear from their partner organizations that entrenched norms are hindering progress across the SDGs. For example, restrictive norms around women's mobility are a barrier to children's vaccination rates (when women aren't allowed to travel without a male companion to a far-away clinic). Feminist and women's rights organizations have been working concertedly on these issues for decades.

But social norms and how they change—including how they impact the progress of the SDGs—are incredibly hard to measure; in fact, there is no agreed upon methodology. The term "social norms" encompasses a wide range of issues. Norms affect everything from vaccination rates, to online and physical violence against female politicians (see Box 18), to men's take up of paternity leave when it's on offer (see Box 19). Among experts in our study, one of the most frequently emphasized data gaps had to do with our poor understanding of harmful gendered behavior online and how people engage with online content (see Box 20).

Box 18

Violence against female politicians online and IRL

Contribution: Sharing privacy-preserved datasets

Research suggests that women are more likely to face gendered forms of harassment and abuse online.[51] The Inter-Parliamentary Union reports that 58.2% of female MPs and parliamentary staff have been the target of online sexist attack on social media.[52] At the same time, we know that social media platforms like Twitter and Facebook are fertile ground for online engagement with voters and for participating in political debate, particularly for candidates with limited resources. Researchers working on digital inclusion and political violence highlight significant difficulties in analyzing the scope and nature of online violence and its relationship to offline variables. Social media data could provide insights into whether users who share political content or are political candidates are more subject to abuse - including threats or acts of violence, online or 'in real life' (IRL). This in turn could reveal barriers to women's full and effective participation in political life. There are real concerns with balancing user privacy and the rights of women and girls online, but these data could support the production of "hard" evidence that could in turn open entry points for accountability.

Box 19

Gender data gaps around men, boys, and changing masculinities

Contribution: Sharing data and strengthening capacity

There is much that we don't understand about men and boys' changing lives and perceptions of the world around them. Promundo is an applied research institute that seeks to promote gender equality by working with men and boys. Together with the International Research Centre on Women, they created the International Men and Gender Equality Survey,[53] which uses survey questions to gain insight into men's and women's understandings of gender-based violence, attitudes toward caregiving, and opinions on gender equality more broadly. The tech community could share social media data and/or strengthen the data science capacity of organizations like Promundo to help upend the assumption that gender equality is only for women and girls.

Read more about the International Men and Gender Equality Survey here:

https://promundoglobal.org/programs/international-men-and-gender-equality-survey-images/

Researchers at Cadi Ayyad University are producing insights into this gap by analyzing comments extracted with Youtube and Twitter APIs to gain information about attitudes regarding violence against women in the Arab States.[54] Measuring changes in behaviour and social norms over time, however, is methodologically challenging. Because these shifts are often context specific, it is hard to have a big dataset that includes meaningful points for comparison. This is exacerbated by a lack of standardization around concepts we are trying to measure, for example: "empowerment."

Tech companies could contribute to our understanding of these difficult-to-measure phenomena through the analysis of trace data. For example, data on people's opinions gathered through LinkedIn Campaign Manager could compliment demographic data to shed light on real time shifts in attitudes, as opposed to relying on temporal snapshots.[55] While there are clear ethical quagmires, as well as concerns around representativeness, trace data have the potential to expand our understanding of shifting social norms over time.

Understanding how social media interactions impact behavior

Contribution: Sharing privacy-preserved data and lending processing power

Emerging qualitative evidence indicate that the online bullying of adolescent girls and LGBTQI youth can lead to self-harm, suicide, and suicidal ideation. Oversight bodies such as the UK Royal College of Psychiatrists are calling on social media companies to provide academics with data on online viewing patterns in order to develop targeted and effective responses to suicide risk.[56] In a related vein, behavioural economists and NGOs are trying to understand how negative images and storylines--as well as sponsored content, 'influencers', and advertisements related to health, beauty, and 'wellness'- influence social norms (attitudes, beliefs and behaviour) offline. And on the flipside, they are testing whether exposure to positive images and storylines can change social norms for the better. Comparative and granular data about where, when, and in relation to what factors discrimination, negative emotional impact, and bullying occurs, could help academics, INGOs, and civic organizations like the Geena Davis Institute on Gender in Media gain insight into the nexus between social media interactions and offline behaviour.

Partnering to secure the rights of women and girls in gender data projects and partnerships

One of the most important insights that emerged from this research is that although women's rights stakeholders and technology companies rarely sit at the same table, they do share some key common concerns about data generation and governance – particularly as these relate to questions of privacy and safety. These shared concerns highlight a common ground between the tech community and gender equality organizations – and thus an entry point for building productive alliances.

To be sure, we also heard about diverging priorities. Amongst public sector and civil society actors, a prevalent concern was that tech companies' incentives and relationships with their users are distinct from those that govern states' relationships with their citizens. While governments have a duty to protect their populations and guarantee human rights, private companies are not held to the same kind of standard.[57] Women's rights organizations are all too familiar with the model of corporate social responsibility that involves offering benefits to women and girls in one region or sector while facilitating or perpetuating abuses of women and girls' rights in another. The very nature of feminist organizations' work is such that they will expect potential partners in the tech community to demonstrate a commitment to upholding the rights of women and girls in the broader remit of their work – not only in data for good initiatives.

Many concerns, however, were less about intractable differences between the goals of the two communities, and more indicative of a lack of mutual understanding and practicable working arrangements – problems that can be addressed through

ongoing dialogue. For example, there was reasonable skepticism regarding the often over-hyped 'promise' of digitally produced data to catalyze positive outcomes for all people, given embedded biases and limitations to representativeness in many datasets. But these concerns don't need to shut down the conversation. Working together, feminist researchers and data scientists in the tech community could come up with hundreds of problems for which big data could be thoughtfully deployed to help understand and resolve. There were also calls for the mass-release of privately held data for the purposes of exposing discrimination and violence online. Here, it is incumbent upon the tech community to better communicate to the gender and development community the human rights-related risks of 'open data.'

This research made clear that productive partnerships between the tech and gender equality communities – partnerships that balance novel insights with genuine care for women's rights – require mutual understanding and the identification of shared priorities. It is also clear that the design of novel systems, data products, and working arrangements that creatively address these ethical concerns will need to be a concerted focus of innovation efforts, if the gender data ecosystem is to thrive in the coming years. In light of this opportunity for innovation, the following section highlights concerns from the gender and development community, as well as areas of common ground.

Addressing Bias and Representativeness

Gender equality advocates receive grand declarations about the promise of 'big data' with skepticism because they are familiar with inequalities in digital access. They know, for example, that social media data presumes access to the internet, which already excludes about 60% of the world's population,[58] and that the gender gap in digital access is almost 12% globally and reaches 32.9% in least-developed countries.[59] This issue of 'non-coverage' understandably

raises red flags for organizations that have mandates grounded in the human rights principle of universality–for example, to extend maternal health coverage to all women, or to ensure that all girls and boys receive a quality education.

Yet bias and limitations to representativeness are not intractable problems. Plenty of feminist thinking is being done about how to surface and

address biases that are 'baked into' even so-called 'raw' data.[60] We see promise in partnerships with feminist academics who are experts in gender bias (and other biases such as those related to race, age, ability, sexuality, geography), and who are experienced in developing and deploying research methodologies for addressing bias. And from the data science side, strides are being taken to ensure that the appropriate uses of freely available datasets are clearly communicated, for example, through accompanying "datasheets" (see Box 21).

Amplifying the appropriate use of open data

A key issue with "open data" is that all too often, datasets are made available without any accompanying information on their origins, limitations, or suitable uses (we found this to be the norm in our analysis of existing gender-related datasets). This is a problem because the data can be used to generate inaccurate answers to important research questions and guide poor policy design. Researchers at Microsoft, Georgia Institute of Technology, Cornell University, the University of Maryland, and the AI Now Institute have a simple solution: they propose that all datasets should be accompanied by a "datasheet" that transparently outlines its "motivation, creation, composition, intended uses, distribution, maintenance, and other information."[61] Their intended application is to reduce opportunities for unintended bias in machine learning systems, but datasheets could be helpful in many other scenarios in which unconventional datasets, or new data products and tools, are made available to gender and development partners or the public at large.

Read about Datasheets for Datasets at:

https://www.microsoft.com/en-us/research/uploads/prod/2019/01/1803.09010.pdf

As the gender data ecosystem matures, conversations about the 'promise' of big data are also taking a more nuanced, humble, and ultimately more useful tone. In particular, there is increasing recognition that the goal should not be to replace conventional forms of data with big data, but to explore what unique insights can be produced when the two are taken together.[62] For example, while not everyone is on social media, there are ways that social network data in areas of high use could complement household survey data to offer novel insights into how families allocate time and money caring for family members living in different households (see Section 2).

Better communication about privacy and security is needed

The preservation of women and girls' privacy and security is an area where the gender and development and tech communities share a lot of common ground. Yet crucial components of the conversation around privacy, security, and risk in data governance and use seem to be getting 'lost in translation.'

For example, feminist researchers and women's and girls' rights advocates frequently ask tech companies to release their data for the purposes of exposing discrimination and violence on digital platforms. From their perspective, evidence of this nature is critical for data-driven advocacy and to support the creation of adequate regulatory frameworks.[63] Yet the tech community's rationale for proceeding with caution here–including rationale that aligns with a women and girls' rights perspective–is often unclear.

For example, there is a need for pointed discussion about the specific ways in which releasing social media message content to expose violence and discrimination could also mean exposing sensitive communications of vulnerable populations: for example, between migrant families organizing safe border crossings and

reunification, or women helping one another to exit violent relationships. There remains too little appreciation among gender advocates of the risks of 'data joining', where anonymized data from different sources is pulled together in ways that can re-identify individuals and expose intimate details of people's lives. Data joining would be a particular concern if companies like Facebook, LinkedIn, and Uber were to all simultaneously release sizable gender datasets—an action called for by many advocates and policy makers.

But for the gender and development community to understand these risks, the tech community needs to do a better job of explaining the kinds of data it has, the opportunities and risks inherent to sharing different kinds of data, and the limits of disaggregation by multiple identity factors (e.g. disaggregating by sex and age while also providing granular resolution in rural areas may not be safe) (see Box 22). This would also help establish some good will and shared understanding. Feminist researchers and women's rights advocates are very likely to be sympathetic to the idea that bridging certain gender data gaps introduces risks for unintended consequences that are untenable from a human rights perspective.

Knowledge-sharing around protection of women and girls' rights in data-sharing

While 'weekend workshops' on 'big data' aren't on their own adequate to strengthen the data science capacity of gender equality organizations (rather, see boxes 9 and 13 on capacity building), there is a need for pointed dialogue on data privacy and security. Tech companies should consider hosting knowledge-sharing dialogues in which they share basic information on the latest technologies for privacy preservation, and nuanced information on the promises and pitfalls of different types of data sharing. This would give gender equality advocates an opportunity not only to learn, but also provide context and feedback from the perspective of the populations that they serve.

Advancing gender equality in all of our work

The largest tech companies invest dramatically more money in data than do most national governments. While many of the experts we spoke with were enthusiastic about partnerships with the tech community to bridge gender data gaps, they were steadfast that private sector data should not displace direly needed investments in national statistical offices and public research institutions.

This report holds that there is room and need for both communities to be working together in this space. There are legitimate concerns around how to build trust and alliances between two communities with such disparate levels of power and resources. The gender and development experts we spoke with were clear: to

establish credibility in the space, they wanted to see the tech community take genuine steps towards upholding women and girls' rights in all of their corporate endeavors. This view is in line with recent Guidance from the Office of the United Nations High Commissioner on Human Rights that establishes all data collectors as duty-bearers. This means that governments and tech companies that collect data are accountable for upholding and advancing human rights in and through their operations.[64]

Recent examples of tech companies taking steps to show humility and establish trust and credibility are encouraging and should build confidence among peer organizations to take further steps in this direction. Take

for example Uber's recent transparency report about sexual violence in ride-shares.[65] As the world's most popular ride-sharing app, Uber's efforts to have an honest reckoning with the safety concerns of women users has opened up space for action to be taken. These efforts can have a ripple effect, too: competing ride-share company Lyft has since pledged that they will release a similar report.

There is also the recent initiative by Riot Games, a gaming company that partnered with academic researchers in psychology, cognitive science, and neuroscience to address widespread harassment and abuse in the chat platform of one of its games, which alone had more than 67 million players every month. Drawing from the team's findings – for instance, that much like laws that govern societies 'offline', online communities benefit from explicitly laid out understanding of the enforceable and shared rules – the company implemented reforms that significantly curbed abusive online messages.[66]

To be sure, such initiatives come with some risks as well as rewards. Joint commitments among tech companies to uphold women and girls' rights on their platforms and in their data-sharing efforts are one way forward. Such commitments would find supportive audiences among many feminist organizations, particularly if paired with facilitated opportunities for ongoing dialogue, meaningful efforts to advance gender equality both within and beyond data for good initiatives, and innovation focused on establishing the technical and operational basis for a growing gender data ecosystem to guarantee women's and girls' rights.

Conclusion

The 2030 Sustainable Development Goals Agenda aims to 'leave no one behind.' It is our hope that this report stimulates collaborative work, between the tech sector and researchers, practitioners and advocates who work everyday for gender equality, toward developing a set of clear, achievable, and ultimately feminist steps for turning this promise into a reality.

This report is rooted in the idea that in order to make substantive change, these two communities need to sit at the same table and learn how to speak each others' language. It will be through meaningful engagement - and an identification of shared goals - that new gender data initiatives will provide new means of understanding and addressing global inequality.

As this report details, there are many opportunities to build alliances between the tech community and the gender and development community to reduce gender inequality. We have identified overlaps and intersections - between the SDGs, between data gaps, between stakeholder concerns, and between potential sources and types of data - in order to flag areas that are ripe for collaboration.

From the outset, these alliances will require recognition of common ground, a substantive vision of partnership, and plenty of dialogue to share technical and thematic expertise. Some academics, policymakers, and civil society workers we spoke to have significant concerns about the involvement of tech companies in addressing gender data gaps. We also heard a great deal of open-mindedness and even excitement, as well as clear-eyed recognition of the time and effort that it will take to build bridges between these communities. We hope that this report will serve as a launch pad for the growing number of individuals and organizations who are committed to building these bridges and to establishing a thriving gender data ecosystem in the years to come.

References

[1] Fuentes, L., & Cookson, T. P. (2019). Counting gender (in)equality? a feminist geographical critique of the 'gender data revolution'. Gender, Place & Culture, 1-22.

[2] We draw on United Nations guidance for this definition. On gender statistics specifically, see: Zheng, S. (2015). What are gender statistics? Gender Statistics Manuel.

[3] For example, most gender data do not account for trans identities, or how to categorize intersex people.

[4] See, for example: UN Women. (2019). Mainstreaming Gender in Environment Statistics for the SDGs and Beyond: Identifying priorities in Asia and the Pacific; UN Women. (2018). Turning Promises into Action: Gender Equality in the 2030 Agenda for Sustainable Development.; Data2X, Grantham, K. (2020). Mapping Gender Data Gaps: An SDG Era Update.; Equal Measures 2030. (2019). Harnessing the Power of Data for Gender Equality: Introducing the 2019 EM2030 SDG Gender Index.

[5] In this report we refer to the "gendered impacts" of things like climate change and unpaid care work. By this we mean the way in which something impacts on the responsibilities, needs, experiences, and opportunities of women, men, girls and boys in different ways.

[6] Liebowitz, D. & Zwingel, S. (2014). Gender Equality Oversimplified: Using CEDAW to Counter the Measurement Obsession. International Studies Review, 16, 362-389.

[7] Krug, EG. et al., eds. (2002) World report on violence and health. Geneva: World Health Organization.

[8] Fisher, A., & Fukuda-Parr, S. (2019). Introduction: Data, Knowledge, Politics and Localizing the SDGs. Journal of Human Development and Capabilities, 20(4), 375-385.

[9] UN Women. (2018). Turning Promises into Action: Gender Equality in the 2030 Agenda for Sustainable Development. New York.

[10] UN Women. (2018). Turning Promises into Action: Gender Equality in the 2030 Agenda for Sustainable Development. New York.

[11] Regarding SDG 16, see: Goetz, A. M., & Jenkins, R. (2016). Gender, Security, and Governance: The case of Sustainable Development Goal 16. Gender & Development, 24(1), 127-137.

[12] Indicator 6.2.1: Proportion of population using safely managed sanitation services, including a hand-washing facility with soap and water.

[13] MacRae, E. R., Clasen, T., Dasmohapatra, M., & Caruso, B. A. (2019). 'It's like a burden on the head': Redefining adequate menstrual hygiene management throughout women's varied life stages in Odisha, India. PloS one, 14(8).

[14] Indicator 7.1.1: Proportion of population with access to electricity; Indicator 7.1.2: Proportion of population with primary reliance on clean fuels and technology.

[15] Ibid.

[16] D'Ignazio, C., & Klein, L. (2020). Data Feminsm: MIT Press; Perez, C. C. (2019). Invisible Women: Data Bias in a World Designed for Men. New York: Adams.

[17] Harnham. (2019). US Data & Analytics Diversity Report.

[18] Seager, J. (2019). Gender Equality and Environmental Sustainability in the Age of Crisis. Paper presented at the Sixty-fourth session of the Commission on the Status of Women (CSW 64), New York.

[19] Moss-Racusin, C., Dovidio, J., Brescoll, V., Graham, M., & Handelsman, J. (2012). Science Faculty's Subtle Gender Biases Favor Male Students. Proceedings of the National Academy of Sciences, 109(41), 16474-16479.

[20] Krivkovich, A., Kutcher, E. & Yee, L. (2014) Breaking down the gender challenge. McKinsey Quarterly, March.

[21] WEF. (2018). Global Gender Gap Report 2018. Geneva: World Economic Forum.

[22] Hooker, S. (2018). Why "data for good" lacks precision. Medium.

[23] See: https://fortune.com/global500/2018/search/?sector=Technology

[24] Rigaud, K. K. et al. (2018). Groundswell: Preparing for Internal Climate Migration. Washington: World Bank.

[25] Rao, N. (2019). A qualitative comparative analysis of women's agency and adaptive capacity in climate change hotspots in Asia and Africa. Nature Climate Change, 9, 964-971.

[26] Castañeda, I., Sabater, L., Owren, C. & Boyer, A. E. (2020). Gender-based violence and environment linkages: the violence of inequality. Gland: International Union for the Conservation of Nature.

[27] Seager, J. (2019). Gender Equality and Environmental Sustainability in the Age of Crisis. Paper presented at the Sixty-fourth session of the Commission on the Status of Women (CSW 64), New York.

[28] Susha, I., Gronlund, A., & Van Tulder, R. (2019). Data Driven Social Partnerships: Exploring an emergent trend in search of research challenges and questions. Government Information Quarterly, 36, 112-128.

[29] For further information on these and other indicators developed by organizations focused on gender and the environment, see: UNEP. (2018). Gender and environment statistics: Unlocking information for action and measuring the SDGs. Nairobi: United Nations Environment Programme.

[30] UN Women (2019). Progress of the World's Women 2019-2020: Families in a Changing World. New York.

[31] 16.2.2: Number of victims of human trafficking per 100,000 population, by sex, age and form of exploitation; 16.2.3: Proportion of young women and men aged 18 29 years who experienced sexual violence by age 18.

[32] Zulver, J. (2019). At Venezuela's border with Colombia, women suffer extraordinary levels of violence. The Washington Post.

[33] Quay, I. (2019). Rapid Gender Analysis and its use in crises: from zero to fifty in five years. Gender and Development, 27(2), 221-236.

[34] McNeill, K. & Jacobs, C. (2019). Half of the world's population lack access to essential health services – are we doing enough? World Economic Forum.

[35] Amos, P. (2008). Safe, Clean, and Affordable... Transport for Development. Transport Sector Board. Washington: World Bank.

[36] Rajé, F., Tight, M., & Pope, F. (2018). Traffic pollution: A search for solutions in a city like Nairobi. Cities, 82, 100-107.

[37] Chawla, S. et al. (2016). Water availability at hospitals in low- and middle-income countries: implications for improving access to safe surgical care. Journal of Surgical Research, 205(1), 169-178.

[38] Ezran, C. et al. (2019). Assessing trends in the content of maternal and child care following a health system strengthening initiative in rural Madagascar: A longitudinal cohort study. PLoS medicine, 16(8).

[39] Razavi, S. (2007). The Political and Social Economy of Care in a Development Context. United Nations Research Institute for Social Development.

[40] Folbre, N. (2018). Developing Care: Recent Research on the Care Economy and Economic Development. Ottawa: IDRC.

[41] Folbre, N., & Yoon, J. (2007). What is child care? Lessons from time-use surveys of major English-speaking countries. Review of Economics of the Household, 5, 223-248.

[42] UN Women. (2019). Progress of the World' Women 2019-2020: Families in a Changing World. New York.

[43] Coffey, C. et al. (2020). Time to Care: Unpaid and underpaid care work and the global inequality crisis. Oxford: Oxfam.

[44] Chopra, D., Nazneen, S. & Krishnan, M. (2019). Women's Economic Engagement and Childcare: Moving from Survival to a "Triple Boon". Policy Brief. Brighton: IDS

[45] Folbre, N. (2018). Developing Care: Recent Research on the Care Economy and Economic Development. Ottawa: IDRC.

[46] ECLAC. (2020). Total Work Time. Gender Equality Observatory.

[47] ECLAC. (2017). Femininity Index of Poor Households. Gender Equality Observatory.

[48] ILO 2018. Men and Women in the Informal Economy: A Statistical Picture. Geneva: International Labour Organization.

[49] Donald, K. & Moussie, R. (2016). Redistributing Unpaid Care Work: Why Tax Matters for Women's Rights. IDS Policy Briefing.

[50] Rodriguez, C., & Itriago, D. (2019). Do Taxes Influence inequality between men and women? Oxfam Research Report – Summary.

[51] See: Amnesty International 2017

[52] Broadband Commission for Sustainable Development (2015), Gender Cyber Violence Against Women and Girls: A World-Wide Wake-Up Call. A Discussion Paper from the UN Broadband Commission for Digital Development Working Group on Broadband and Gender.

[53] As of 2018, IMAGES studies had been carried out in more than 27 countries around the world. Studies inspired by IMAGES have been conducted in at least another 14 countries.

[54] Zahir, J et al. (2020). Mining the Web for Insights on Violence Against Women. Data2x.

[55] Cesare, N., Lee, H., McCormick, T., Spiro, E., & Zagheni, E. (2018). Promises and pitfalls of using digital traces for demographic research.

Demography, 55(5), 1979-1999.

[56] Campbell, D. (2020). Social media firms 'should hand over data amid suicide risk.' The Guardian.

[57] In some of our conversations with women's rights stakeholders, it was human rights principles - and not the more common refrain of 'data ethics' - that were highlighted as key to the data governance conversation. Some experts noted that unlike internationally recognized human rights, 'ethics' are not legally-binding and might be vulnerable to manipulation by anti-democratic states or actors who use the language of 'ethics' to target marginalized populations.

[58] Verhulst, S., & Young, A. (2017). Open Data in Developing Economies: Toward Building an Evidence Base on What Works and How. ODI.

[59] UN Women (2018). Gender equality and big data: Making gender data visible. New York.

[60] D'Ignazio, C. & Klein, L. (2020) Data Feminism. Cambridge: MIT Press; Masters, C.L. (2015). Women's ways of structuring data. Ada: A Journal of Gender, New Media, and Technology, 8.

[61] Gebru, T. et al. (2018). Datasheets for Datasets. Proceedings of the 5th Workshop on Fairness, Accountability, and Transparency in Machine Learning, Stockholm, Sweden, PMLR80, 2018.

[62] Data2x. (2019). Big Data, Big Impact? Towards Gender-Sensitive Data Systems.

[63] O'Connell, A., & Sweetman, C. (2018). Introduction: Gender, development and ICTs. Gender & Development, 26(2), 217-229.

[64] Office of the High Commissioner on Human Rights (2018). A Human Rights-Based Approach to Data. Available at: https://www.ohchr.org/Documents/Issues/HRIndicators/GuidanceNoteonApproachtoData.pdf

[65] Sidiqqui, F. (2019). Uber discloses 3,000 reports of sexual assault on US rides last year in long-awaited safety study. Washington Post.

[66] Hudson, L. (2014). Curbing Online Abuse Isn't Impossible. Here's Where We Start. WIred.